8-9-2022

Classification models for 2,4-D formulations in damaged Enlist crops through the application of FTIR spectroscopy and machine learning algorithms

Benjamin Blackburn
Mississippi State University, Benjamin.Blackburn1@uga.edu

Classification models for 2,4-D formulations in damaged Enlist crops through the application of FTIR spectroscopy and machine learning algorithms

By

Benjamin Blackburn

Approved by:

Darrell L. Sparks (Major Professor)
Ashli Brown-Johnson
Daniel B. Reynolds
Michael A. Caprio
Natraj Krishnan (Graduate Coordinator)
Scott T. Willard (Dean, College of Agriculture and Life Sciences)

A Document Type.
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Agricultural Life Sciences
in the Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology

Mississippi State, Mississippi

August 2022

Name: Benjamin Blackburn

Date of Degree: August 9, 2022

Institution: Mississippi State University

Major Field: Agricultural Life Sciences

Major Professor: Darrell Sparks

Title of Study: Classification models for 2,4-D formulations in damaged Enlist crops through the application of FTIR spectroscopy and machine learning algorithms

Pages in Study 129

Candidate for Degree of Master of Science

With new 2,4-Dichlorophenoxyacetic acid (2,4-D) tolerant crops, increases in off-target movement events are expected. New formulations may mitigate these events, but standard lab techniques are ineffective in identifying these 2,4-D formulations. Using Fourier-transform infrared spectroscopy and machine learning algorithms, research was conducted to classify 2,4-D formulations in treated herbicide-tolerant soybeans and cotton and observe the influence of leaf treatment status and collection timing on classification accuracy. Pooled Classification models using k-nearest neighbor classified 2,4-D formulations with over 65% accuracy in cotton and soybean. Tissue collected 14 DAT and 21 DAT for cotton and soybean respectively produced higher accuracies than the pooled model. Tissue directly treated with 2,4-D also performed better than the pooled model. Lastly, models using timing and treatment status as factors resulted in higher accuracies, with cotton 14 DAT New Growth and Treated models and 28 DAT and 21 DAT Treated soybean models achieving the best accuracies.

DEDICATION

I want to dedicate this thesis to my parents, Laura and Henry Blackburn, and my brothers, William and Daniel. To Laura, the strongest and most caring mother in the entire world. Solo hay una. To Henry, the most intelligent and kindest man I know who never gives up. To William, the best older brother someone could ask for and who I do not thank enough. And to Daniel, my twin brother who lives so far away now but has always been by my side. I couldn't have done it without the support and love you all have shown me. TQM.

ACKNOWLEDGEMENTS

I want to acknowledge all of my committee members. First, Darrell Sparks, thank you for all your help in being my undergraduate and graduate degrees advisor. Ashli Brown in accepting an undergraduate researcher on a whim and treating me like one of her own. Daniel Reynolds for taking on and teaching a biochemist how to be a plant and soil scientist. And Michael Carpio for teaching me the ins and outs of RStudio without asking anything in return. I will be forever thankful for all the guidance and assistance you have shown me.

I would also like to acknowledge Graham Oakley, Beau Varner, Matthew Taylor, Samuel Reeves, Cody Wilhite, Nicole Glenn, Holden Herrod, Jake Weathersby, and Doug Messer. You have all helped a city boy from Pascagoula become a little country. Thank you for teaching me how to drive a tractor and a four-wheeler, backpack spray, plant row crops, enjoy Brooksville and Ramsey as much as possible and survive what seemed like an endless amount of volatility and drift trials. None of this would've been possible without the help you have all given me. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THESIS INTRODUCTION

In 2019, agriculture, food, and other related industries accounted for around 5 percent of the United States gross domestic product, or around 1.1 trillion dollars, and food accounts for around 12 percent of U.S. household expenses (*Ag and Food Statistics Charting the Essentials, February 2020*, 2020). A significant problem for many farmers and scientists, and particularly agriculture, is Herbicide-resistant (HR) weeds (Perotti et al., 2020), which have cost farmers $187 million in additional herbicide treatments costs in Australia alone (Llewellyn et al., 2016). Herbicide resistance is the inherited ability of a weed or crop to survive a herbicide application. This HR can be caused by mutations found in an individual weed that then reproduces and produces offspring that are also HR. Using the same mode of action (MOA) without consideration of other MOAs or other weed management options can result in more HR weeds in a population (Vencill et al., 2022). At the time of writing, there are currently 266 HR weed species, with 55 of those HR weed species being resistant to Glyphosate (Heap, 2022). Herbicide-tolerant (HT) crops are new biotechnologies that can help mitigate HR. These HT cultivars are found in multiple row crops like soybean (*Glycine max* L.), cotton (*Gossypium hirsutum* L.), and corn (*Zea mays* L.). These cultivars have been genetically modified to resist multiple herbicides like glyphosate, 2,4-D, dicamba, or glufosinate. In 2019, over 90 percent of planted U.S. corn, upland cotton, and soybeans were genetically engineered, with 94 percent of domestic soybeans, 95 percent of domestic cotton, and 89 percent of domestic corn being

herbicide-tolerant (*Recent Trends in GE Adoption*, 2020). One new line of HT cultivars is the Enlist® weed control system ('Enlist crops', Corteva Agriscience, Indianapolis, IN 46268) which is resistant to 2,4-D and glyphosate in all cultivars, with the addition of glufosinate tolerance in the cotton and soybean cultivars. 2,4-D was first discovered in the 1940s and is a synthetic auxin herbicide mimicking natural auxins, which are phytohormones that control growth (Song, 2014). Although 2,4-D is one of the most commonly used herbicides globally, its exact mode of action is not fully understood (Tu et al., 2001). According to the EPA, 2,4-D functions by causing unregulated cell division in vascular tissue by increasing the cell wall's plasticity, increasing the biosynthesis of proteins, and increasing ethylene production (*2,4-D Technical Fact Sheet*, n.d.). 2,4-D tolerance is produced by employing transgenes encoding aryloxyalkanoate dioxygenase enzymes, or AADs, that can metabolize 2,4-D into its non-active metabolites (Wright et al., 2010). These HT crops will allow farmers and applicators to apply multiple herbicides with multiple MOAs, which can help mitigate the spread of HR weeds (Beckie, 2011; Riar et al., 2013). Farmers and applicators have also implemented Enlist cultivars to protect against possible off-target movement (OTM) damage caused by dicamba or 2,4-D, as both soybean and cotton are susceptible to auxin herbicides, and for the ability to use 2,4-D as a postemergence (POST) herbicide to control weeds that are HR to glyphosate or other MOAs (Lyon et al., 1993; Shyam et al., 2021; Singh & Sharma, 2000). Although using 2,4-D as a POST herbicide can certainly help farmers combat HR weeds and help mitigate the spread of these weeds, a problem has arisen that could negate the potential of 2,4-D: off-target movement (OTM).

Off-target movement is the movement of herbicides from the targeted crop to other off-target crops, including susceptible crops. Off-target movement mainly occurs because of primary

and secondary drift. Primary drift, also known as spray drift, occurs when wind or improper application equipment blows the herbicide away from the intended target site. Primary drift is mainly influenced by meteorological factors like wind, application factors like nozzle type and size, and formulation factors like viscosity (Carlsen et al., 2006). Primary drift can typically be minimized by following the guidelines found in many 2,4-D labels, which recommend large droplet sizes, applying 2,4-D when wind speeds are below 15 mph, not applying 2,4-D during atmospheric inversions, and using the proper equipment (*Nufarm Weedar 64 Broadleaf Herbicide*, n.d.; *Unison*, n.d.). Secondary drift, or vapor drift, occurs after application when a liquid herbicide evaporates and converts into a gaseous vapor that moves off the application site into an unknown area. Although many factors influencing secondary drift are still not characterized, vapor pressure, temperature, uptake of the chemical, pH of the environment, and atmospheric factors all influence secondary drift (Bish et al., 2021). In 2021 alone, there were 3,461 dicamba incidents, affecting more than 1 million acres of soybean reported in the U.S. (Tindall et al., 2021). Dicamba is another synthetic auxin herbicide producing many OTM events; however, in a herbicide drift survey of around 300 Midwestern growers, forty-four percent named 2,4-D as a herbicide responsible for OTM injury (Tindall et al., 2021). OTM of 2,4-D produces similar visual symptomology in both soybeans and cotton, including leaf cupping and stunting, chlorosis, altered height, epinastic response in both the stem and petiole, and callus formation in the stems (Andersen et al., 2004; J. T. Buol et al., 2019; Egan et al., 2014; Marple et al., 2008a). During OTM events, small concentrations of the applied herbicides will fall into susceptible crops, but even minuscule amounts of 2,4-D can cause injury to susceptible crops like non-tolerant soybeans and cotton (Lenny Wells et al., 2019; Marple et al., 2008b; Robinson et al., 2013), with some crops like grapes seeing damages with 2,4-D rates of $1/300^{th}$ of an

applied rate (Mohseni-Moghadam et al., 2016). To mitigate the possibility of 2,4-D OTM occurring, different formulations of 2,4-D have been constructed to reduce the volatility of 2,4-D. A herbicide formulation is a compound containing the same parent compound, in this case, 2,4-D, with different salt groups attached to it, like an ester or choline salt, as well as different adjuvants. 2,4-D has many different formulations like 2,4-D choline, 2,4-D acid, and 2,4-D dimethylamine, which all have different properties like solubility and vapor pressure (M. A. Peterson et al., 2016). The choline formulation of 2,4-D is the only formulation labeled to be used in Enlist crops, and it is the least volatile formulation of 2,4-D (Sosnoskie et al., 2015a)**.** Although 2,4-D choline is the only formulation labeled for use over HT crops, other formulations of 2,4-D can be used for other purposes like rangeland pasture, farmsteads, and burndown applications. Because of the readability of other formulations of 2,4-D, the potential use of other 2,4-D formulations, besides the choline formulation, for controlling broadleaf weeds in Enlist crops can become a possible problem. Although these herbicides are still 2,4-D formulations and will not damage HT crops, these other formulations are not labeled for usage in HT crops, thus making them an illegal application. These other formulations of 2,4-D are also more volatile and can potentially damage nearby susceptible crops. Unfortunately, the visual symptomology caused by these different 2,4-D formulations is almost identical at field level (J. Buol, 2019), meaning one could not visually determine which formulation of 2,4-D caused damage to a crop based on the visual symptomology alone.

Another method of combating HR weeds is using tank mixtures to minimize the over-reliance of just one MOA and rotating herbicide MOAs (Knezevic et al., 2009; D. E. Peterson, 1999). Tank mixing is mixing multiple herbicides, typically with different MOAs, in a tank and applying that mixture to control weeds. Tank mixtures of 2,4-D formulations and glyphosate are

very common and kill weeds more effectively than when using 2,4-D and glyphosate alone (Joseph et al., 2018; Merchant et al., 2014). Pre-mixtures (premix) of 2,4-D and glyphosate, meaning 2,4-D and glyphosate already mixed in a container instead of tank mixing, have also been introduced and can be more effective than tank mixtures (Ford et al., 2014; Palma-Bautista et al., 2021). 2,4-D and glyphosate premixes have also produced lower downwind depositions when compared to 2,4-D and glyphosate tank mixtures, depending on the nozzle used (Havens et al., 2018b). Although these procedures can mitigate OTM, they cannot ultimately eliminate OTM.

With the introduction of new biotechnologies and formulations, an increase in OTM and damage to susceptible crops are likely to occur. Although there is no way to visually distinguish between the various 2,4-D formulations in the field via symptomology, recent studies involving Fourier transform infrared spectroscopy (FTIR) and chemometric analysis have shown promise in identifying 2,4-D formulations in injured plant tissue samples. FTIR spectroscopy and principal component analysis-linear discriminant analysis (PCA-LDA) have been used to classify different formulations of 2,4-D and dicamba in injured cotton and soybean tissue leaves at a set rate mimicking OTM and collected at various days after treatment (Reid, 2017). FTIR and LDA alone have also been used to create classification models for 2,4-D and dicamba with various rates that mimic OTM and collected various days after treatment in dicamba-tolerant crops (J. Buol, 2019). In addition, other machine learning algorithms have been coupled with spectrometers to create pesticide classification models. For example, near-infrared spectroscopy and support vector machines have been used to create classification models for cypermethrin, Matrine, and avermectin residues on purple cabbage with near perfect accuracy (Li et al., 2021).

Taking all this information for consideration, the objectives for this thesis are as follows:

1. The creation of classification models capable of identifying various formulations of 2,4-D from HT cotton and HT soybeans using FTIR spectroscopy and machine learning algorithms.
2. To determine which machine learning algorithm is the best in constructing the classification models.
3. To establish if the timing between treatment and collection of the tissue samples and the treatment status of the leaf tissue will influence the accuracy of the classification models.

References

2,4-D Technical Fact Sheet. (n.d.). Retrieved April 9, 2020, from http://npic.orst.edu/factsheets/archive/2,4-Dtech.html

Ag and Food Statistics Charting the Essentials, February 2020. (2020).

Andersen, S. M., Clay, S. A., Wrage, L. J., & Matthees, D. (2004). Soybean Foliage Residues of Dicamba and 2,4-D and Correlation to Application Rates and Yield. *Agronomy Journal*, *96*(3), 750–760. https://doi.org/10.2134/AGRONJ2004.0750

Beckie, H. J. (2011). Herbicide-resistant weed management: focus on glyphosate. *Pest Management Science*, *67*(9), 1037–1048. https://doi.org/10.1002/PS.2195

Bish, M., Oseland, E., & Bradley, K. (2021). Target pesticide movement: a review of our current understanding of drift due to inversions and secondary movement Off-target pesticide movement: a review of our current understanding of drift due to inversions and secondary movement. *Weed Technol*. https://doi.org/10.1017/wet.2020.138

Buol, J. (2019). Stewarding 2,4-D- and dicamba- based weed control technologies in cotton and soybean production systems [Mississippi State University]. In *Theses and Dissertations*. https://scholarsjunction.msstate.edu/td/5049

Buol, J. T., Reynolds, D. B., Dodds, D. M., Mills, J. A., Nichols, R. L., Bond, J. A., Jenkins, J. N., & Dubien, J. L. (2019). The effect of cotton growth stage on response to a sublethal concentration of 2,4-D. *Weed Technology*, *33*(2), 321–328. https://doi.org/10.1017/WET.2019.9

Carlsen, S. C. K., Spliid, N. H., & Svensmark, B. (2006). Drift of 10 herbicides after tractor spray application. 2. Primary drift (droplet drift). *Chemosphere*, *64*(5), 778–786. https://doi.org/10.1016/J.CHEMOSPHERE.2005.10.060

Egan, J. F., Barlow, K. M., & Mortensen, D. A. (2014). A Meta-Analysis on the Effects of 2,4-D and Dicamba Drift on Soybean and Cotton. *Weed Science*, *62*(1), 193–206. https://doi.org/10.1614/WS-D-13-00025.1

Ford, L., Soltani, N., Robinson, D. E., Nurse, R. E., McFadden, A., & Sikkema, P. H. (2014). Evaluation of 2,4-D Amine, Glyphosate, 2,4-D Amine plus Glyphosate DMA and 2,4-D Choline/Glyphosate DMA for Their Efficacy on Glyphosate Susceptible and Resistant Canada Fleabane Populations. *Agricultural Sciences*, *05*(11), 1053–1060. https://doi.org/10.4236/AS.2014.511114

Havens, P. L., Hillger, D. E., Hewitt, A. J., Kruger, G. R., Marchi-Werle, L., & Czaczyk, Z. (2018). Field Measurements of Drift of Conventional and Drift Control Formulations of 2,4-D Plus Glyphosate. *Weed Technology*, *32*(5), 550–556. https://doi.org/10.1017/WET.2018.55

Heap, I. (2022, February 14). *The International Herbicide-Resistant Weed Database*. https://www.weedscience.org/Pages/SOASummary.aspx

Joseph, D. D., Marshall, M. W., Sanders, C. H., Joseph, D. D., Marshall, M. W., & Sanders, C. H. (2018). Efficacy of 2,4-D, Dicamba, Glufosinate and Glyphosate Combinations on Selected Broadleaf Weed Heights. *American Journal of Plant Sciences*, *9*(6), 1321–1333. https://doi.org/10.4236/AJPS.2018.96097

Knezevic, S. Z., Datta, A., Scott, J., Klein, R. N., & Golus, J. (2009). Problem Weed Control in Glyphosate-Resistant Soybean with Glyphosate Tank Mixes and Soil-Applied Herbicides. *Weed Technology*, *23*(4), 507–512. https://doi.org/10.1614/WT-09-012.1

Lenny Wells, M., Prostko, E. P., & Wendell Carter, O. (2019). Simulated Single Drift Events of 2,4-D and Dicamba on Pecan Trees. *HortTechnology*, *29*(3), 360–366. https://doi.org/10.21273/HORTTECH04265-19

Li, M., Lu, L., & Zhang, X. (2021). Qualitative Determination of Pesticide Residues in Purple Cabbage Based on Near Infrared Spectroscopy. *Journal of Physics: Conference Series*, *1884*(1), 12–15. https://doi.org/10.1088/1742-6596/1884/1/012015

Llewellyn, R., Ronning, D., Clarke, M., Mayfield, A., Walker, S., & Ouzman, J. (2016). *IMPACT OF WEEDS ON AUSTRALIAN GRAIN PRODUCTION: The cost of weeds to Australian grain growers and the adoption of weed management and tillage practices*. www.grdc.com.au/bookshop

Lyon, B. R., Cousins, Y. L., Llewellyn, D. J., & Dennis, E. S. (1993). Cotton plants transformed with a bacterial degradation gene are protected from accidental spray drift damage by the herbicide 2,4-dichlorophenoxyacetic acid. *Transgenic Research 1993 2:3*, *2*(3), 162–169. https://doi.org/10.1007/BF01972610

Marple, M. E., Al-Khatib, K., & Peterson, D. E. (2008a). Cotton Injury and Yield as Affected by Simulated Drift of 2,4-D and Dicamba. *Weed Technology*, *22*(4), 609–614. https://doi.org/10.1614/WT-07-095.1

Marple, M. E., Al-Khatib, K., & Peterson, D. E. (2008b). Cotton Injury and Yield as Affected by Simulated Drift of 2,4-D and Dicamba. *Weed Technology*, *22*(4), 609–614. https://doi.org/10.1614/WT-07-095.1

Merchant, R. M., Culpepper, A. S., Eure, P. M., Richburg, J. S., & Braxton, L. B. (2014). Controlling Glyphosate-Resistant Palmer Amaranth (Amaranthus palmeri) in Cotton with Resistance to Glyphosate, 2,4-D, and Glufosinate. *Weed Technology*, *28*(2), 291–297. https://doi.org/10.1614/WT-D-13-00104.1

Mohseni-Moghadam, M., Wolfe, S., Dami, I., & Doohan, D. (2016). Response of Wine Grape Cultivars to Simulated Drift Rates of 2,4-D, Dicamba, and Glyphosate, and 2,4-D or Dicamba Plus Glyphosate. *Weed Technology*, *30*(3), 807–814. https://doi.org/10.1614/WT-D-15-00106.1

Nufarm Weedar 64 Broadleaf Herbicide. (n.d.). Retrieved February 21, 2022, from http://www.cdms.net/ldat/ld08K005.pdf

Palma-Bautista, C., Cruz-Hipólito, H. E., Alcántara-de la Cruz, R., Vázquez-García, J. G., Yanniccari, M., & De Prado, R. (2021). Comparison of premix glyphosate and 2,4-D formulation and direct tank mixture for control of Conyza canadensis and Epilobium ciliatum. *Environmental Pollution*, *281*, 117013. https://doi.org/10.1016/J.ENVPOL.2021.117013

Perotti, V. E., Larran, A. S., Palmieri, V. E., Martinatto, A. K., & Permingeat, H. R. (2020). Herbicide resistant weeds: A call to integrate conventional agricultural practices, molecular biology knowledge and new technologies. *Plant Science*, *290*, 110255. https://doi.org/10.1016/J.PLANTSCI.2019.110255

Peterson, D. E. (1999). The Impact of Herbicide-Resistant Weeds on Kansas Agriculture. *Weed Technology*, *13*(3), 632–635. https://doi.org/10.1017/S0890037X00046315

Peterson, M. A., Mcmaster, S. A., Riechers, D. E., Skelton, J., & Stahlman, P. W. (2016). 2,4-D Past, Present, and Future: A Review. *Weed Technology*, *30*, 303–345. https://doi.org/10.1614/WT-D-15-00131.1

Recent Trends in GE Adoption. (2020). https://www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-us/recent-trends-in-ge-adoption/

Reid, C. (2017). Monitoring Aspergillus Flavus Progression and Aflatoxin Accumulation in Inoculated Maize (Zea Mays L.) Hybrids [Mississippi State University]. In *Theses and Dissertations*. https://scholarsjunction.msstate.edu/td/3195

Riar, D. S., Norsworthy, J. K., Steckel, L. E., Stephenson, D. O., Eubank, T. W., Bond, J., & Scott, R. C. (2013). Adoption of Best Management Practices for Herbicide-Resistant Weeds in Midsouthern United States Cotton, Rice, and Soybean. *Weed Technology*, *27*(4), 788–797. https://doi.org/10.1614/WT-D-13-00087.1

Robinson, A. P., Davis, V. M., Simpson, D. M., & Johnson, W. G. (2013). Response of Soybean Yield Components to 2,4-D. *Weed Science*, *61*(1), 68–76. https://doi.org/10.1614/WS-D-12-00077.1

Shyam, C., Chahal, P. S., Jhala, A. J., & Jugulam, M. (2021). Management of glyphosate-resistant Palmer amaranth (Amaranthus palmeri) in 2,4-D–, glufosinate-, and glyphosate-resistant soybean. *Weed Technology*, *35*(1), 136–143. https://doi.org/10.1017/WET.2020.91

Singh, M., & Sharma, S. D. (2000). EFFECT OF GLYPHOSATE AND ITS 2,4-D FORMULATION ON SOME DIFFICULT TO CONTROL WEEDS. *Proceedings of the Florida State Horticultural Society.*, 63–67.

Song, Y. (2014). Insight into the mode of action of 2,4-dichlorophenoxyacetic acid (2,4-D) as an herbicide. *Journal of Integrative Plant Biology*, *56*(2), 106–113. https://doi.org/10.1111/JIPB.12131

Sosnoskie, L. M., Culpepper, A. S., Braxton, L. B., & Richburg, J. S. (2015). Evaluating the Volatility of Three Formulations of 2,4-D When Applied in the Field. *Weed Technology*, *29*(2), 177–184. https://doi.org/10.1614/wt-d-14-00128.1

Tindall, K., Becker, J., Orlowski, J., Hawkins, C., & Kells, B. (2021). *Status of Over-the-Top Dicamba: Summary of 2021 Usage, Incidents and Consequences of Off-Target Movement, and Impacts of Stakeholder-Suggested Mitigations (DP# 464173: PC Code 128931)*. https://www.regulations.gov/document/EPA-HQ-OPP-2020-0492-0021

Tu, M., Hurd, C., Randall, J. M., & Rice, B. (2001). Weed Control Methods Handbook: Tools and Techniques for Use in Natural Areas. *The Nature Conservancy*, 87–96.

Unison. (n.d.). Retrieved February 21, 2022, from https://s3-us-west-1.amazonaws.com/agrian-cg-fs1-production/pdfs/Unison_MSDS1t.pdf

Vencill, W. K., Nichols, R. L., Webster, T. M., Soteres, J. K., Mallory-Smith, C., Burgos, N. R., Johnson, W. G., & Mcclelland, M. R. (2022). *Herbicide Resistance: Toward an Understanding of Resistance Development and the Impact of Herbicide-Resistant Crops*. https://doi.org/10.1614/WS-D-11-00206.1

Wright, T. R., Shan, G., Walsha, T. A., Lira, J. M., Cui, C., Song, P., Zhuang, M., ArnoldN, N. L., Lin, G., Yau, K., Russell, S. M., Cicchillo, R. M., Peterson, M. A., Simpson, D. M., Zhou, N., Ponsamuel, J., & Zhang, Z. (2010). Robust crop resistance to broadleaf and grass herbicides provided by aryloxyalkanoate dioxygenase transgenes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(47), 20240–20245. https://doi.org/10.1073/PNAS.1013154107/-/DCSUPPLEMENTAL

CHAPTER II

DESIGNING CLASSIFICATION MODELS FOR FORMULATIONS OF 2,4-D APPLIED TO ENLIST CROPS UTILIZING FOURIER TRANSFORM INFRARED SPECTROSCOPY (FTIR) AND MACHINE LEARNING ALGORITHMS

Introduction

Herbicide-resistant (HR) weeds have greatly influenced agriculture and are a significant problem for farmers and scientists (Perotti et al., 2020). Herbicide-resistant weeds are caused by mutations found in an individual weed that makes it resistant to a particular herbicide, and when it reproduces, its offspring will also be HR. Using the same mode of action (MOA) without pivoting to other MOAs or other weed management options can result in more HR weeds in a population (Vencill et al., 2022). Glyphosate-resistant weeds make up a large bulk of the HR weeds, with 55 glyphosate-resistant weeds out of 266 HR weed species (Heap, 2022). Herbicide-tolerant (HT) crops are new biotechnologies that can help mitigate HR. These HT cultivars are found in soybean (*Glycine max* L.), cotton (*Gossypium hirsutum* L.), and corn (*Zea mays* L.) crops. These cultivars have been genetically modified to resist multiple herbicides like glyphosate, 2,4-D, dicamba, or glufosinate. The Enlist® weed control system (‘Enlist crops’,Corteva Agriscience, Indianapolis, IN 46268) is a new biotechnology that instills 2,4-D and glyphosate tolerance in all cultivars, with the addition of glufosinate tolerance in the cotton and soybean cultivars. This 2,4-D tolerance is produced by employing transgenes encoding aryloxyalkanoate dioxygenase enzymes, Or AADs, that can metabolize 2,4-D into its non-active

metabolites (Wright et al., 2010). These HT biotechnologies will allow farmers and applicators to spray 2,4-D post-emergent (POST) to control weeds that are hard to kill or glyphosate-resistant and protect their crops from 2,4-D OTM (Lyon et al., 1993; Shyam et al., 2021; Singh & Sharma, 2000). Off-target movement is the movement of herbicides from the targeted crop to other off-target crops, including susceptible crops. With the introduction of these new 2,4-D tolerant crops, increased use of 2,4-D is expected, leading to 2,4-D OTM and damage to susceptible crops (Mortensen et al., 2012; Sharkey et al., 2021). To reduce the potential of OTM, different formulations of 2,4-D have been introduced. These 2,4-D formulations include 2,4-D choline, 2,4-D amine, and 2,4-D ester, which all have varying properties, like vapor pressure and solubility (Peterson et al., 2016). The choline formulation of 2,4-D is the only 2,4-D formulation labeled for POST application in Enlist crops and happens to be the least volatile formulation of 2,4-D (Sosnoskie et al., 2015a)**.** This 2,4-D choline formulation has also exhibited lower downwind and upwind deposition compared to other 2,4-D formulations and 2,4-D glyphosate tank mixtures (Havens et al., 2018; Mortensen et al., 2012). Although the less volatile choline formulation of 2,4-D is the only labeled 2,4-D formulation for use in Enlist crops, other formulations of 2,4-D are still labeled and used in other applications besides row crops, like rangeland pasture, farmsteads, and burndown applications. These formulations can still cause OTM injury during these applications and threaten non-tolerant crops, as these formulations are still more volatile than the choline formulation of 2,4-D. However, these formulations would not damage HT crops, introducing the possibility of applicators making illegal applications of off-labeled 2,4-D formulations as these formulations are also not labeled for use in Enlist crops, thus making their application illegal.

The OTM damages caused by these various formulations of 2,4-D to non-tolerant crops vary slightly if at all, thus making 2,4-D formulation identification via visual symptomology almost impossible (Buol, 2019). Sosnoskie et al. (2014) found an increased amount of injury and reduced cotton height when exposed to 2,4-D ester compared to 2,4-D amine or 2,4-D choline. In soybeans, visual symptomology was very similar regarding 2,4-D amine and 2,4-D ester with rates mimicking OTM (Thompson et al., 2007). Standard analytical techniques like high-performance liquid chromatography and gas chromatography/mass spectroscopy have been used to determine 2,4-D acid residue in water samples (Orooji et al., 2021) and determine multiple acetanilide herbicides from various cereal crops (Yaping Zhang et al., 2011). Although these analytical techniques effectively determine and quantify 2,4-D in samples, they are not effective at identifying or classifying specific formulations of the herbicide because the salt group (choline, ethylhexyl ester, dimethylamine salt, etc.) will be cleaved from the formulation during sample extraction making formulation identification impossible (Reid, 2017). Recent studies have shown the feasibility of using Fourier-Transform infrared spectroscopy (FTIR) as an alternative analytical tool to identify the specific formulation, as it requires little to no samples preparation.

FTIR spectroscopy is the measurement of the interaction of infrared radiation with matter. A broadband light source containing the full spectrum of wavelengths and a Michelson interferometer are used to collect spectral data from a wide range of wavelengths. The Michelson interferometer comprises two mirrors, one stationary and one moving, and a beamsplitter. Light from the source is directed at the beamsplitter, which then splits the light, directing 50% of it to the stationary mirror and the other 50% to the moving mirror. The light from both mirrors is then reflected back to the beamsplitter, focused on the sample, and refocused to the detector. An

interferogram is constructed by altering the optical path length of the two arms and recording the signal for the various path lengths. The interferogram plots intensity of the signal vs. the distance of the moving mirror. A Fourier transformation is applied to the interferogram to produce a spectrum plotting absorbance vs. wavenumber (cm^{-1}), which is reciprocal to wavelength. Many bonds and chemical structures have specific peaks and bands that can be compared to unknown samples to identify the chemical makeup of a sample. FTIR spectroscopy has been applied to many analyses, including waste management, measuring air pollution, and even soil analysis (Simonescu, 2012). FTIR spectroscopy can be coupled with supervised machine learning algorithms to create classification models to identify various chemicals or samples. Machine learning is the automated recognition of patterns in data or the process of programming a computer to learn from input data (Shai & Shai, 2014). The input data is considered the training dataset, a section of the data used to train machine learning models and determine the best parameters for each algorithm. Once the best parameter is found, the models will be tested using another data section, called the test dataset. Supervised machine learning is when the training dataset contains labels, or the values for the input and output (Shai & Shai, 2014). The test dataset is missing this label, allowing the particular algorithm to classify or identify the label based on the samples used to train the model, thus producing a machine learning model. Principal component analysis with linear discriminant analysis (PCA-LDA) classification models have been used with FTIR spectroscopy to classify various 2,4-D formulations from injured cotton tissue samples collected at 0, 3, 7, and 28 days after treatment (DAT) that were treated with 8 g 2,4-D ae ha^{-1} with 90% classification accuracy (Reid, 2017). These models used a small number of samples for their construction and only at a rate that mimicked a large amount of OTM, but they also showed the possibility of coupling FTIR spectroscopy and machine learning

to classify 2,4-D formulations. Buol (2019) continued this process by constructing classification models using PCA-LDA and LDA only and FTIR spectroscopy to classify various formulations of 2,4-D from injured cotton and soybean tissue samples collected 7, 14, 21, 28, and 56 DAT and treated with 33, 17, 8, 4, 2, 1 g 2,4-D ae ha^{-1} with 2,4-D classification accuracies of up to 80%. These models helped solidify the feasibility of using FTIR and machine learning to classify various formulations of 2,4-D by using hundreds of samples and various concentrations of 2,4-D. However, Reid (2017) and Buol (2019) only employed rates that mimic OTM and collected tissue samples from non-tolerant soybeans and cotton without examining applied rates in HT crops.

Objective

To construct classification models for the identification of 2,4-D formulations present in HT soybean and cotton tissue samples, experiments were conducted utilizing multiple machine learning algorithms and FTIR spectra from tissue samples that were directly treated with an applied rate of 2,4-D and tissue samples that grew post-treatment of 2,4-D collected at various days after treatment.

Methods and Materials

Experimental Design, Treatments, and Data Sampling

Experiments were conducted in 2020 at the R.R. Foil Plant Science Research center in Starkville, Mississippi, to construct classification models to identify 2,4-D formulations found in HT cotton and soybean tissue. The soybean variety 481E19 was seeded at 130,000 live plants per acre, and the cotton variety PHY 490 W3FE was seeded at 45,000 live plants per acre. Treatments were arraigned in a randomized complete block design with six treatments and four

replications, including a non-treated check. Experimental units consisted of plots containing two 76-cm by 12.2-m rows for cotton and four 76-cm by 12.2-m rows for soybean. Samples were only collected from the middle two rows. Sampling units consisted of individual leaf tissue samples collected from the plot. Experimental factors consisted of 2,4-D formulations, days between treatment and collection, and individual leaf tissue treatment status. All treatments consisted of a tank mix of glyphosate (Roundup Powermax II™ , Bayer CropSciences) and 2,4-D formulations except for the non-treated check, which only contained glyphosate, and the treatment involving Enlist Duo (Enlist Duo™ herbicide with Colex-D® technology, Corteva Agriscience). 2,4-D formulations used included 2,4-D Acid ('Unison', Unison® Novel Broadleaf, Helena Chemical Company, Collierville, TN 38017), 2,4-D Dimethylamine salt ('Weedar', Weedar® 64, Nufarm Agricultural Products, Alsip, IL 60803), 2,4-D Ethylhexyl Ester ('Weedone', Weedone® LV4 EC, Nufarm Agricultural Products, Alsip, IL 60803), 2,4-D Choline salt ('Enlist 1', Enlist One™ with Colex-D® technology, Corteva Agriscience), and 2,4-D Choline salt/glyphosate premixture ('Enlist 2', Enlist Duo™ herbicide with Colex-D® technology, Corteva Agriscience). The chemical structure of 2,4-D and the various 2,4-D formulations used in this experiment is shown in figure 2.1. Before the application, wooden stakes were inserted into the soil at 2.44-m sections of each plot to mark each temporal section period, 0, 7, 14, 21, and 28 days after treatment (DAT). Also, prior to the application, electric tape was tied to the highest exposed leaf in ten individual cotton plants in each temporal section for each plot. This was also repeated for soybeans, except at the highest exposed. This distinguished tissue samples that were directly treated with the 2,4-D formulations and those that grew post-treatment and were not directly treated with the 2,4-D formulations. Soybeans and cotton grow upwards or up the stem, so leaf treatment status can be determined by tying a piece

of electric tape prior to treatment application (Albers, 1993; Purcell et al., 2014). A broadcast application of glyphosate at 0.87 kg ae ha^{-1} was applied to the experimental area to control emerging weeds. All active ingredients were applied at a rate of 1.06 kg ae ha^{-1} except for Enlist 2, which was applied at 2.19 kg ae ha^{-1}. All herbicides were applied via a compressed air tractor-mounted research sprayer calibrated at 140 L ha^{-1}. The sprayer boom was equipped with TTI11005 (TeeJet Technologies, Glendale Heights, Il 60139) spray tips which were used to apply the treatments to the middle two rows of cotton and soybean plots. Treatment application occurred at the V3 growth stage in soybeans (McWilliams et al., 1999) and when cotton achieved three true leaves. During application, spray equipment was washed with ammonia (WipeOut® XS, Helena Chemical Company, Collierville, TN 38017) and water between each treatment to prevent any cross-contamination. No OTM damage from 2,4-D was observed between the plots at the location. Leaf tissue sample collection occurred at 0, 7, 14, 21, and 28 DAT in each temporal section period, depending on the sampling date, and samples were collected from the middle two rows. Tissue samples were hand-harvested using latex gloves which were changed between each plot, and rubber boots to minimize cross-contamination. Between 10 and 20 leaf tissue samples were collected from various cotton and soybean plants from the various plots per collection. The leaf samples were stored in 3.78 L plastic freezer bags (Ziploc®, SC Johnson & Son, Inc., Racine, WI 53403) depending on the treatment status of the leaves collected. A third of the bags were labeled as “treated,” meaning only leaf tissue samples that were directly treated with the 2,4-D formulations and located below the tied electric tape were placed in these bags. Another third of all bags were labeled as “new growth,” and only tissue samples that grew post-2,4-D treatment and were located above the tied electric tape were placed in these bags. The last third of all bags were labeled as “composite” and comprised a

combination of tissue samples that were directly treated with 2,4-D and tissue samples that grew post-treatment of 2,4-D. Following the collection, all sample bags were placed into ice-filled coolers, transported to the Mississippi State Chemical Laboratory, and stored in a -80°C Thermo Scientific (Thermo Fisher Scientific, Waltham, MA 02451) TSC2090D chest freezer. After an initial freeze lasting 2 to 3 days, the samples were moved into a -20°C walk-in freezer for long-term storage.

Sample Processing, Data Collection, and Data Preprocessing

Sample processing and data collection followed previous and similar studies (Buol, 2019; Reid, 2017). Samples were thawed and processed by grinding the leaf tissue in a mortar and pestle. The ground tissue samples were returned to the original sample bag and placed in the -20°C walk-in freezer. Latex gloves were worn and changed between each sample, and the mortar and pestle were rinsed with water and cleaned with a solution of 70% ethanol and 30% water. After processing the samples, data collection began by thawing out the samples and analyzing them using a Thermo Scientific (Thermo Fisher Scientific) Nicolet 6700 FTIR optical spectrometer equipped with a liquid nitrogen-cooled MCT High-D detector, KBr beamsplitter, and Smart ARK accessory (Figure 2.2). Roughly 1 g of leaf tissue from each sample bag was inserted onto a ZnSe trough plate with an incidence angle of 60°. Ten reflections of infrared light were passed through the plate in each scan, and an infrared spectrum was produced from 32 scans of each tissue sample. The 1 g of tissue sample was returned to the sample bag, which was then mixed up, and this process was repeated three more times, producing four spectra generated from each sample bag. After a sample bag was used to produce four spectra, the ZnSe plate, smart ARK accessory, and scoopula used to place the tissue onto the plate were thoroughly cleaned with the ethanol/water solution. Latex gloves were also worn and changed after each

sample. A background spectrum with no sample placed onto a clean ZnSe plate was collected every hour. All spectra were collected at 4000 to 650 cm-1 frequencies and compared to other spectra to ensure proper spectra were collected using Omnic 7.3 (Thermo Fisher Scientific). Automatic baseline correction was performed on each spectrum to correct the tilting or slanting of the spectra in Omnic 7.3. Following the automatic baseline correction in Omnic 7.3, all spectra were exported to The Unscrambler X 10.5 (Camo Analytics, Magnolia, TX 77354) software to visualize all the spectra and convert the spectra files into Microsoft Excel (Microsoft Corporation, Redmond, WA 98052) files. Once all the spectra files were converted into excel files, they were exported into RStudio (RStudio Team, Boston, MA) for preprocessing and analysis. RStudio was used instead of Unscrambler X 10.5 because of Unscrambler X's limitations and the lack of machine learning implementation compared to RStudio. All raw spectra (Figure 2.3 and figure 2.4) were scaled to a mean of 0 and a standard deviation of 1 after importation into RStudio. The spectra were then smoothed and derived using the Savitzky-Golay algorithm via the prospectr package (Stevens & Ramirez-Lopez, 2011). The Savitzky-Golay filter is used to filter out noise by applying a polynomial to a fixed number of data points, then moving the polynomial, thus dropping a data point on one end and adding another data point to the other end, until it applies the polynomial to all data points (Savitzky & Golay, 1964). All spectra were first smoothed using a first-order polynomial and a window size of 11, then derived using a first-order polynomial with a first derivative and a window size of 3 (Figure 2.5-2.8). Following the Savitzky-Golay transformation, feature selection was implemented using the Boruta package (Kursa & Rudnicki, 2010). Feature selection is used to remove irrelevant and unimportant data that can increase run time, reduce accuracy, and interfere with the completion of the model (Cai et al., 2018). Boruta functions by creating duplicates of all the features, called

shadow features, and shuffles them within columns to remove the association between the shadow features and the response variables. Next, a random forest classifier is applied to the data with the shadow features included. It will then calculate the importance of each feature by using the mean decrease accuracy. The algorithm will then compare the *z*-scores of the real features versus the maximum *z*-score of the shadow features. If the *z*-scores of the real features are higher than the shadow features, Boruta classifies this as a "hit." It repeats the process until the number of hits for a feature becomes unlikely to be caused by chance and then assigns the feature as important or unimportant. The maximum number of importance source runs was set at 200, and the number of trees for the random forest was set at 1000. All of the features labeled as important were analyzed using different machine learning algorithms, while the features labeled as not important were excluded.

Data Analysis Using Machine Learning Algorithms

Data analysis was conducted using the mlr3 package (Lang et al., 2019). Nested cross-validation with five outer loops with three repeats and five inner loops with three repeats were utilized to set the parameters, train, and test all the models. It then went through thirty tuning iterations, producing six thousand five hundred fifty models per machine learning algorithm. Machine learning algorithms used to construct classification models include Naive Bayes (NBayes), principal component analysis-linear discriminant analysis (PCA-LDA), XGBoost (XGB), ranger- random forest (Ranger), support vector machines (SVM), and k-nearest neighbors (KNN).

Packages utilized for analysis include readxl (Wickham & Bryan, 2019), ramify (Greenwell, 2016), prospectr (Stevens & Ramirez-Lopez, 2021), knitr (Xie, 2021), Boruta (Kursa & Rudnicki, 2010), corrplot (Wei & Simko, 2021), summaryplots (Comtois, 2021),

futures (Bengtsson, 2021), devtools (Wickham et al., 2021), mlr3 (Lang et al., 2019), mlr3viz (Lang et al., 2021), mlr3learners (Lang et al., 2021), mlr3pipelines (Pfisterer et al., 2021), mlr3tuning (Becker et al., 2021), ggplot2 (Wickham, 2016), mlr3benchmark (Raphael, 2021), mlr3extralearners (Sonabend & Schratz, 2021), mlr3filters (Schratz et al., 2021), tidyverse (Wickham et al., 2019), multcomp (Hothorn et al., 2008), data.table (Dowland Srinivasan, 2021), flextable (Gohel, 2021), captioner (Alathea, 2015), caret (Kuhn, 2021), e1071 (Meyer et al., 2021), pROC (Robin et al., 2011), signal (signal developers, 2013), pracma (Borchers, 2021), stats (R Core Team, 2020) and emmeans (Length, 2022).

Naive Bayes

The Naive Bayes classifier is based on the Bayes theorem and functions by assuming the presence of a feature in a class is unrelated to the presence of other features. Naive Bayes mainly looks at the probability and likelihood for classification. Naive Bayes functions by finding the probability of X occurring if Y has also occurred. Some parameters for Naive Bayes include the value of the Laplace smoothing and the probability threshold.

Linear Discriminant Analysis

Linear discriminant analysis is used to find a linear combination of features that classifies multiple classes or events. It is very similar to linear regression but used for classification instead of predictive analysis. Linear discriminant analysis is often paired with principal component analysis, which is an unsupervised learning technique used to reduce the dimensions of a dataset. The number of principal components used were based on cross-validation rather than variance explained. PCA-LDA is also mainly used for dimension reduction, but previous experiments using PCA-LDA to correctly classify 2,4-D and dicamba formulations at a concentration

mimicking off-target drift from tissue harvested multiple weeks after 2,4-D application produced classification models with accuracies of up to 90% (Reid, 2017)

XGBoost

XGBoost and random forest are similar algorithms that are both decision-tree-based but with some key differences. XGBoost is an ensemble technique that uses the gradient descent algorithm to essentially correct the mistakes made by previous trees generated by the model, learn from them, and improve the performance of the next step. This allows new models to be produced that predict the errors of the previous models and the residuals of the previous models and adds them together to make the final prediction. Parameters for XGBoost include the number of iterations, controlling the learning rate of the model, and the maximum depth of trees.

Ranger

Ranger is a form of the random forest learning method, a supervised decision tree-based classification model in which the trees are constructed independently of each other. Decision trees are a series of steps that answer questions and provide costs, probabilities, or other outcomes based on the question. One key difference between a random forest and a decision tree is that the process of finding the root node and the feature splitting node will occur randomly and is not based on rules in random forest models. A random subset of features is used at each split in the tree and not the entire dataset. Each tree produces a prediction in a random forest model, and the class with the most votes becomes part of the prediction model. Ranger parameters include Mtry or the number of variables randomly sampled for each tree, the number of possible trees, and the minimum size of the terminal node or node size.

Support-Vector Machines

The support-vector machine algorithm will graph a training dataset of two different classes and find a hyperplane separating them. SVM will use the points closest to the hyperplane, support vectors, to compute the largest distance between the hyperplane and the support vectors, with this maximum distance line being labeled the optimal hyperplane. A test set is then graphed into the same space as the training set. A prediction is made based on which side of the hyperplane the test points fall into. If the dataset is not linearly separable, dimensions can be added to help fit the hyperplane. By using a one-to-rest approach, SVM can be used in multi-class classification. The hyperplane is constructed for each class by maximizing the separation between one class and all other classes at once. The cost and gamma parameters are critical parameters in SVMs. The cost parameter is used to control the tradeoff between variance and bias, where the model can be trained to purposefully misclassify some of the training data to help the model classify the test data. A high cost will allow the hyperplane to classify all training points correctly, overfitting the model. A lower cost will misclassify some training set points to help better classify the test set. Gamma describes the weight of a single training point. A high gamma will add more weight to the points closest to the line and ignore points further away, while a low gamma will give more weight to points further away, causing the hyperplane to be more linear.

K-Nearest Neighbors

K-nearest neighbors is a classification method very similar to SVM. KNN marks a training data set as belonging to one of two categories. A test set is then introduced and mapped into the same place as the training set. Classification will be based on the data points nearest to the test dataset, called neighbors. One of the critical differences between SVM and KNN is using

a hyperplane to separate the data in SVM, while KNN does not use a hyperplane. A critical parameter for KNN is k, the number of neighbors considered. If k is set to one, it will only use the closest neighbor to classify the data point. If k is set to three, it will then use the three closest neighbors to classify the data points.

Comparing Machine Learning Models

All machine learning models will produce a classification matrix containing the predicted formulation of 2,4-D versus the actual reference 2,4-D formulation, the sensitivity of all formulations, and the positive predictive for all formulations. Following the construction of all the classification models, the models were fit into a binomial generalized linear model using the stats package (R Core Team, 2020), and mean separated and compared using Tukey's Honest Significance Difference (HSD) test from the emmeans package (Length, 2022) with a significance level of 0.05 to determine which model was the most accurate. Tukey HSD is used to determine the honest significant difference between two means using a studentized range and is a conservative test compared to other comparison methods (Abdi & Williams, 2010). For example, Tukey's HSD has been used to compare multiple machine learning model's abilities to estimate the dry biomass weight of wood chip residues (Fuente et al., 2021) and in comparing machine learning models performances in assessing triggering factors for debris flow (Yonghong Zhang et al., 2019).

Results and Discussion

Cotton

Two thousand four hundred forty-six cotton spectra were used in data analysis. Prior to feature selection, absorbance values for 1728 wavenumbers were collected, with 434

wavenumbers being labeled as important and used in machine learning analysis. Reid (2017) and Buol (2019) only used wavenumbers found in the "fingerprint region" between 1800 and 800 cm^{-1}. However, wavenumbers found outside the "fingerprint region" were labeled as important following feature selection and included in the analysis. The wavelengths labeled as most important from Boruta included wavelengths at around 1620 cm^{-1}, 1695 cm^{-1}, 1320 cm^{-1}, and 1050 cm^{-1}. The peaks at 1620 cm^{-1} and 1695 cm^{-1} are most likely produced from the aromatic rings found in auxin herbicides and primary and secondary amines, which have been identified as important in previous models (Buol, 2019; Reid, 2017). The peak at 1320 cm^{-1} most likely represents a carboxylic acid group and was also labeled as important in previous models (Buol, 2019; Reid, 2017). Lastly, the peak at 1050 cm^{-1} is most likely produced from the alcohol group found in 2,4-D. The confusion matrices from classification models generated using cotton spectra and various machine learning techniques can be found in tables 2.1 - 2.6. The classification model using NBayes produced an overall accuracy of 19% (Table 2.1). The PCA-LDA classification model produced an overall accuracy of 37% (Table 2.2). The classification model using XGBoost resulted in an overall accuracy of 56% (Table 2.3), while the Ranger classification model's overall accuracy was slightly higher than XGBoost, with an overall accuracy of 65% (Table 2.4). The SVM classification model resulted in an overall accuracy of 73% (Table 2.5). Lastly, the classification model using KNN resulted in the highest overall accuracy, with an overall accuracy of 81% (Table 2.6). The classification model using KNN produced the highest overall accuracy, and the classification model using Nbayes produced the worst accuracy, while the classification model using PCA-LDA resulted in the second least accurate model (Figure 2.9). Reid (2017) and Buol (2019) used another variety of soybean and cotton while only selecting tissue samples that exhibited damage, tissue most likely directly

applied with 2,4-D, which could explain the differences in accuracy comparing their LDA classification and the LDA classification model presented here**.** The KNN classification model best classified the control (85%) and performed inadequately when classifying Enlist 1 (78%), but all formulation sensitivities were above 70%.

Soybean

Two thousand four hundred seventy-five soybean spectra were used in data analysis. Prior to feature selection, absorbance values for 1728 wavenumbers were collected, with 331 wavenumbers being labeled as important and used in machine learning analysis. Reid (2017) and Buol (2019) only used wavenumbers found in the "fingerprint region" between 1800 and 800 cm^{-1}. However, wavenumbers found outside the "fingerprint region" were labeled as important following feature selection and included in the analysis. The wavelengths labeled as most important from Boruta included wavelengths at 1550 cm^{-1}, 1500 cm^{-1}, and 1670 cm^{-1}. These peaks are most likely produced from the aromatic rings found in auxin herbicides and primary and secondary amines, which have also been identified as important in previous models (Buol, 2019; Reid, 2017). The confusion matrices from classification models generated using cotton spectra and various machine learning techniques can be found in tables 2.7-2.12. The classification model using NBayes produced an overall accuracy of 18% (Table 2.7). The classification model using PCA-LDA also performed poorly, producing an overall accuracy of 32% (Table 2.8). The classification model using XGBoost resulted in an overall accuracy of 49% (Table 2.9), while the Ranger classification model's overall accuracy was slightly higher than XGBoost, with an overall accuracy of 54% (Table 2.10). The SVM classification model resulted in an overall accuracy of 60% (Table 2.11). Lastly, the classification model using KNN resulted in the highest overall accuracy, with an overall accuracy of 68% (Table 2.12). The

classification model using KNN produced the highest overall accuracy, and the classification model using Nbayes performed the worst, while the classification model using PCA-LDA resulted in the second least accurate model (Figure 2.10). Reid (2017) and Buol (2019) used another variety of soybean and cotton while only selecting tissue samples that exhibited damage, tissue most likely directly applied with 2,4-D, which could explain the differences in accuracy comparing their LDA classification and the LDA classification model presented here. The KNN classification model performed best in classifying Unison (71%) and worst when classifying Weedar (66%) and the control (67%), but all formulation sensitivities were above 60%.

Conclusions and Regulatory Implications

With the introduction and implementation of the Enlist weed control system, an increase in 2,4-D applications will likely occur; for example, the usage of 2,4-D has steadily increased since 2012 (*Estimated Annual Agricultural Pesticide Use*, 2021). With this increase in 2,4-D usage, damage to susceptible crops from OTM will almost certainly increase. This increase in OTM damage will likely be caused by an off-labeled application of a more volatile yet cheaper formulation of 2,4-D instead of the Enlist formulations. The choline formulation of 2,4-D is significantly less volatile than other 2,4-D formulations and produces little to no damage to susceptible row crops upwind or downwind of a treated plot (Kalsing et al., 2018; Sosnoskie et al., 2015b; Werle et al., 2021), but is typically more expensive than other formulations of 2,4-D. This difference in cost can lead to applicators spraying other more volatile, less stable, and illegal formulations of 2,4-D onto their crops to save money, increasing the probability of an OTM drift event occurring. The lack of proper analytical tools to correctly identify formulations of 2,4-D rapidly and at a low cost has hampered the ability to manage these OTM drift events. These classification models are meant to fix these problems and better manage these events.

These classification models were constructed only using one year's worth of data at one location, so more locations and years are most likely needed to observe if location and year influence accuracy. Next, the difference in classification accuracies of 2,4-D in the soybean and cotton models could be caused by physiological differences in the different crops, but more research involving 2,4-D and HT crops is needed. Also, for the construction of the models, all the leaf tissue samples were used to train and test the models, but none were used to validate the model. Previous models built using a single concentration of 2,4-D on susceptible cotton produced high accuracies with unknown samples (Reid, 2017), but these models were built with susceptible cotton tissue with a rate mimicking a large OTM drift event. Other models built using multiple concentrations of 2,4-D mimicking drift events and multiple timing evaluations produced lower accuracies with unknown samples (Buol, 2019). However, the models were built using another type of resistant row crop and only using the wavelengths constituting the fingerprint regions, 1800 to 800 cm^{-1}. In the following chapter, models will be built evaluating the effect of evaluation timing and treatment status of tissue on the accuracy of the classification models presented here.

This study shows how, with reasonable accuracy, using FTIR spectroscopy and machine learning models to classify various 2,4-D formulations found in 2,4-D tolerant soybeans and cotton leaf tissue samples is possible. Out of all machine learning algorithms examined, k-nearest neighbor produced the highest overall accuracy, with 80% formulation classification accuracy in cotton tissue treated with 2,4-D and 67% formulation classification accuracy in soybean tissue treated with 2,4-D. While 80% accuracy for a cotton classification model and 67% accuracy for a soybean classification model does not seem exceptional, a no-information model or if the model was guessing the formulation of 2,4-D found in the samples produced an

accuracy of 17%. This displays the possibility of using FTIR spectroscopy and machine learning models to correctly identify various formulations of 2,4-D from HT cotton and soybean tissue samples.

Finally, these models are also meant to be used alongside other investigative tools like sales records, application records, and other models used to mimic OTM drift damage. For example, during an investigative examining an OTM drift event, leaf tissue samples from that drift event will be collected and analyzed using previously constructed models that mimic drift damage to determine which formulation of 2,4-D, if any, caused that damage. Following that analysis, leaf tissue samples from neighboring farms and locations would be analyzed using the models presented here to determine which formulation of 2,4-D farmers and applicators used near the drift event and if any of those formulations found using this model match the formulation that caused the drift damage. Additionally, in many states, application records are required or at least recommended when applying 2,4-D or any other auxin herbicides, and those application records will also be used as an investigative tool. Lastly, many states also require sellers to keep sales receipts of regulated herbicides, including 2,4-D, for multiple years. Using the models constructed here, previously constructed models mimicking OTM drift, sales receipts, and application records, OTM drift events involving 2,4-D can be better monitored and investigated.

Tables

Table 2.1 Confusion matrix generated from the classification model using Naïve Bayes and transformed cotton spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	99	107	89	100	109	110	16%
Enlist 2	223	289	248	237	278	268	19%
No 2,4-D	58	53	80	72	82	76	19%
Unison	421	369	423	452	333	317	20%
Weedar	95	134	134	129	156	129	20%
Weedone	280	296	250	243	254	348	21%
Sensitivity	8%	23%	7%	37%	13%	28%	19%[†]

Hyperparameters: Naïve Bayes Threshold= 0.09928702; Epsilon= 0.08752653
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.2 Confusion matrix generated from the classification model using PCA-LDA and transformed cotton spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite) [a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	463	208	108	196	142	154	36%
Enlist 2	177	428	133	141	158	175	35%
No 2,4-D	100	97	458	162	167	128	41%
Unison	166	184	187	438	131	189	34%
Weedar	131	172	199	117	448	115	38%
Weedone	139	159	139	179	166	487	38%
Sensitivity	39%	34%	37%	36%	37%	39%	37%[†]

Hyperparameters: ldapca.rank= 196.7904
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.3 Confusion matrix generated from the classification model using XGBoost and transformed cotton spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite) [a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	640	131	83	150	72	70	56%
Enlist 2	127	695	71	103	121	154	55%
No 2,4-D	99	76	727	96	131	104	59%
Unison	149	66	130	699	119	74	57%
Weedar	62	112	109	105	650	116	56%
Weedone	99	168	104	80	119	730	56%
Sensitivity	54%	56%	59%	57%	54%	58%	56%[†]

Hyperparameters: Alpha= 1.386928; Eta=0.1784706; colsample_bytree= 0.196827; Lambda= 3.823641; nrounds= 573.525; subsample= 0.8062951; max_depth=49.72615; colsample_bylevel= 0.5393224

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value

[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter

[†]Overall accuracy of the classification model

Table 2.4 Confusion matrix generated from the classification model using Ranger and transformed cotton spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite) [a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	704	88	48	114	61	73	65%
Enlist 2	122	826	52	89	103	146	62%
No 2,4-D	79	37	852	79	97	82	69%
Unison	129	45	96	790	90	49	66%
Weedar	56	108	71	65	761	88	66%
Weedone	86	144	105	96	100	810	60%
Sensitivity	60%	66%	70%	64%	63%	65%	65%[†]

Hyperparameters: Mtry=5.825705
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.5 Confusion matrix generated from the classification model using SVM and transformed cotton spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite) [a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	803	70	33	87	43	48	74%
Enlist 2	87	898	38	67	67	99	71%
No 2,4-D	82	86	979	104	91	82	69%
Unison	93	32	70	854	54	53	74%
Weedar	51	67	53	61	899	73	75%
Weedone	60	95	51	60	58	893	73%
Sensitivity	68%	72%	80%	69%	74%	72%	73%[†]

Hyperparameters: filt2.filter.nfeat= 379.1929; SVM cost= 5.758028; SVM gamma= -1.850011
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.6 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite) [a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	919	50	37	70	45	70	77%
Enlist 2	55	1006	32	43	45	66	81%
No 2,4-D	37	34	1038	47	39	18	86%
Unison	77	47	55	1000	70	38	78%
Weedar	45	39	30	38	977	47	83%
Weedone	43	72	32	35	36	1009	82%
Sensitivity	78%	81%	85%	81%	81%	81%	81%[†]

Hyperparameters: filt2.filter.nfeat= 372.861; k= 2.544557
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.7 Confusion matrix generated from the classification model using Naïve Bayes and transformed soybean spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	473	430	463	471	425	389	18%
Enlist 2	187	160	163	179	152	150	16%
No 2,4-D	64	88	80	92	75	83	17%
Unison	69	69	63	81	77	72	19%
Weedar	114	131	114	106	154	126	21%
Weedone	341	373	359	319	323	413	19%
Sensitivity	38%	13%	6%	6%	13%	33%	18%[†]

Hyperparameters: Naïve Bayes Threshold= 0.123539; Epsilon= 0.3138411
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.8 Confusion matrix generated from the classification model using PCA-LDA and transformed soybean spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	443	164	171	137	186	193	34%
Enlist 2	196	377	194	163	152	222	29%
No 2,4-D	157	145	348	183	135	167	31%
Unison	177	220	205	496	144	183	35%
Weedar	123	149	147	121	418	146	38%
Weedone	152	196	177	148	171	322	28%
Sensitivity	35%	30%	28%	40%	35%	26%	32%[†]

Hyperparameters: ldapca.rank= 232.2527
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.9 Confusion matrix generated from the classification model using XGBoost and transformed soybean spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	621	152	144	106	121	113	49%
Enlist 2	125	561	137	134	122	106	47%
No 2,4-D	129	110	525	138	100	117	47%
Unison	123	167	146	622	86	83	51%
Weedar	132	141	134	139	674	182	48%
Weedone	118	120	156	109	103	632	51%
Sensitivity	50%	45%	42%	50%	56%	51%	49%[†]

Hyperparameters: Alpha= 1.47321; Eta= 0.05576059; colsample_bytree= 0.1659995; Lambda= 8.531977; nrounds= 989.3554; subsample= 0.960932; max_depth= 25.18549; colsample_bylevel= 0.0386287

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value

[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter

[†]Overall accuracy of the classification model

Table 2.10 Confusion matrix generated from the classification model using Ranger and transformed soybean spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	726	179	161	132	122	110	51%
Enlist 2	95	593	115	105	92	90	54%
No 2,4-D	108	103	606	123	71	95	55%
Unison	81	132	95	634	71	66	59%
Weedar	140	140	120	137	756	179	51%
Weedone	98	104	145	117	94	693	55%
Sensitivity	58%	47%	49%	51%	63%	56%	54%[†]

Hyperparameters: Mtry= 2
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.11 Confusion matrix generated from the classification model using SVM and transformed soybean spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	745	128	95	91	74	71	62%
Enlist 2	117	740	98	83	92	96	60%
No 2,4-D	117	106	696	114	86	120	56%
Unison	80	107	121	771	104	72	61%
Weedar	123	108	104	106	783	133	58%
Weedone	66	62	128	83	67	741	65%
Sensitivity	60%	59%	56%	62%	65%	60%	60%[†]

Hyperparameters: filt2.filter.nfeat= 315.9735; SVM cost= 6.373757; SVM gamma= -1.922477
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 2.12 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	859	109	65	90	88	67	67%
Enlist 2	94	846	82	64	70	61	70%
No 2,4-D	93	107	829	76	74	118	64%
Unison	76	71	101	880	97	61	68%
Weedar	59	60	65	62	799	78	71%
Weedone	67	58	100	76	78	848	69%
Sensitivity	69%	68%	67%	71%	66%	69%	68%[†]

Hyperparameters: filt2.filter.nfeat= 324.4514; k= 4.020513
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Figures

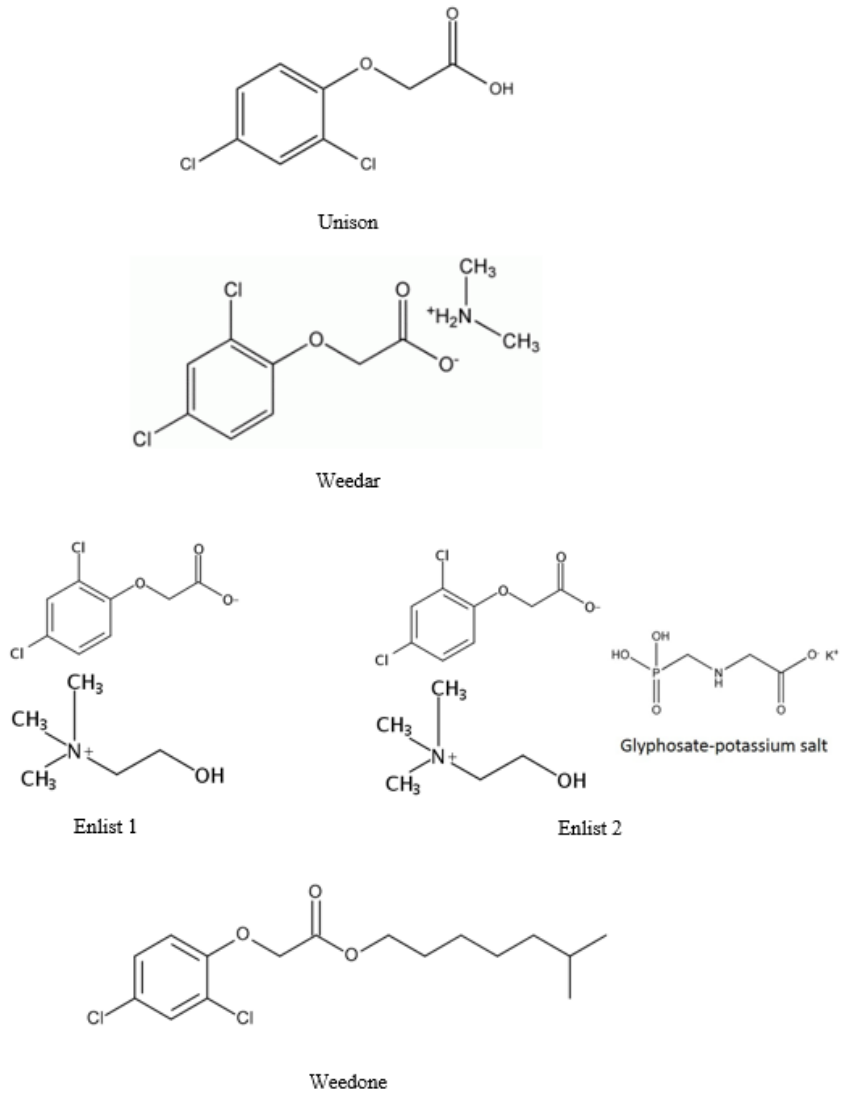

Figure 2.1 Chemical structures of 2,4-D and formulated salts (Unison, Weedar, Weedone, Enlist 1, and Enlist 2).[a]

[a]Abbreviations: Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture

Figure 2.2 Nicolet 6700 FTIR optical spectrometer equipped with a liquid nitrogen-cooled MCT High-D detector, KBr beamsplitter, and Smart ARK accessory

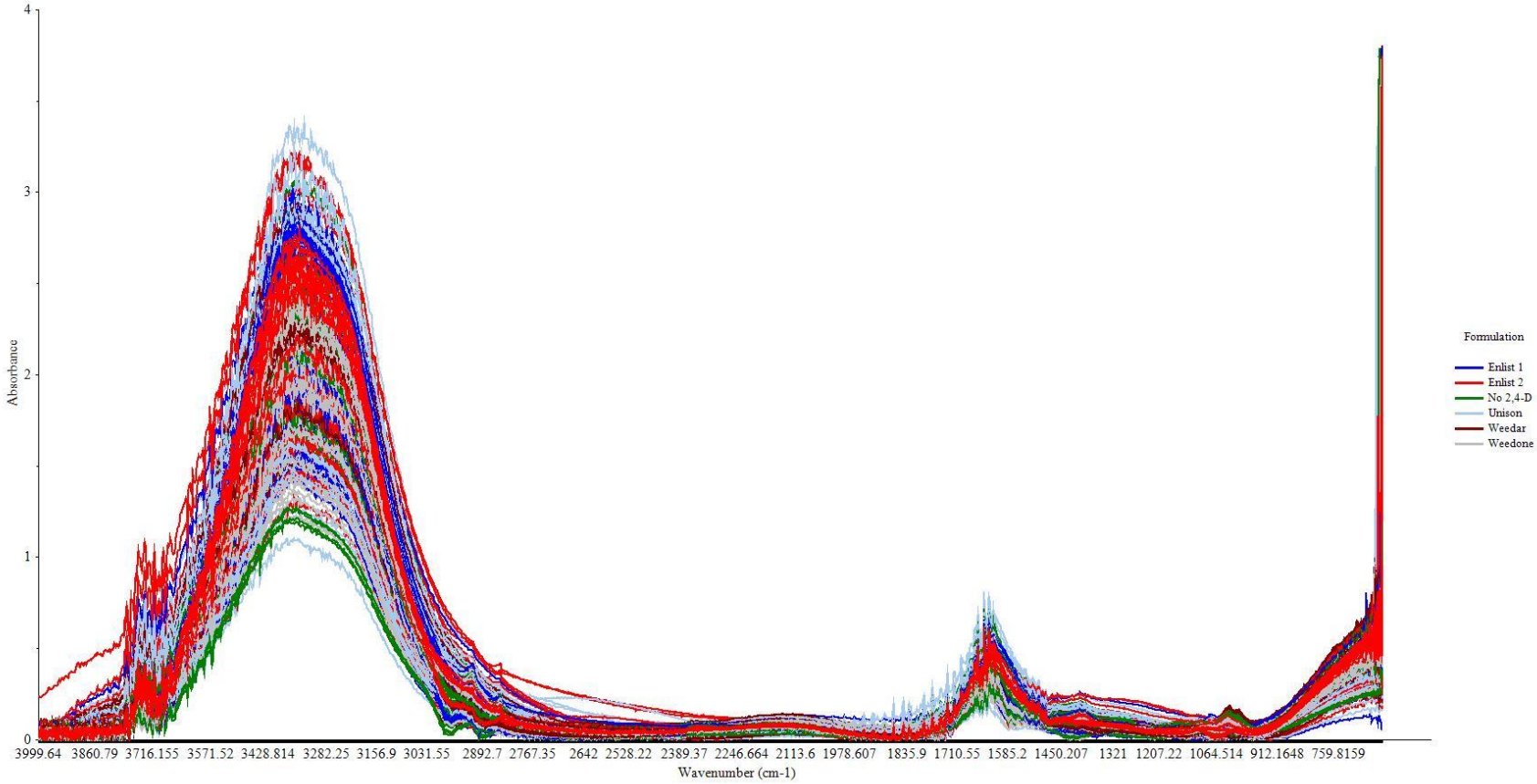

Figure 2.3 Raw spectra (4000 to 650 cm^{-1}) from cotton tissue treated with no 2,4-D, Unison, Weedar, Weedone, Enlist 1 or Enlist 2, pooled over sample timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a]

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture

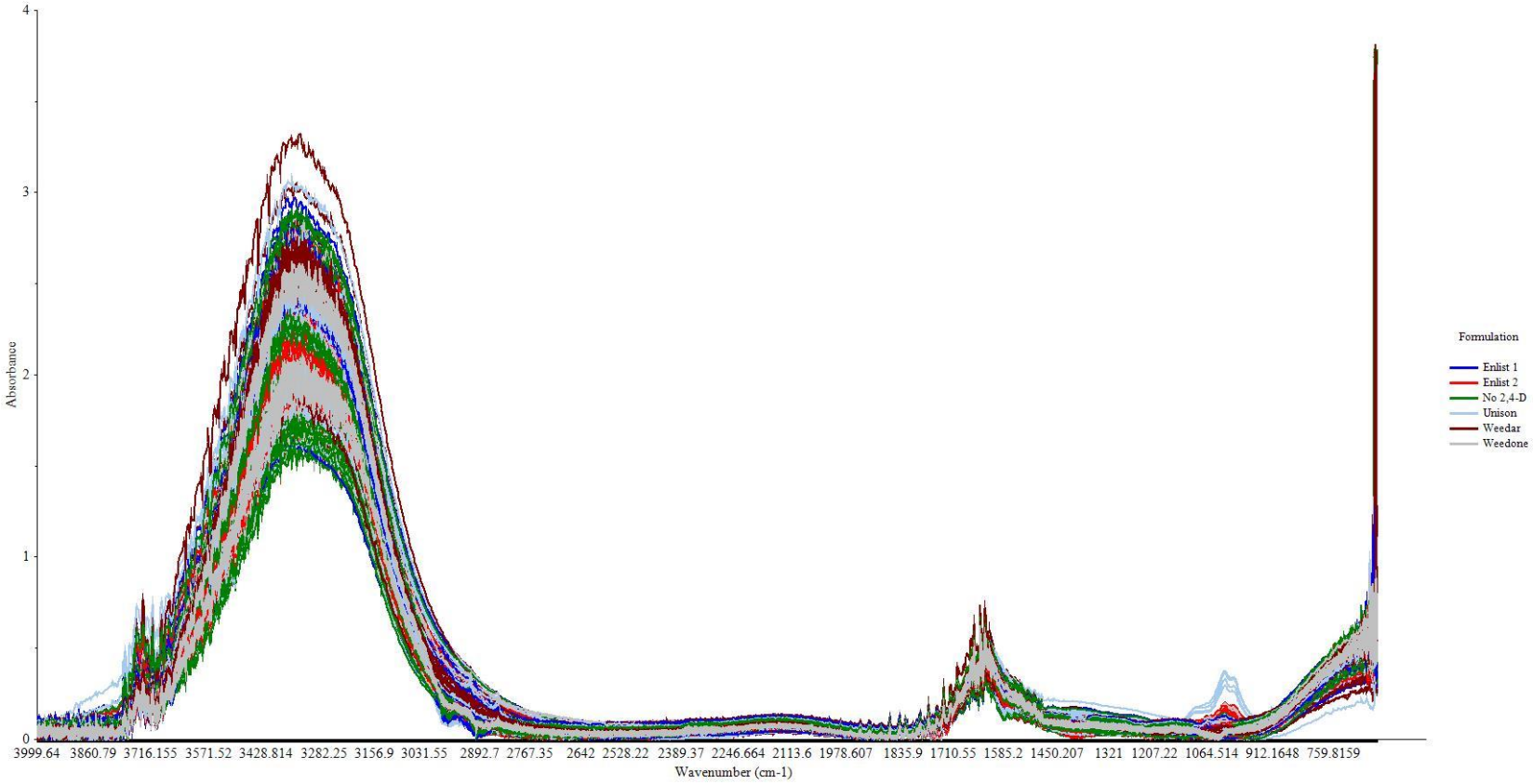

Figure 2.4 Raw spectra (4000 to 650 cm^{-1}) from soybean tissue treated with no 2,4-D, Unison, Weedar, Weedone, Enlist 1 or Enlist 2, pooled over sample timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a]

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture

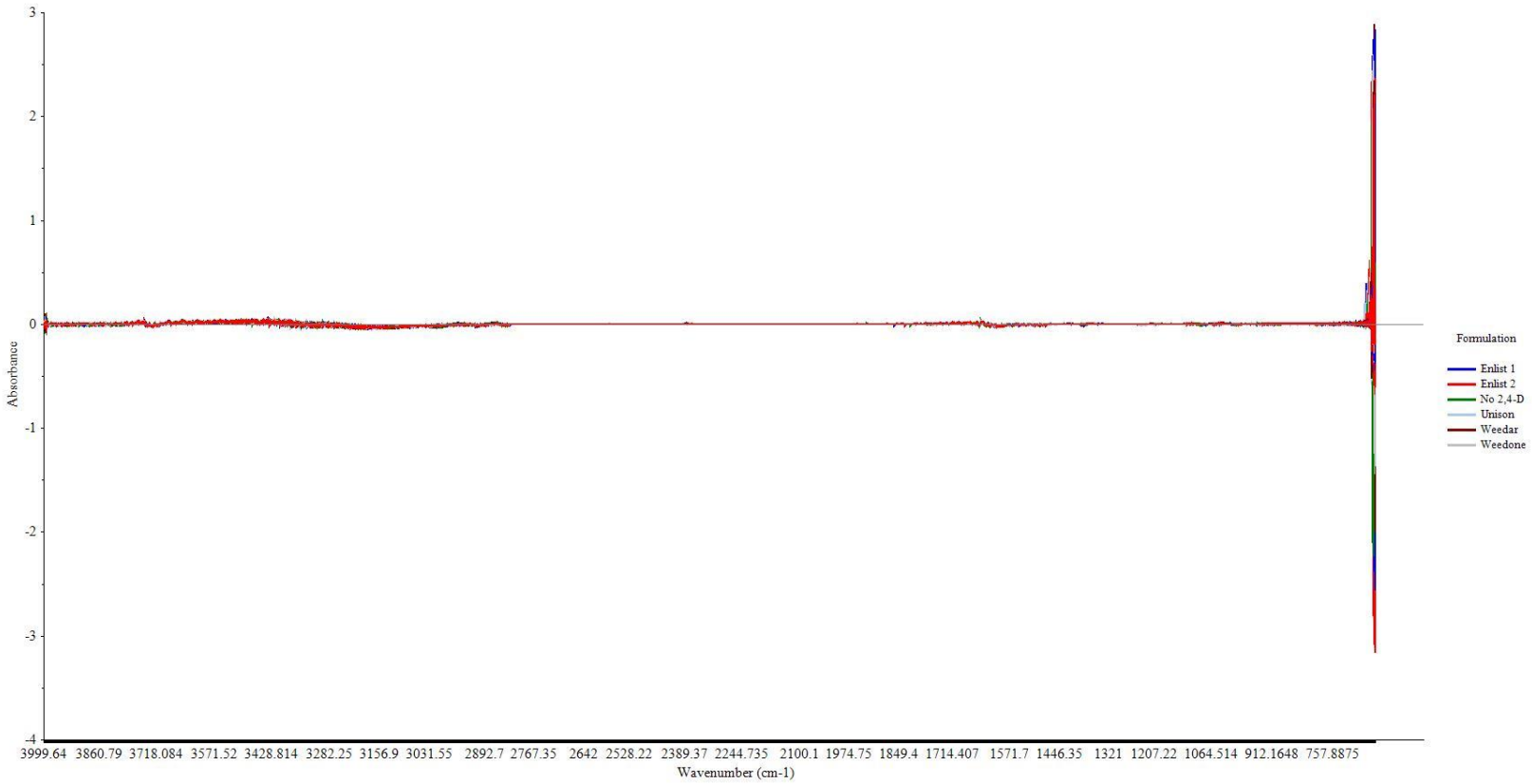

Figure 2.5 Transformed spectra (4000 to 650 cm^{-1}) from cotton tissue treated with no 2,4-D, Unison, Weedar, Weedone, Enlist 1 or Enlist 2, pooled over sample timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture

[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter

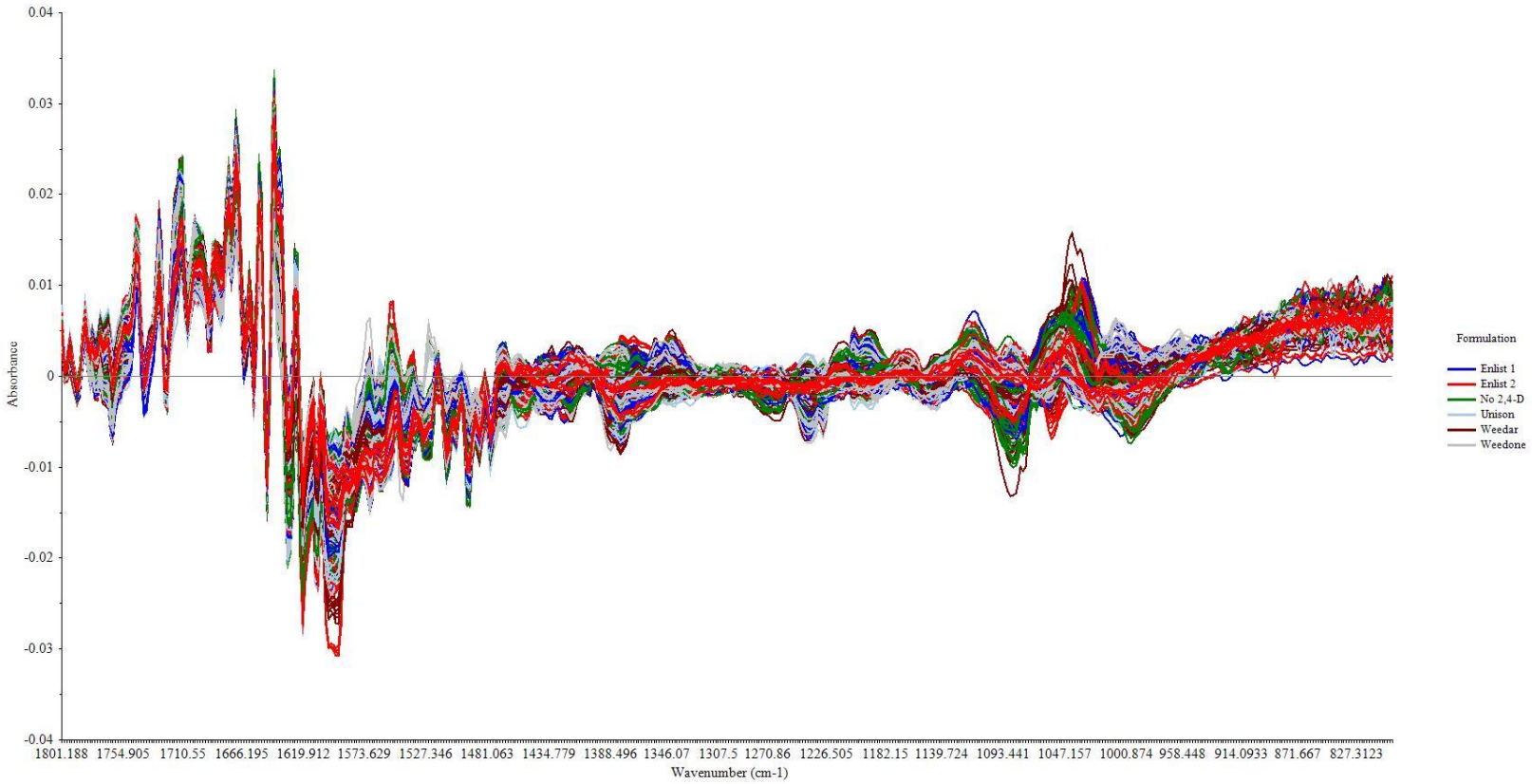

Figure 2.6 Transformed fingerprint (1800-800 cm^{-1}) spectra from cotton tissue treated with no 2,4-D, Unison, Weedar, Weedone, Enlist 1 or Enlist 2, pooled over sample timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture

[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter

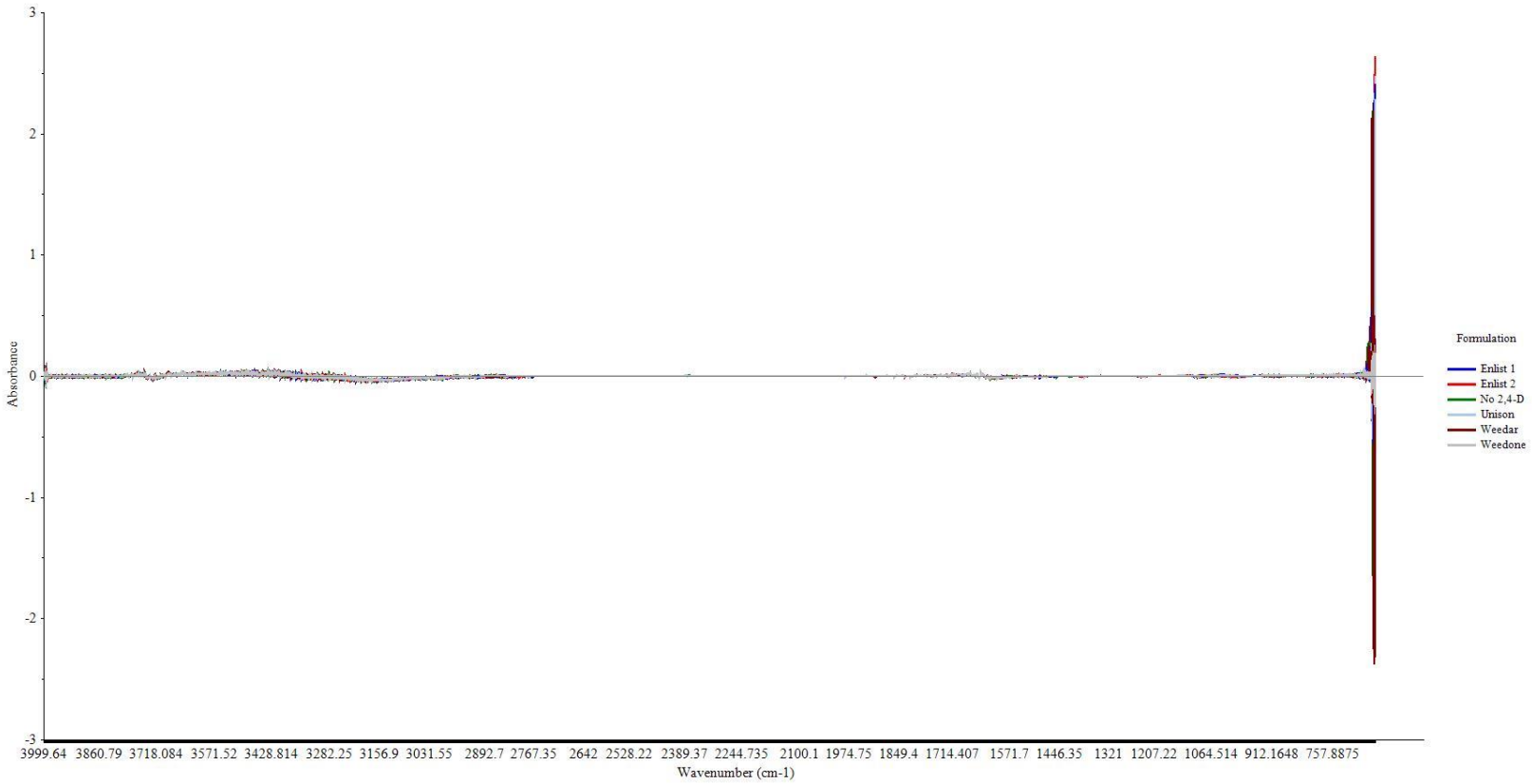

Figure 2.7 Transformed spectra (4000 to 650 cm^{-1}) from soybean tissue treated with no 2,4-D, Unison, Weedar, Weedone, Enlist 1 or Enlist 2, pooled over sample timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture

[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter

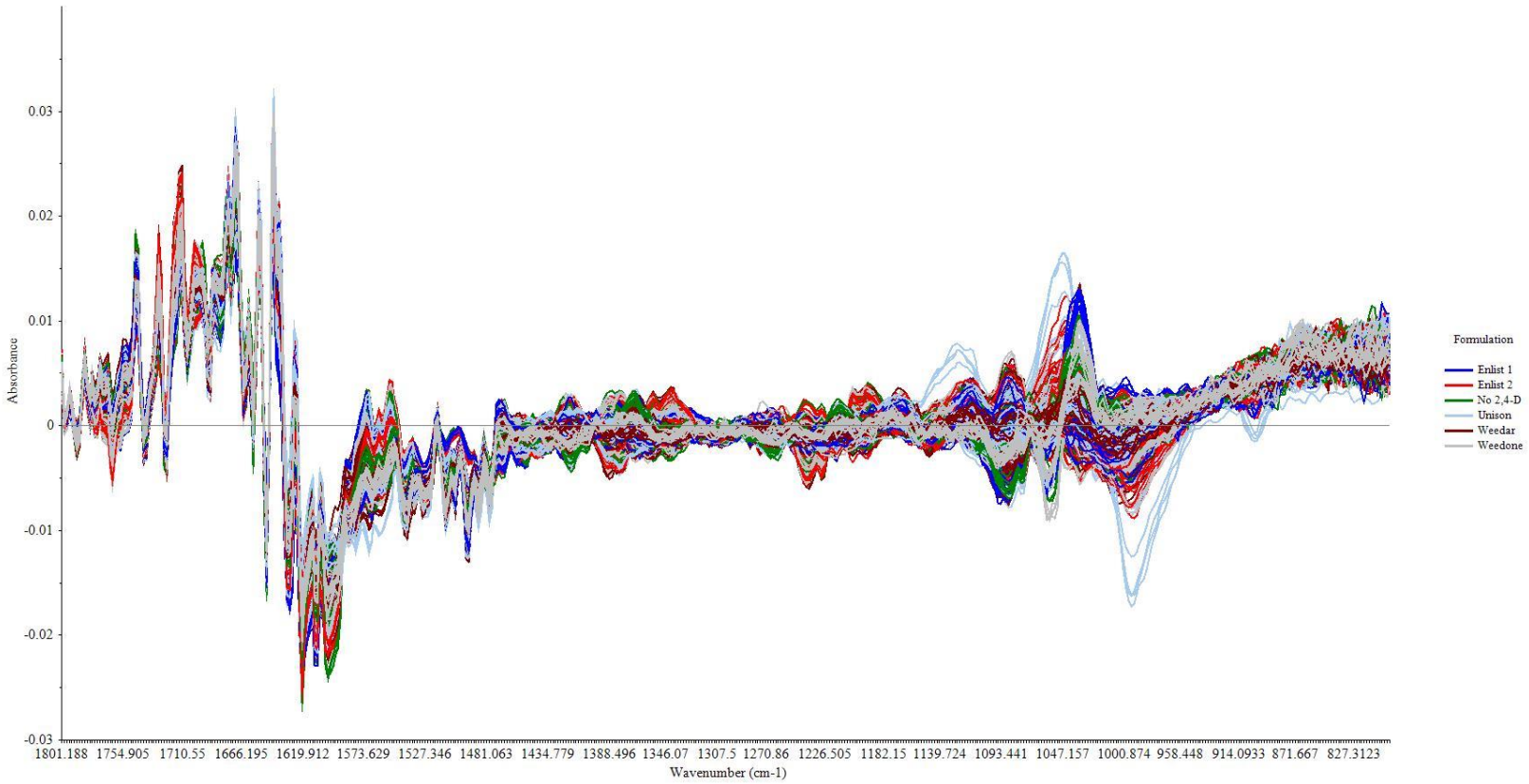

Figure 2.8 Transformed fingerprint (1800-800 cm^{-1}) spectra from soybean tissue treated with no 2,4-D, Unison, Weedar, Weedone, Enlist 1 or Enlist 2, pooled over sample timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture

[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter

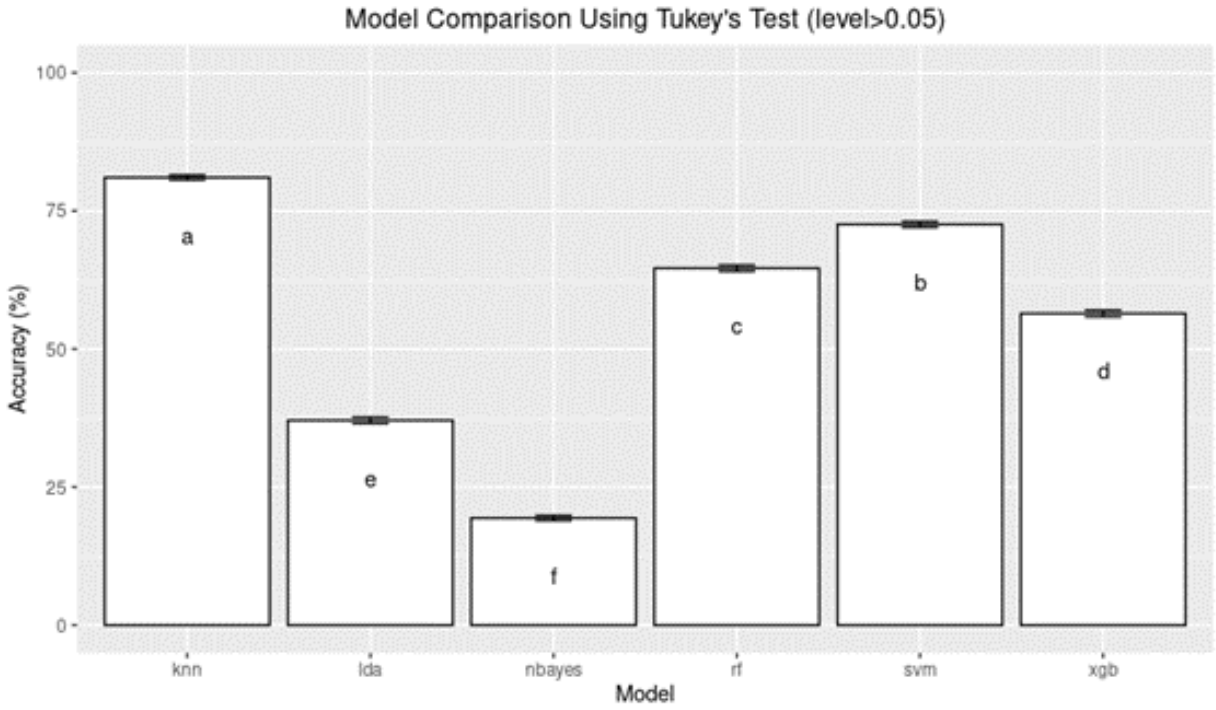

Figure 2.9 Comparison of supervised machine learning models using Tukey's honest significance difference test with transformed cotton spectra pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, Composite, and New Growth).[a,b]

[a]Abbreviations: knn, k-nearest neighbor; lda, principal component analysis-linear discriminant analysis; nbayes, Naïve Bayes; rf, Ranger; svm, support vector machine; xgb, XGBoost
[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

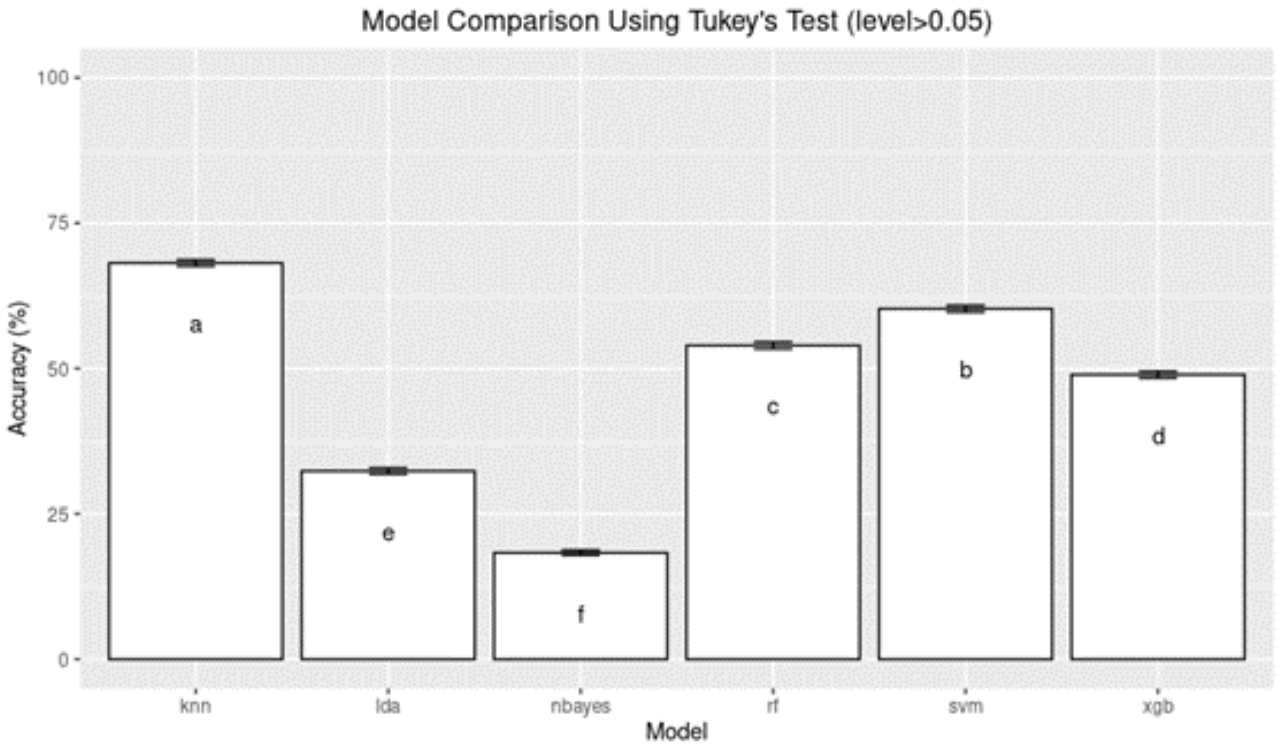

Figure 2.10 Comparison of supervised machine learning models using Tukey's honest significance difference test with transformed soybean spectra pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, Composite, and New Growth).[a,b]

[a]Abbreviations: knn, k-nearest neighbor; lda, principal component analysis-linear discriminant analysis; nbayes, Naïve Bayes; rf, Ranger; svm, support vector machine; xgb, XGBoost
[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

References

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encyclopedia of Research Design*, *3*(1), 1–5.

Albers, D. (1993). *Cotton Plant Development and Plant Mapping | MU Extension*. University of Missouri Extension. https://extension.missouri.edu/publications/g4268

Antoine Stevens and Leornardo Ramirez-Lopez (2021). An introduction to the prospectr package. R package Vignette R package version 0.2.2.

Binder M, Pfisterer F, Lang M, Schneider L, Kotthoff L, Bischl B (2021). "mlr3pipelines - Flexible Machine Learning Pipelines in R."_Journal of Machine Learning Research_, *22*(184), 1-7. <URL: https://jmlr.org/papers/v22/21-0281.html>.

Brandon Greenwell (2016). ramify: Additional Matrix Functionality. R package version 0.3.3.

https://CRAN.R-project.org/package=ramify

Buol, J. (2019). Stewarding 2,4-D- and dicamba- based weed control technologies in cotton and soybean production systems [Mississippi State University]. In *Theses and Dissertations*. https://scholarsjunction.msstate.edu/td/5049

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79. https://doi.org/10.1016/j.neucom.2017.11.077

David Gohel (2021). flextable: Functions for Tabular Reporting. R package version 0.6.10. https://CRAN.R-project.org/package=flextable

David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel and Friedrich Leisch (2021). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-9. https://CRAN.R-project.org/package=e1071

Dominic Comtois (2021). summarytools: Tools to Quickly and Neatly Summarize Data. R package version 1.0.0. https://CRAN.R-project.org/package=summarytools

Estimated Annual Agricultural Pesticide Use. (2021). https://water.usgs.gov/nawqa/pnsp/usage/maps/show_map.php?year=2019&map=24D&hilo=L

Fuente, R. D. la, Cancino, J., & Acuña, E. (2021). Comparison of machine learning methods for dry biomass estimation based on green logging residues chips. *International Journal of Forest Engineering*, *32*(2), 174–184. https://doi.org/10.1080/14942119.2021.1892415

H. Bengtsson, A Unifying Framework for Parallel and Distributed Processing in R using Futures, The R Journal, 2021, doi:10.32614/RJ-2021-048

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Hadley Wickham and Jennifer Bryan (2019). readxl: Read Excel Files. R package version 1.3.1.

https://CRAN.R-project.org/package=readxl

Hadley Wickham, Jim Hester and Winston Chang (2021). devtools: Tools to Make Developing R Packages Easier. R package version 2.4.2. https://CRAN.R-project.org/package=devtools

Hans W. Borchers (2021). pracma: Practical Numerical Math Functions. R package version 2.3.3. https://CRAN.R-project.org/package=pracma

Havens, P. L., Hillger, D. E., Hewitt, A. J., Kruger, G. R., Marchi-Werle, L., & Czaczyk, Z. (2018). Field Measurements of Drift of Conventional and Drift Control Formulations of 2,4-D Plus Glyphosate. *Technology*, *32*(5), 550–556. https://doi.org/10.2307/26567622

Heap, I. (2022, February 14). *The International Herbicide-Resistant Weed Database*. https://www.weedscience.org/Pages/SOASummary.aspx

Kalsing, A., Rossi, C. V. S., Lucio, F. R., Zobiole, L. H. S., Cunha, L. C. V. D., & Minozzi, G. B. (2018). Effect of Formulations and Spray Nozzles on 2,4-D Spray Drift under Field Conditions. *Weed Technology*, *32*(4), 379–384. https://doi.org/10.1017/WET.2018.18

Lang M, Binder M, Richter J, Schratz P, Pfisterer F, Coors S, Au Q, Casalicchio G, Kotthoff L, Bischl B (2019). "mlr3: A modern object-oriented machine learning framework in R." _Journal of Open Source Software_. doi: 10.21105/joss.01903 (URL:https://doi.org/10.21105/joss.01903), <URL: https://joss.theoj.org/papers/10.21105/joss.01903>.

Letaw Alathea (2015). captioner: Numbers Figures and Creates Simple Captions. R package version 2.2.3. https://CRAN.R-project.org/package=captioner

Lyon, B. R., Cousins, Y. L., Llewellyn, D. J., & Dennis, E. S. (1993). Cotton plants transformed with a bacterial degradation gene are protected from accidental spray drift damage by the herbicide 2,4-dichlorophenoxyacetic acid. *Transgenic Research 1993 2:3*, *2*(3), 162–169. https://doi.org/10.1007/BF01972610

Marc Becker, Michel Lang, Jakob Richter, Bernd Bischl and Daniel Schalk (2021). mlr3tuning: Tuning for 'mlr3'. R package version 0.9.0. https://CRAN.R-project.org/package=mlr3tuning

Matt Dowle and Arun Srinivasan (2021). data.table: Extension of `data.frame`. R package version 1.14.2. https://CRAN.R-project.org/package=data.table

Max Kuhn (2021). caret: Classification and Regression Training. R package version 6.0-88. https://CRAN.R-project.org/package=caret

McWilliams, D. A., Berglund, D. R., & Endres, G. J. (1999). *Soybean Growth and Management Quick Guide*. https://library.ndsu.edu/ir/bitstream/handle/10365/5453/a1174.pdf?sequence=1

Michel Lang, Patrick Schratz, Raphael Sonabend, Marc Becker and Jakob Richter (2021). mlr3viz: Visualizations for 'mlr3'. R package version 0.5.7. https://CRAN.R-project.org/package=mlr3viz

Michel Lang, Quay Au, Stefan Coors and Patrick Schratz (2021). mlr3learners: Recommended Learners for 'mlr3'. R package version 0.5.1. https://CRAN.R-project.org/package=mlr3learners

Miron B. Kursa, Witold R. Rudnicki (2010). Feature Selection with the Boruta Package. Journal of Statistical Software, 36(11), 1-13. URL http://www.jstatsoft.org/v36/i11/.

Mortensen, D. A., Franklin Egan, J., Maxwell, B. D., Ryan, M. R., & Smith, R. G. (2012). Navigating a Critical Juncture for Sustainable Weed Management. *BioScience*, *62*(1). https://doi.org/10.1525/bio.2012.62.1.12

Orooji, N., Takdastan, A., Yengejeh, R. J., Jorfi, S., & Davami, A. H. (2021). A quick and inexpensive method to determine 2,4-dichlorophenoxyacetic acid residues in water samples by hplc. *Desalination and Water Treatment*, *217*, 329–338. https://doi.org/10.5004/DWT.2021.26905

Patrick Schratz, Michel Lang, Bernd Bischl and Martin Binder (2021). mlr3filters: Filter Based Feature Selection for 'mlr3'. R package version 0.4.2. https://CRAN.R-project.org/package=mlr3filters

Perotti, V. E., Larran, A. S., Palmieri, V. E., Martinatto, A. K., & Permingeat, H. R. (2020). Herbicide resistant weeds: A call to integrate conventional agricultural practices, molecular biology knowledge and new technologies. *Plant Science*, *290*, 110255. https://doi.org/10.1016/J.PLANTSCI.2019.110255

Peterson, M. A., Mcmaster, S. A., Riechers, D. E., Skelton, J., & Stahlman, P. W. (2016). 2,4-D Past, Present, and Future: A Review. *Weed Technology*, *30*, 303–345. https://doi.org/10.1614/WT-D-15-00131.1

Purcell, L. C., Salmeron, M., & Ashlock, L. (2014). Soybean Growth and Development. In *Arkansas Soybean Production Handbook* (pp. 1–8).

Raphael Sonabend and Patrick Schratz (2021). mlr3extralearners: Extra Learners For mlr3. R package version 0.5.8.

Reid, C. (2017). Monitoring Aspergillus Flavus Progression and Aflatoxin Accumulation in Inoculated Maize (Zea Mays L.) Hybrids [Mississippi State University]. In *Theses and Dissertations*. https://scholarsjunction.msstate.edu/td/3195

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Russell V. Lenth (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.2. https://CRAN.R-project.org/package=emmeans

Savitzky, A., & Golay, M. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://pubs.acs.org/sharingguidelines

Shai, S.-S., & Shai, B.-D. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. https://books.google.com/books?id=Hf6QAwAAQBAJ&dq=what+is+machine+learning&lr=&source=gbs_navlinks_s

Sharkey, S. M., Williams, B. J., & Parker, K. M. (2021). Herbicide Drift from Genetically Engineered Herbicide-Tolerant Crops. *Environmental Science & Technology*, *55*(23), 15559–15568. https://doi.org/10.1021/ACS.EST.1C01906

Shyam, C., Chahal, P. S., Jhala, A. J., & Jugulam, M. (2021). Management of glyphosate-resistant Palmer amaranth (Amaranthus palmeri) in 2,4-D–, glufosinate-, and glyphosate-resistant soybean. *Weed Technology*, *35*(1), 136–143. https://doi.org/10.1017/WET.2020.91

signal developers (2013). signal: Signal processing. URL: http://rforge.rproject.org/projects/signal/.

Simonescu, C. M. (2012). Application of FTIR Spectroscopy in Environmental Studies. In M. Farrukh (Ed.), *Advanced Aspects of Spectroscopy* (pp. 49–53). IntechOpen. https://doi.org/10.5772/48331

Singh, M., & Sharma, S. D. (2000). EFFECT OF GLYPHOSATE AND ITS 2,4-D FORMULATION ON SOME DIFFICULT TO CONTROL WEEDS. *Proceedings of the Florida State Horticultural Society.*, 63–67.

Sonabend Raphael and Florian Pfisterer (2021). mlr3benchmark: Analysis and Visualisation of Benchmark Experiments. R package version 0.1.3. https://CRAN.R-project.org/package=mlr3benchmark

Sosnoskie, L. M., Culpepper, A. S., Braxton, L. B., & Richburg, J. S. (2015a). Evaluating the Volatility of Three Formulations of 2,4-D When Applied in the Field. *Weed Technology*, *29*(2), 177–184. https://doi.org/10.1614/wt-d-14-00128.1

Taiyun Wei and Viliam Simko (2021). R package 'corrplot': Visualization of a Correlation Matrix (Version 0.92). Available from https://github.com/taiyun/corrplot

Thompson, M. A., Steckel, L. E., Ellis, A. T., & Mueller, T. C. (2007). Soybean Tolerance to Early Preplant Applications of 2,4-D Ester, 2,4-D Amine, and Dicamba. *Weed Technology*, *21*(4), 882–885. https://doi.org/10.1614/WT-06-188.1

Torsten Hothorn, Frank Bretz and Peter Westfall (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal 50(3), 346--363.

Vencill, W. K., Nichols, R. L., Webster, T. M., Soteres, J. K., Mallory-Smith, C., Burgos, N. R., Johnson, W. G., & Mcclelland, M. R. (2022). *Herbicide Resistance: Toward an Understanding of Resistance Development and the Impact of Herbicide-Resistant Crops*. https://doi.org/10.1614/WS-D-11-00206.1

Werle, R., Mobli, A., Striegel, S., Arneson, N., Dewerff, R., Brown, A., & Oliveira, M. (2021). Large-scale evaluation of 2,4-D choline off-target movement and injury in 2,4-D-susceptible soybean. *Weed Technology*, 1–7. https://doi.org/10.1017/WET.2021.62

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Wright, T. R., Shan, G., Walsha, T. A., Lira, J. M., Cui, C., Song, P., Zhuang, M., ArnoldN, N. L., Lin, G., Yau, K., Russell, S. M., Cicchillo, R. M., Peterson, M. A., Simpson, D. M., Zhou, N., Ponsamuel, J., & Zhang, Z. (2010). Robust crop resistance to broadleaf and grass herbicides provided by aryloxyalkanoate dioxygenase transgenes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(47), 20240–20245. https://doi.org/10.1073/PNAS.1013154107/-/DCSUPPLEMENTAL

Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, p. 77. DOI:10.1186/1471-2105-12-77 http://www.biomedcentral.com/1471-2105/12/77/

Yihui Xie (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.36.

Zhang, Yaping, Yang, J., Shi, R., Su, Q., Yao, L., & Li, P. (2011). Determination of acetanilide herbicides in cereal crops using accelerated solvent extraction, solid-phase extraction and gas chromatography-electron capture detector. *Journal of Separation Science*, *34*(14), 1675–1682. https://doi.org/10.1002/JSSC.201100058

Zhang, Yonghong, Ge, T., Tian, W., & Liou, Y. A. (2019). Debris Flow Susceptibility Mapping Using Machine-Learning Techniques in Shigatse Area, China. *Remote Sensing*, *11*(23), 2801. https://doi.org/10.3390/RS11232801

CHAPTER III

THE INFLUENCE OF THE TIMING BETWEEN APPLICATION AND COLLECTION AND APPLICATION STATUS OF LEAF TISSUE IN 2,4-D CLASSIFICATION ACCURACY

Introduction

2,4-D was discovered by both British and American scientists in the 1940s and remained a secret due to World War II until its commercialization in 1945 (G. E. Peterson, 1967; M. A. Peterson et al., 2016). 2,4-D mimics phytohormones that control growth, called auxins, and is structurally similar to indole-3-acetic acid (IAA), the leading natural auxin (Song, 2014) (Figure 3.1). Despite the longevity of 2,4-D, the exact mode of action is not entirely understood (Tu et al., 2001). According to the EPA, 2,4-D functions by causing unregulated cell division in vascular tissue by increasing the cell wall's plasticity, increasing the biosynthesis of proteins, and increasing ethylene production (*2,4-D Technical Fact Sheet*, n.d.). However, recent studies have brought insight into the mode of action of 2,4-D at the molecular level. Song (2017) determined that at low concentrations of auxin, Aux/IAA proteins, which are negative regulators of auxin-responsive genes (Mockaitis & Estelle, 2008; Tiwari et al., 2001), will bind to ARF, which then binds to auxin response elements and inhibit auxin response genes (Guilfoyle et al., 2007). These auxin response genes play roles in plant development and have been differently expressed during abiotic and biotic stress, showing a connection between auxin and stress signaling pathways (Ghanashyam & Jain, 2009). At high concentrations of natural IAA, or when exposed to synthetic auxins, auxin will enter the cell and bind Aux/IAA proteins to F-box protein

TIR_1, which acts as auxin receptors and mediates the degradation of Aux/IAA (Song, 2014). This will then free up ARF and cause the activation of auxin response genes (Tan et al., 2007). With the activation of auxin response genes, an overproduction of chemicals will occur, like ethylene, abscisic acid (ABA), reactive oxygen species, and a decreased production of nitric oxide (Song, 2014). Ethylene is a plant hormone that mediates plant growth and development and adaptive responses to different stresses like drought, floods, and pathogen attacks (Chang, 2016). However, an overproduction of ethylene may produce herbicide-related responses like epinasty and senescence, and the production of ABA (Song, 2014). Abscisic acid is another plant hormone with roles in stress response and pathogen defense (Alazem & Lin, 2017), but an overproduction of ABA will lead to stomatal closure and plant death (Song, 2014). Lastly, 2,4-D application leads to an overproduction of reactive oxygen species, which could cause epinasty and senescence as well as activating responses against stress conditions (Pazmiño et al., 2011), and a reduction in nitric oxide, although the molecular mechanism of reduced nitric oxide by 2,4-D is unknown (Song, 2014).

The process of auxin overdose, like when a synthetic auxin herbicide is applied to a plant, occurs in three phases; stimulation, inhibition, and decay (Grossmann, 2010). In the stimulation phase, which occurs within the first hours of application, metabolic pathways activate, like the stimulation of ethylene biosynthesis, followed by unregulated growth and the build-up of abscisic acid (Grossmann, 2010). Following the stimulation phase, the inhibition phase occurs within 24 hours of application and contains the growth inhibition of roots and the shoot, decreased internode elongation, stomatal closure, and an overproduction of reactive oxygen species (Grossmann, 2010). Lastly, the decay phase occurs, characterized by chloroplast damage and chlorosis, destruction of the membrane and vascular system integrity, causing necrosis and

plant death (Grossmann, 2010). During 2,4-D OTM events, small concentrations of 2,4-D will move and damage susceptible crops, with even 1/300th of an applied rate of 2,4-D damaging grapes (Mohseni-Moghadam et al., 2016). As previously mentioned, OTM of 2,4-D produces similar visual symptomology in both soybeans and cotton, including leaf cupping and stunting, chlorosis, altered height, epinastic response in both the stem and petiole, and callus formation in the stems (Andersen et al., 2004; Buol et al., 2019; Egan et al., 2014; Marple et al., 2008a). These visual 2,4-D OTM symptomologies can occur within hours if the plants are proliferating or within a few days if the plants are growing more slowly (Brown et al., n.d.). Reid (2017) found differing accuracies of 2,4-D classification models, depending on the timing between treatment and application, with classification models constructed using samples collected 7 DAT and immediately after application, producing 90% accuracy. Buol (2019) concluded that developing various classification models based on the timing between treatment and collection could produce better accuracy than a pooled general model.

2,4-D translocation and uptake in herbicide-tolerant (HT) crops like cotton and soybean have recently been examined. Skelton et al. (2017) conducted experiments comparing the translocation, uptake, and metabolism of 2,4-D in HT and non-HT soybeans. Skelton et al. (2017) found no significant difference in the amount or rate of uptake of 2,4-D in HT and non-HT soybeans. Skelton et al. (2017) also examined if adjuvants (ADJ) and glyphosate influenced uptake in HT and non-HT soybeans. Treatments containing ADJ, including Enlist 2, had greater uptake of 2,4-D than treatments not containing ADJ, including 2,4-D + glyphosate. Enlist 2 had greater uptake compared to 2,4-D + glyphosate at 1 hour after application, but uptake was the same from 3 to 24 hours after application. Enlist 2 had a lower uptake than 2,4-D + ADJ and 2,4-D + ADJ + glyphosate but a faster rate of uptake. A significant difference in the translocation of

radiolabeled 2,4-D was observed for the different soybean lines, with more radiolabeled 2,4-D remaining in the treated leaf in HT soybeans compared to non-HT soybeans. In both HT and non-HT soybean lines, a majority (>90%) of radiolabeled 2,4-D remained in the treated leaf from 1 hour after application to 24 hours after application (Table 3.1). Skelton et al. (2017) also compared 2,4-D acid and 2,4-D-ethylhexyl ester uptake in HT soybeans. They found the amount of uptake to be equivalent for both formulations, but the rate of uptake of the ethylhexyl ester formulation to be much higher than the acid formulation at 1 and 3 hours after application but equivalent at 6 and 24 hours after application. Once again, less radiolabeled 2,4-D translocated out of treated leaves in HT soybeans than non-HT soybeans. Skelton et al. (2017) theorized that this difference could be physiologically connected to the detoxification of 2,4-D and the storage of metabolites in treated leaves of HT soybeans, while non-HT soybeans are not able to detoxify or sequester 2,4-D irreversibly.

Perez et al. (2021) conducted experiments on the translocation and uptake of 2,4-D in HT chromosome substitution (CS) lines of cotton versus a susceptible parent cotton line (TM-1). The CS lines' uptake was 15-23%, while the TM-1 line produced an uptake of around 1% 24 hours after treatment. When examining the translocation of 2,4-D in these different cotton lines, the CS lines all produced a similar pattern, in which a vast majority of 2,4-D (>90%) remained in the treated leaf, while only 23% of the radiolabeled 2,4-D remained in the treated leaf in the TM-1 line (Figure 3.2)**.** The remaining 77% of the radiolabeled 2,4-D translocated above and below the treated leaf in TM-1 cotton, while only about 5% of 2,4-D left the treated leaf in CS cotton (Table 3.1). A greater amount of 2,4-D was transported below the treated leaf (42%) than above the treated leaf (35%) in the TM-1 line, while a majority of the CS lines had roughly the same amount of 2,4-D transported above and below the treated leaf (5%). When measuring the amount

of movement of 2,4-D every twelve hours for forty-eight hours using all four cotton lines tested, each line produced different results. In the CS-T04-15 line, 2,4-D movement into the tissue above the treated leaf increased between 12 and 24 hours and stagnated between 24 and 48 hours. Regarding the movement of 2,4-D to below the treated leaf, there was no significant change in the movement of 2,4-D throughout the time interval. For CS-T07 and TM-1, radiolabeled 2,4-D slowly moved to tissue above and below the treated leaf at the same rate. Lastly, in the CS-B15sh line, the radiolabeled 2,4-D movement was similar to CS-T07 and TM-1 for the first 24 hours, but movement from the treated leaf to above the treated leaf increased after 24 hours. Perez et al. believed the differences in uptake from the CS lines and TM-1 might be caused by the cells at the treatment site dying quickly and influencing the movement of 2,4-D from the treated area. Perez et al. concluded that the small amount of radiolabeled 2,4-D translocation could be caused by the treated leaf sequestering 2,4-D and its derived metabolites.

Previous classification models constructed by Reid (2017) and Buol (2019) were built using leaf tissue samples that were "damaged" and did not examine whether the application status of the "damaged" leaf samples would affect the accuracy of their classification models. However, recent research has shown that in HT soybeans and cotton, most of the radiolabeled 2,4-D remains in the treated leaf and a minimal amount moved above or below the treated tissue (Perez et al., 2021; Skelton et al., 2017).

Objective

To ascertain the influence of leaf tissue treatment status and the timing between treatment and collection of leaf tissue samples in correctly classifying 2,4-D formulations, experiments were conducted utilizing KNN classification models and FTIR spectrometry of soybean and cotton tissue samples that were directly treated with, not directly treated with, or a combination

of both treated with and not directly treated with, various formulations of 2,4-D and collected at multiple sample timing periods.

Methods and Materials

Experimental Design, Treatments, and Data Sampling

The same leaf tissue samples used in the previous chapter were used in this chapter. All experimental Design, treatment, and data sampling were the same.

Sample Processing, Data Collection, and Data Preprocessing

The same leaf tissue samples used in the previous chapter were used in this chapter. All sample processing, data collection, and Data preprocessing were the same.

Data Analysis Using Machine Learning Algorithms

The same leaf tissue samples used in the previous chapter were used in this chapter. Data analysis was conducted using the mlr3 package (Lang et al., 2019). Nested cross-validation with five outer loops with three repeats and five inner loops with three repeats were used to set the parameters, train, and test all the models. It then went through thirty tuning iterations, producing six thousand five hundred fifty models per machine learning algorithm. After creating a general pooled model in the previous chapter, showing the possibility of using machine learning algorithms and FTIR spectroscopy to classify 2,4-D formulations, multiple classification models were constructed using subsets of the data depending on the timing and treatment status of the tissue. Only the k-nearest neighbor algorithm was used in constructing these models, as it produced the highest accuracy in constructing the pooled general model. Before constructing the models, the subset function found in the base RStudio package was used to subset the data based

on leaf tissue sample timing and treatment status. Succeeding this sub-setting, multiple k-nearest neighbor models were constructed using these various filtered data frames.

Comparing Machine Learning Models

All machine learning models will produce a classification matrix containing the predicted formulation of 2,4-D versus the actual reference 2,4-D formulation, the sensitivity of all formulations, and the positive predictive for all formulations. Following the construction of all the classification models, the models were fit into a binomial generalized linear model using the stats package (R Core Team, 2020), mean separated and compared using Tukey's Honest Significance Difference (HSD) test from the emmeans package (Length, 2022) with a significance level of 0.05 to determine if timing and treatment status influence the models' accuracy. Tukey HSD is used to determine the honest significant difference between two means using a studentized range and is a conservative test compared to other comparison methods (Abdi & Williams, 2010). For example, Tukey's HSD has been used to compare multiple machine learning model's abilities to estimate the dry biomass weight of wood chip residues (Fuente et al., 2021) and in comparing machine learning models performances in assessing triggering factors for debris flow (Yonghong Zhang et al., 2019).

Results and Discussion

Cotton

Two thousand four hundred forty-six cotton spectra were used in data analysis. Prior to feature selection, absorbance values for 1728 wavenumbers were collected, with 434 wavenumbers being labeled as important and used in machine learning analysis. The general

pooled classification model, using all cotton tissue samples without subsampling, produced an overall accuracy of 80%, with all formulation sensitivities being above 75% (Table 3.2).

Timing Model Comparisons

One hundred ninety-two spectra resulting from tissue samples collected immediately following treatment were used to construct a 0 DAT classification model. The model performed best when identifying Weedone (85%) and performed worst when identifying Unison (64%), resulting in an overall accuracy of 76% (Table 3.3)**.** A classification model built using tissue samples collected 7 days after initial treatment was used to construct a 7 DAT classification model with 567 spectra. This model produced an overall accuracy of 83% and achieved the worst accuracy in identifying Enlist 1 and the control (80%) and achieved the best accuracy in identifying Weedone **(**88%) (Table 3.4). A subset of 564 spectra produced using tissue samples collected 14 days after treatment was utilized to construct a 14 DAT classification model resulting in an overall accuracy of 88% (Table 3.5)**.** Enlist 1 was best classified (91%)**,** and Weedone resulted in the worst classification accuracy (85%) (Table 3.5). A 21 DAT classification model was built employing 561 spectra produced from tissue samples collected 21 days after treatment. The classification model performed the best classifying the control (91%) and worst when classifying Enlist 1 (74%), resulting in an overall classification accuracy of 81% (Table 3.6). Lastly, 563 spectra from tissue samples collected 28 days after treatment were utilized to construct a 28 DAT classification model, resulting in an overall accuracy of 78%, with all formulation sensitivities being above 75% (Table 3.7).

Treatment Status Model Comparisons

A total of 953 spectra produced by tissue samples directly treated with 2,4-D were used to construct a Treated classification model. This model produced an overall accuracy of 86%, performing worst when classifying Enlist 1 (84%) and best when classifying Weedar (89%) (Table 3.8). Seven hundred fifty-five cotton spectra were used to create the New Growth classification model, using tissue samples that grew post-treatment and were not directly treated with 2,4-D. This model performed best when classifying the control (93%) and worst when classifying Unison (80%) and resulted in an overall accuracy of 85% (Table 3.9). Finally, for the classification model built using a combination of tissue samples directly treated with 2,4-D and tissue samples that grew post-treatment and were not directly treated with 2,4-D, 739 spectra were used to construct the Composite classification model. The overall accuracy for the composite model is 79%, and the model classified Weedone with the highest accuracy (84%) and classified Enlist 1 with the lowest accuracy (74%) (Table 3.10).

Timing and Treatment Status Model comparisons

7 DAT Model Results

A classification model constructed from 192 spectra produced by tissue samples collected 7 days after treatment and directly treated with 2,4-D was used to construct a 7 DAT Treated classification model. The overall accuracy of the classification model was 83%, with the model best-classifying Weedone (90%) and inadequately classifying Enlist 2 (80%) (Table 3.11). A subset of 192 spectra collected from tissue samples that were collected 7 days after treatment and were a combination of tissue samples that were directly treated with 2,4-D and tissue samples that grew post-treatment was used to construct a 7 DAT Composite classification model. This model's overall classification accuracy was 80%, and the model poorly identified the

control (63%) and best identified Weedone (90%) (Table 3.12). One hundred eighty-three spectra produced from tissue samples collected 7 days after treatment that grew post-treatment and were not directly treated with 2,4-D were used to construct a 7 DAT New Growth classification model. The model achieved the best classification accuracy when identifying the control (96%) and achieved the worst classification accuracy when identifying Unison and Enlist 1 (80%), thus resulting in an overall accuracy of 87% (Table 3.13).

14 DAT Model Results

A subset of 188 spectra collected from tissue samples directly treated with 2,4-D and collected 14 days after treatment was used to construct a 14 DAT Treated classification model producing an overall accuracy of 96% (Table 3.14). This model performed best when classifying Weedone (100%) and worst when classifying Unison (90%) (Table 3.14). Next, a classification model was constructed using 188 spectra produced from tissue samples collected 14 days after treatment and a combination of tissue samples directly treated with 2,4-D and tissue samples that grew post-treatment and not directly treated with 2,4-D. This model, labeled as 14 DAT Composite classification model, resulted in an overall accuracy of 82% and best classified Enlist 1 (88%) while poorly classifying Unison and the control (76%) (Table 3.15). Finally, Tissue samples were collected 14 days after treatment and grew post-treatment and were not directly applied with 2,4-D generated 188 spectra that were utilized to construct a 14 DAT New Growth classification model. This classification model resulted in an overall accuracy of 97% and perfectly classified the control and Unison (100%) and inadequately classified Weedar (93%) (Table 3.16).

21 DAT Model Results

A subset of 189 spectra, generated from tissue samples collected 21 days after treatment and directly treated with 2,4-D, were used to create a 21 DAT Treated classification model producing an overall accuracy of 92%, with all formulation sensitivities being over 85% (Table 3.17). A classification model using tissue samples collected 21 days after treatment and a combination of tissue directly treated with 2,4-D and tissue that grew post-treatment produced 180 spectra that were used to construct a 21 DAT Composite classification model. This model resulted in an overall accuracy of 81% and performed best when classifying the control (90%) and worst when classifying Enlist 2 (67%) (Table 3.18). Finally, tissue samples collected 21 days after treatment that grew post-treatment of 2,4-D generated 192 spectra which were then used to construct a 21 DAT New Growth classification model. This model best classified the control (97%) and Weedar (93%) while poorly classifying Weedone (80%) and Enlist 1 (78%), resulting in an overall accuracy of 86% (Table 3.19).

28 DAT Model Results

A subset of 192 spectra generated from tissue samples collected 28 days after treatment and directly treated with 2,4-D was used to develop a 28 DAT Treated classification model. This model achieved an overall accuracy of 92% and best identified Unison (100%), Enlist 1 (100%), and Weedar (97%) while poorly identifying Weedone (81%) and Enlist 2 (83%) (Table 3.20). Next, a 28 Composite classification model was constructed using 179 spectra generated from tissue samples collected 28 days after treatment that were both directly treated with 2,4-D and grew post-treatment. This classification model best classified the control (91%) and poorly classified Enlist 1 (60%), resulting in an overall accuracy of 77% (Table 3.21). Lastly, Tissue samples collected 28 days after treatment that grew post-treatment produced 192 spectra were

used to construct a 28 DAT New Growth classification model. This classification model achieved an overall accuracy of 81% while best classifying the control (85%) and poorly classifying Weedar (59%) (Table 3.22).

Model Comparison Results

Using Tukey's HSD, the various timing, treatment status, and general pooled models were compared. Regarding timing model comparison, the 14 DAT classification model was more accurate than all other models, including the pooled general model (Figure 3.3). These findings are similar to Reid (2017), in which the classification models using tissue samples collected 14 and 28 DAT produced higher classification accuracies than other models using earlier collected tissue samples. For treatment status model comparison, the Treated classification and New Growth models were more accurate than the Composite classification and the general pooled model (Figure 3.4). Research conducted by Skelton (2017) and Perez (2021) displayed that for HT soybeans and cotton, a vast majority of 2,4-D remains in the treated leaf, at least for the first 24 hours. This could explain why the treated model performed better than the pooled general model. However, Skelton (2017) and Perez (2021) did not observe 2,4-D in those crops past two days, so more 2,4-D could translocate after two days to tissue that grew after treatment, which could explain why the model built using new growth tissue also performed better than the pooled general model. Lastly, when comparing timing and treatment status model**s,** five classification models were more accurate than the pooled overall general model, including the 7 DAT New Growth model, 14 DAT New Growth model, 14 DAT Treated model, 21 DAT Treated model, and 28 DAT Treated model, with the 14 DAT New Growth model and the 14 DAT Treated model producing the highest accuracies (Figure 3.5). Three out of the five models that produced higher accuracy than the pooled general model were models constructed using tissue samples

directly treated with 2,4-D. Once again, these results are more likely caused by a majority of the 2,4-D remaining in the treated tissue rather than translocating to other plant areas.

Soybean

Two thousand four hundred forty-six cotton spectra were used in data analysis. Prior to feature selection, absorbance values for 1728 wavenumbers were collected, with 331 wavenumbers being labeled as important and used in machine learning analysis. The general pooled classification model, using all cotton tissue samples without subsampling, produced an overall accuracy of 68%, with all formulation sensitivities being above 65% (Table 3.23).

Timing Model Comparisons

One hundred ninety-two spectra resulting from tissue samples collected immediately following treatment were used to construct a 0 DAT classification model. The model performed best when identifying Enlist 1 (88%) and performed worst when identifying Weedar (48%), resulting in an overall accuracy of 71% (Table 3.24). Five hundred sixty-seven spectra produced from tissue samples collected 7 days after initial treatment were used to construct a 7 DAT classification model. This model produced an overall accuracy of 69% and achieved the worst accuracy in identifying Enlist 1 (64%) and the best accuracy in identifying Unison (77%) (Table 3.25). A subset of 564 spectra produced using tissue samples collected 14 days after treatment was utilized to construct a 14 DAT classification model resulting in an overall accuracy of 68% (Table 3.26). Unison resulted in the best classification accuracy (72%) and Enlist 2 resulted in the worst classification accuracy (65%) (Table 3.26). A classification model employing 561 spectra created from tissue samples collected 21 days after treatment was used to build a 21 DAT classification model. The classification model performed best when classifying

Weedone (86%) and worst when classifying Enlist 2 and Weedar (72%), resulting in an overall classification accuracy of 77% (Table 3.27). Lastly, 563 spectra from tissue samples collected 28 days after treatment were used to construct a 28 DAT classification model, resulting in an overall accuracy of 70%, with the model best classifying Weedar (82%) and poorly classifying the control (66%) **(**Table 3.28).

Treatment Status Model Comparisons

Nine hundred fifty-three spectra produced by tissue samples directly treated with 2,4-D were used to construct a Treated classification model. This model produced an overall accuracy of 80%, performing poorly when classifying Weedar (72%) and best when classifying Enlist 2 (84%) (Table 3.29). Seven hundred fifty-five cotton spectra were used to create the New Growth classification model, using tissue samples that grew post-treatment and were not directly treated with 2,4-D. This model performed best when classifying Enlist 1 (81%) and worst when classifying the control **(**70%), resulting in an overall accuracy of 75% (Table 3.30). Finally, tissue samples directly treated with 2,4-D and tissue samples that grew post-treatment and were not directly treated with 2,4-D generated 739 spectra which were utilized to construct the Composite classification model. The overall accuracy for the composite model was 74%, with the model classifying Enlist 1 with the highest accuracy (86%) and poorly classifying the control (69%) (Table 3.31).

Timing and Treatment Status Model comparisons

7 DAT Model Results

A classification model constructed from 192 spectra produced by tissue samples collected 7 days after treatment and directly treated with 2,4-D was used to construct a 7 DAT Treated

classification model. The overall accuracy of the classification model was 79%, with the model best-classifying Unison (94%) and poorly classifying Weedar (64%) (Table 3.32). A subset of 192 spectra manufactured from tissue samples collected 7 days after treatment and were a combination of tissue samples that were directly treated with 2,4-D and tissue samples that grew post-treatment were used to construct a 7 DAT Composite classification model. This model's overall classification accuracy was also 79%, and the model poorly identified the control and Weedone (74%) while correctly identifying Enlist 1 (89%) and Unison (88%) best (Table 3.33). Tissue samples collected 7 days after treatment that grew post-treatment and were not directly treated with 2,4-D yielded 183 spectra which were used to construct a 7 DAT New Growth classification model. The model achieved the best classification accuracy when identifying Weedone (81%) and Enlist 1 (80%) and achieved the worst classification accuracy when identifying Enlist 2 (66%), resulting in an overall accuracy of 74% (Table 3.34).

14 DAT Model Results

A subset of 188 spectra induced from tissue samples directly treated with 2,4-D and collected 14 days after treatment was utilized to assemble a 14 DAT Treated classification model producing an overall accuracy of 83% (Table 3.35). This model performed best when classifying Weedone and Enlist 1 (89%) and worst when classifying Enlist 2 (76%) (Table 3.35). Next, a classification model was erected using 188 spectra produced from tissue samples collected 14 days after treatment and a combination of tissue samples directly treated with 2,4-D and tissue samples that grew post-treatment and not directly treated with 2,4-D. This model, labeled as 14 DAT Composite classification model, resulted in an overall accuracy of 72% and best classified the control (86%) while poorly classifying Unison (59%) (Table 3.36). Lastly, tissue samples harvested 14 days after treatment and were not directly applied with 2,4-D generated 188 spectra

and were utilized to construct a 14 DAT New Growth classification model. This classification model resulted in an overall accuracy of 74% and best classified the control (82%) and poorly classified Weedone (63%) (Table 3.37).

21 DAT Model Results

A subset of 189 spectra, generated from tissue samples that were collected 21 days after treatment and directly treated with 2,4-D, were used to assemble a 21 DAT Treated classification model producing an overall accuracy of 89%, with the model best classifying Enlist 2 (97%) and inadequately classifying Enlist 1 (80%) (Table 3.38). A classification model using 180 spectra produced via tissue samples harvested 21 days after treatment and containing tissue directly treated with 2,4-D and tissue that grew post-treatment were utilized to build a 21 DAT Composite classification model. This model resulted in an overall accuracy of 82% and performed best when classifying Unison (92%) and worst when classifying Enlist 2 and Weedar (77%) (Table 3.39). Tissue samples collected 21 days after treatment that grew post-treatment of 2,4-D generated 192 spectra were used to construct a 21 DAT New Growth classification model. This model best classified Enlist 1 (89%) and Weedar (88%) while poorly classifying Enlist 2 (75%) and the control (76%), resulting in an overall accuracy of 80% (Table 3.40).

28 DAT Model Results

A subset of 192 spectra generated from tissue samples collected 28 days after treatment that were directly treated with 2,4-D was used to develop a 28 DAT Treated classification model. This model achieved an overall accuracy of 92% and best identified Weedar (100%) and Unison (94%), with all formulation sensitivities being at or above 90% (Table 3.41). Next, a 28

Composite classification model was assembled employing 179 spectra generated from tissue samples collected 28 days after treatment that were both directly treated with 2,4-D and grew post-treatment. This classification model best classified Enlist 1 (83%) and poorly classified the control (62%), resulting in an overall accuracy of 76% (Table 3.42). Lastly, Tissue samples harvested 28 days after treatment that grew post-treatment produced 192 spectra which were used to build a 28 DAT New Growth classification model. This classification model achieved an overall accuracy of 79% while best classifying Weedar (89%) and poorly classifying the control (69%) (Table 3.43).

Model Comparison Results

Using Tukey's HSD, the various timing, treatment status, and general pooled models were compared. Regarding timing model comparison, the 21 DAT classification model was more accurate than all other models, including the pooled general model (Figure 3.6). Unfortunately, Reid (2017) did not harvest samples 21 DAT, so a comparison cannot be made. For the treatment status model comparison, the Treated classification model was more accurate than all other treatment status models, including the general model (Figure 3.7). Research conducted by Skelton (2017) and Perez (2021) showed that for HT soybeans and cotton, a vast majority of 2,4-D remains in the treated leaf, at least for the first 24 hours. This could explain why the treated model performed better than all other treatment status models. In addition, the model assembled using treated cotton tissue also performed better than the pooled general model and composite model, similar to the soybean results. Lastly, when comparing timing and treatment status models, all the classification models were more accurate than the pooled overall general model, except for the 14 DAT composite model, 14 DAT New Growth model, and 7 DAT New Growth model, with the 21 DAT Treated and 28 DAT Treated models producing the highest overall

accuracies (Figure 3.8). While most classification models produced higher accuracy than the pooled general model, the models with the highest overall accuracy were constructed with tissue directly treated with 2,4-D. These results could be caused by a majority of the 2,4-D remaining in the treated tissue rather than translocating to other plant areas. However, more studies on 2,4-D translocation in HT crops must be completed for a definitive cause.

Conclusions and Regulatory Implications

As previously stated, OTM drift events are more than likely going to increase with the mass adoption of the Enlist weed control systems and the increase in 2,4-D application, particularly off-labeled 2,4-D formulations that are generally cheaper than labeled 2,4-D formulations but much more volatile. These OTM drift events can decrease yields depending on growth stage, the amount of herbicide that drifted, and the number of times that an OTM drift event occurred (Everitt & Keeling, 2009; Marple et al., 2008b; Oakley, 2021). The general pooled model constructed in the previous displays how effective FTIR spectroscopy and machine learning algorithms could be in construction classification models to identify 2,4-D formulations. With the various classification models constructed, farmers and regulators can collect the most optimized leaf tissue samples for greater accuracy than other models. In addition, with a majority of timing and treatment status models producing higher accuracies than the general pooled model, the timing between treatment and harvest and the treatment status of the leaf tissue both influence the accuracy of a 2,4-D formulation classification model. These findings concur with Buol (2019), who theorized that producing multiple classification models based on timing would produce high accuracies than an overall model and better classify 2,4-D formulations.

Similar to the general pooled models constructed in the previous chapter, these models only used one year and one location's worth of data. More research is needed to evaluate for a

possible yearly or location interaction. More samples could also be utilized to validate the accuracy of these models, as no validation was conducted for the timing and treatment status models. Lastly, this experiment could also be reproduced on a smaller scale to precisely monitor which leaf tissue was applied with herbicide. It could also be reproduced with treatments involving 2,4-D formulations without the active ingredient to determine whether the models use the active ingredient or adjuvant when classifying 2,4-D formulations.

When investigating a possible 2,4-D OTM event and attempting to achieve the most optimal results, regulators and investigators should collect tissue 21 & 28 days after treatment in soybean and 14 days after treatment in cotton, from petioles at the bottom of the plant and from multiple individuals to avoid collecting tissue samples with different treatment statuses for both soybeans and cotton. With this new information, 2,4-D formulations can be identified accurately and coupled with previous models constructed to determine which 2,4-D formulation caused a specific 2,4-D OTM drift event, sales receipts, and application records, 2,4-D OTM drift events can be better monitored and investigated.

Tables

Table 3.1 Translocation of ^{14}C material from whole plant assay in Enlist and non-transgenic (NT) soybean lines.[a]

		Hours after application			
		1	3	6	24
Soybean line	Plant portion	% ^{14}C-material in plant portion			
NT	Treated leaf	99.5 c	98.8 c	96.4 d	93.0 d
	above	0.2	0.8	2.5	3.6
	below	0.3	0.4	1.0	3.5
Enlist	Treated leaf	99.6 c	99.3 c	98.9 c	98.3 c
	above	0.1	0.2	0.4	0.6
	below	0.3	0.4	0.6	1.2

[a]Significant differences were determined between soybean lines in the amount ^{14}C-material remaining in the treated leaf (p = <0.0001), while no significant differences were found in either soybean line between amounts of ^{14}C found above or below the treated leaf (p =0.56). ^{14}C-material is expressed as a percentage of 2,4-D uptake. Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

Table 3.2 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	886	46	47	70	50	65	76%
Enlist 2	64	1017	25	46	43	66	81%
No 2,4-D	47	34	1026	54	35	19	84%
Unison	80	37	60	987	68	37	78%
Weedar	50	44	32	34	975	57	82%
Weedone	49	70	34	42	41	1004	81%
Sensitivity	75%	81%	84%	80%	80%	80%	80%[†]

Hyperparameters: filt2.filter.nfeat= 426.8003; k= 5.031281
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.3 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue collected at 0 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	68	5	5	6	3	1	77%
Enlist 2	5	78	0	5	13	0	77%
No 2,4-D	1	0	81	4	0	0	94%
Unison	16	2	4	61	7	3	66%
Weedar	6	9	3	8	68	10	65%
Weedone	0	2	3	12	5	82	79%
Sensitivity	71%	81%	84%	64%	71%	85%	76%[†]

Hyperparameters: filt2.filter.nfeat= 353.9535; k= 2.368775
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.4 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue collected at 7 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	229	8	12	13	3	7	84%
Enlist 2	13	234	12	3	5	18	82%
No 2,4-D	10	4	227	14	8	3	85%
Unison	10	17	13	239	24	4	78%
Weedar	15	7	11	11	224	3	83%
Weedone	11	18	10	5	3	253	84%
Sensitivity	80%	81%	80%	84%	84%	88%	83%[†]

Hyperparameters: filt2.filter.nfeat= 421.5464; k= 1.61777
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.5 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue collected at 14 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	230	4	11	7	8	8	86%
Enlist 2	3	258	0	9	8	24	85%
No 2,4-D	5	2	249	11	9	4	89%
Unison	6	6	17	256	4	4	87%
Weedar	6	4	8	5	255	4	90%
Weedone	2	14	0	0	7	244	91%
Sensitivity	91%	90%	87%	89%	88%	85%	88%[†]

Hyperparameters: filt2.filter.nfeat= 421.8581; k= 1.550807
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.6 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue collected at 21 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	195	20	12	6	6	23	74%
Enlist 2	20	225	1	12	10	9	81%
No 2,4-D	11	2	258	6	6	8	89%
Unison	13	10	6	236	9	17	81%
Weedar	6	13	7	0	230	8	87%
Weedone	19	18	1	28	6	226	76%
Sensitivity	74%	78%	91%	82%	86%	78%	81%[†]

Hyperparameters: filt2.filter.nfeat= 261.7887; k= 3.748688
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.7 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue collected at 28 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	215	12	6	34	10	17	73%
Enlist 2	12	225	11	8	6	6	84%
No 2,4-D	0	16	219	0	18	10	83%
Unison	37	8	13	217	28	11	69%
Weedar	6	8	12	9	220	15	81%
Weedone	6	19	12	8	9	226	81%
Sensitivity	78%	78%	80%	79%	76%	79%	78%[†]

Hyperparameters: filt2.filter.nfeat= 358.0201; k= 5.32356
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.8 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	385	18	12	5	10	14	87%
Enlist 2	13	411	11	10	9	16	87%
No 2,4-D	9	9	418	25	13	4	87%
Unison	26	15	16	413	11	13	84%
Weedar	20	9	14	11	428	19	85%
Weedone	6	18	9	16	9	414	88%
Sensitivity	84%	86%	87%	86%	89%	86%	86%[†]

Hyperparameters: filt2.filter.nfeat= 316.6248; k= 4.372543
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.9 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	
Enlist 1	306	11	8	14	10	21	83%
Enlist 2	23	332	4	12	6	21	83%
No 2,4-D	11	11	357	12	0	5	90%
Unison	13	8	7	305	28	11	82%
Weedar	5	6	4	21	311	5	88%
Weedone	14	16	4	17	5	321	85%
	82%	86%	93%	80%	86%	84%	85%[†]

Hyperparameters: filt2.filter.nfeat= 205.0934; k= 2.54763
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.10 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	256	21	16	23	17	10	75%
Enlist 2	23	308	11	19	16	23	77%
No 2,4-D	15	3	293	13	34	7	80%
Unison	28	22	15	298	10	10	78%
Weedar	13	12	13	11	280	10	83%
Weedone	10	18	12	8	15	324	84%
Sensitivity	74%	80%	81%	80%	75%	84%	79%[†]

Hyperparameters: filt2.filter.nfeat= 210.1004; k= 2.64406
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.11 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 7 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	79	1	2	2	4	0	90%
Enlist 2	1	77	6	5	2	5	80%
No 2,4-D	4	5	80	4	1	1	84%
Unison	3	6	8	80	3	2	78%
Weedar	9	5	0	2	78	2	81%
Weedone	0	2	0	3	8	86	87%
Sensitivity	82%	80%	83%	83%	81%	90%	83%[†]

Hyperparameters: filt2.filter.nfeat= 414.2801; k= 3.063691
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.12 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 7 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	68	3	8	11	2	0	74%
Enlist 2	0	81	4	2	1	5	87%
No 2,4-D	7	3	59	3	3	2	77%
Unison	18	5	6	78	6	0	69%
Weedar	3	2	9	1	87	3	83%
Weedone	0	2	7	1	0	86	90%
Sensitivity	71%	84%	63%	81%	88%	90%	80%[†]

Hyperparameters: filt2.filter.nfeat= 179.2818; k= 2.723594
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.13 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 7 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	77	0	0	3	0	6	90%
Enlist 2	9	85	0	1	3	3	84%
No 2,4-D	0	0	92	9	0	1	90%
Unison	3	7	4	74	6	0	79%
Weedar	4	0	0	4	63	2	86%
Weedone	3	4	0	2	0	84	90%
Sensitivity	80%	89%	96%	80%	88%	88%	87%[†]

Hyperparameters: filt2.filter.nfeat= 261.4763; k= 3.492143
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.14 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 14 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	83	0	0	0	0	0	100%
Enlist 2	0	90	0	4	2	0	94%
No 2,4-D	0	0	94	0	2	0	98%
Unison	0	2	0	86	0	0	98%
Weedar	0	4	2	5	92	0	89%
Weedone	1	0	0	1	0	96	98%
Sensitivity	99%	94%	98%	90%	96%	100%	96%[†]

Hyperparameters: filt2.filter.nfeat= 374.7596; k= 1.775486
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.15 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 14 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	74	0	5	0	3	2	88%
Enlist 2	4	82	0	6	3	10	78%
No 2,4-D	1	0	71	10	5	0	82%
Unison	0	5	13	73	2	0	78%
Weedar	3	1	4	7	86	6	80%
Weedone	2	8	0	0	0	78	89%
Sensitivity	88%	85%	76%	76%	87%	81%	82%[†]

Hyperparameters: filt2.filter.nfeat= 168.1977; k= 3.945832
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.16 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 14 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	83	0	0	0	1	1	98%
Enlist 2	1	90	0	0	0	3	96%
No 2,4-D	0	0	96	0	0	0	100%
Unison	0	0	0	96	3	0	97%
Weedar	0	0	0	0	89	0	100%
Weedone	0	6	0	0	3	92	91%
Sensitivity	99%	94%	100%	100%	93%	96%	97%[†]

Hyperparameters: filt2.filter.nfeat= 175.8271; k= 2.450246
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.17 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 21 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	81	8	3	1	2	3	83%
Enlist 2	0	85	1	0	0	3	96%
No 2,4-D	3	1	85	4	3	0	89%
Unison	0	2	0	91	0	0	98%
Weedar	0	0	4	0	91	4	92%
Weedone	3	0	3	0	0	86	93%
Sensitivity	93%	89%	89%	95%	95%	90%	92%[†]

Hyperparameters: filt2.filter.nfeat= 252.8226; k= 3.926834
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.18 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 21 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	64	12	3	1	0	4	76%
Enlist 2	13	64	3	5	1	2	73%
No 2,4-D	0	0	84	0	3	0	97%
Unison	0	10	2	81	10	10	72%
Weedar	2	8	1	3	61	1	80%
Weedone	2	2	0	6	0	82	89%
Sensitivity	79%	67%	90%	84%	81%	83%	81%[†]

Hyperparameters: filt2.filter.nfeat= 348.6883; k= 1.377049
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.19 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 21 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	75	5	1	1	6	7	79%
Enlist 2	0	81	0	3	0	2	94%
No 2,4-D	0	0	93	3	0	0	97%
Unison	1	3	2	78	0	7	86%
Weedar	4	1	0	0	89	3	92%
Weedone	16	6	0	11	1	77	69%
Sensitivity	78%	84%	97%	81%	93%	80%	86%[†]

Hyperparameters: filt2.filter.nfeat= 234.7751; k= 4.999414
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.20 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 28 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	96	0	3	0	0	3	94%
Enlist 2	0	80	3	0	0	5	91%
No 2,4-D	0	5	86	0	3	0	91%
Unison	0	0	1	96	0	2	97%
Weedar	0	1	3	0	93	8	89%
Weedone	0	10	0	0	0	78	89%
Sensitivity	100%	83%	90%	100%	97%	81%	92%[†]

Hyperparameters: filt2.filter.nfeat= 415.3922; k= 5.491673
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.21 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 28 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	50	4	3	10	5	5	65%
Enlist 2	7	80	0	0	4	1	87%
No 2,4-D	0	0	74	0	10	0	88%
Unison	23	8	3	70	9	7	58%
Weedar	2	0	0	3	65	3	89%
Weedone	2	4	1	1	6	77	85%
Sensitivity	60%	83%	91%	83%	66%	83%	77%[†]

Hyperparameters: filt2.filter.nfeat= 210.4421; k= 2.281379
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.22 Confusion matrix generated from the classification model using KNN and transformed cotton spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 28 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	80	4	0	4	5	12	76%
Enlist 2	7	79	6	1	3	0	82%
No 2,4-D	0	10	82	1	0	0	88%
Unison	4	0	3	77	17	3	74%
Weedar	3	3	4	10	71	1	77%
Weedone	2	0	1	3	0	80	93%
Sensitivity	83%	82%	85%	80%	74%	83%	81%[†]

Hyperparameters: filt2.filter.nfeat= 357.5847; k= 2.685309
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.23 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	851	107	71	76	82	73	68%
Enlist 2	93	826	93	75	72	65	67%
No 2,4-D	100	103	830	85	63	100	65%
Unison	81	73	101	883	99	64	68%
Weedar	66	72	57	68	809	85	70%
Weedone	57	70	90	61	81	846	70%
Sensitivity	68%	66%	67%	71%	67%	69%	68%[†]

Hyperparameters: filt2.filter.nfeat= 307.4615; k= 4.810381
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.24 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue collected at 0 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	84	0	12	4	17	1	71%
Enlist 2	0	75	8	11	0	3	77%
No 2,4-D	2	2	50	1	7	6	74%
Unison	2	16	6	79	13	6	65%
Weedar	7	3	9	1	46	4	66%
Weedone	1	0	11	0	13	76	75%
Sensitivity	88%	78%	52%	82%	48%	79%	71%[†]

Hyperparameters: filt2.filter.nfeat= 214.7996; k= 2.465622
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.25 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue collected at 7 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	200	30	11	12	28	18	67%
Enlist 2	27	186	13	8	15	13	71%
No 2,4-D	24	24	197	20	15	26	64%
Unison	3	6	20	222	21	8	79%
Weedar	19	24	21	17	197	24	65%
Weedone	15	21	26	9	9	196	71%
Sensitivity	69%	64%	68%	77%	69%	69%	69%[†]

Hyperparameters: filt2.filter.nfeat= 235.2964; k= 3.732914
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.26 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue collected at 14 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	204	43	20	22	21	15	63%
Enlist 2	32	187	20	16	10	23	65%
No 2,4-D	13	18	180	14	18	27	67%
Unison	23	13	17	208	23	5	72%
Weedar	8	13	14	21	204	19	73%
Weedone	8	14	31	7	12	187	72%
Sensitivity	71%	65%	64%	72%	71%	68%	68%[†]

Hyperparameters: filt2.filter.nfeat= 292.8866; k= 4.183178
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.27 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue collected at 21 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	225	18	10	12	8	3	82%
Enlist 2	32	206	5	15	24	9	71%
No 2,4-D	9	13	231	21	13	13	77%
Unison	13	24	20	219	7	9	75%
Weedar	8	17	6	6	179	6	81%
Weedone	1	10	16	15	18	248	81%
Sensitivity	78%	72%	80%	76%	72%	86%	77%[†]

Hyperparameters: filt2.filter.nfeat= 247.7708; k= 5.228313
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.28 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue collected at 28 DAT and treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over tissue treatment status (Treated, New Growth, and Composite).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	201	14	17	14	13	20	72%
Enlist 2	24	193	29	13	14	8	69%
No 2,4-D	16	42	189	10	6	29	65%
Unison	11	15	27	200	13	18	70%
Weedar	14	13	1	19	235	13	80%
Weedone	22	11	25	32	7	200	67%
Sensitivity	70%	67%	66%	69%	82%	69%	70%[†]

Hyperparameters: filt2.filter.nfeat= 275.2221; k= 2.435127
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.29 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	378	30	23	25	40	8	75%
Enlist 2	20	402	12	24	17	15	82%
No 2,4-D	27	7	377	16	16	15	82%
Unison	26	23	19	396	27	10	79%
Weedar	17	7	18	11	340	40	79%
Weedone	12	8	25	8	31	392	82%
Sensitivity	79%	84%	80%	82%	72%	82%	80%[†]

Hyperparameters: filt2.filter.nfeat= 318.1033; k= 2.954399
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.30 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	311	23	18	16	17	9	79%
Enlist 2	24	275	20	16	19	20	74%
No 2,4-D	12	30	267	19	14	20	74%
Unison	10	12	17	283	16	14	80%
Weedar	18	21	20	21	280	15	75%
Weedone	9	23	42	29	20	291	70%
Sensitivity	81%	72%	70%	74%	77%	79%	75%[†]

Hyperparameters: filt2.filter.nfeat= 272.4868; k= 2.796651
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.31 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	330	50	11	28	27	27	70%
Enlist 2	21	281	14	8	10	12	81%
No 2,4-D	1	14	266	33	26	21	74%
Unison	9	9	42	287	15	17	76%
Weedar	11	19	25	18	259	26	72%
Weedone	12	17	26	10	32	281	74%
Sensitivity	86%	72%	69%	75%	70%	73%	74%[†]

Hyperparameters: filt2.filter.nfeat= 251.5531; k= 4.005087
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.32 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 7 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	64	14	6	0	12	2	65%
Enlist 2	6	79	5	0	5	3	81%
No 2,4-D	9	2	80	0	7	5	78%
Unison	6	1	0	90	4	3	87%
Weedar	11	0	0	3	63	2	80%
Weedone	0	0	5	3	8	81	84%
Sensitivity	67%	82%	83%	94%	64%	84%	79%[†]

Hyperparameters: filt2.filter.nfeat= 305.7559; k= 5.311139
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.33 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 7 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	85	4	0	0	3	7	86%
Enlist 2	4	69	4	0	3	3	83%
No 2,4-D	0	10	71	9	4	7	70%
Unison	0	2	8	84	3	0	87%
Weedar	5	9	7	0	77	8	73%
Weedone	2	5	6	3	3	71	79%
Sensitivity	89%	70%	74%	88%	83%	74%	79%[†]

Hyperparameters: filt2.filter.nfeat= 316.0921; k= 4.982105
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.34 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 7 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	77	0	7	8	2	4	79%
Enlist 2	2	63	8	1	1	0	84%
No 2,4-D	9	18	68	3	4	8	62%
Unison	3	4	2	68	7	0	81%
Weedar	5	4	6	14	73	6	68%
Weedone	0	7	5	2	6	75	79%
Sensitivity	80%	66%	71%	71%	78%	81%	74%[†]

Hyperparameters: filt2.filter.nfeat= 174.713; k= 5.187904
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.35 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 14 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	85	3	6	12	2	3	77%
Enlist 2	0	73	0	3	1	4	90%
No 2,4-D	0	2	79	0	6	0	91%
Unison	11	2	0	77	6	1	79%
Weedar	0	11	5	4	75	3	77%
Weedone	0	5	0	0	6	85	89%
Sensitivity	89%	76%	88%	80%	78%	89%	83%[†]

Hyperparameters: filt2.filter.nfeat= 295.7347; k= 1.800445
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.36 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 14 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	66	23	6	13	16	9	50%
Enlist 2	15	69	0	8	4	2	70%
No 2,4-D	2	2	83	9	0	2	85%
Unison	3	1	5	57	6	0	79%
Weedar	4	0	0	9	69	10	75%
Weedone	6	1	2	0	1	73	88%
Sensitivity	69%	72%	86%	59%	72%	76%	72%[†]

Hyperparameters: filt2.filter.nfeat= 240.3397; k= 3.235399
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.37 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 14 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	74	8	4	11	0	2	75%
Enlist 2	14	69	3	5	5	9	66%
No 2,4-D	8	9	68	0	6	7	69%
Unison	0	7	2	79	0	2	88%
Weedar	0	0	4	1	76	11	83%
Weedone	0	3	15	0	9	53	66%
Sensitivity	77%	72%	71%	82%	79%	63%	74%[†]

Hyperparameters: filt2.filter.nfeat= 314.8957; k= 4.029961
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.38 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 21 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	77	1	0	1	0	0	97%
Enlist 2	7	90	0	0	3	1	89%
No 2,4-D	1	0	85	9	0	0	89%
Unison	8	0	11	81	1	0	80%
Weedar	1	0	0	0	75	6	91%
Weedone	2	2	0	5	5	89	86%
Sensitivity	80%	97%	89%	84%	89%	93%	89%[†]

Hyperparameters: filt2.filter.nfeat= 249.2773; k= 4.08888
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.39 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 21 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	75	6	1	0	0	0	91%
Enlist 2	15	76	3	2	1	3	76%
No 2,4-D	3	3	77	0	9	3	81%
Unison	0	1	2	88	0	0	97%
Weedar	0	9	5	1	65	6	76%
Weedone	3	4	8	5	9	84	74%
Sensitivity	78%	77%	80%	92%	77%	88%	82%[†]

Hyperparameters: filt2.filter.nfeat= 195.3071; k= 4.066186
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.40 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 21 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	85	8	2	1	4	0	85%
Enlist 2	2	72	0	3	3	12	78%
No 2,4-D	3	7	73	12	2	5	72%
Unison	0	2	6	74	1	3	86%
Weedar	6	2	3	0	71	0	87%
Weedone	0	5	12	6	0	76	77%
Sensitivity	89%	75%	76%	77%	88%	79%	80%[†]

Hyperparameters: filt2.filter.nfeat= 204.5759; k= 3.212075
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.41 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue directly treated with No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 28 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	86	3	0	0	0	0	97%
Enlist 2	5	86	0	0	0	1	93%
No 2,4-D	0	0	86	2	0	1	97%
Unison	5	4	3	90	0	0	88%
Weedar	0	3	0	3	96	7	88%
Weedone	0	0	7	1	0	87	92%
Sensitivity	90%	90%	90%	94%	100%	91%	92%[†]

Hyperparameters: filt2.filter.nfeat= 328.9369; k= 3.498013
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.42 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue both directly treated with and grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 28 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	80	3	1	0	6	4	85%
Enlist 2	1	76	5	1	3	4	84%
No 2,4-D	1	8	60	7	0	15	66%
Unison	3	5	17	79	4	6	69%
Weedar	8	0	2	5	77	4	80%
Weedone	3	4	11	4	6	63	69%
Sensitivity	83%	79%	62%	82%	80%	66%	76%[†]

Hyperparameters: filt2.filter.nfeat= 315.2236; k= 4.820841
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Table 3.43 Confusion matrix generated from the classification model using KNN and transformed soybean spectra produced from tissue that grew post-treatment of No 2,4-D, Unison, Weedar, Weedone, Enlist 1, and Enlist 2 and collected 28 DAT.[a,b]

Prediction	Enlist 1	Enlist 2	No 2,4-D	Unison	Weedar	Weedone	PPV
Enlist 1	79	0	9	3	3	4	81%
Enlist 2	4	73	13	0	2	4	76%
No 2,4-D	7	12	66	4	2	5	69%
Unison	2	0	1	75	4	5	86%
Weedar	1	8	1	2	85	2	86%
Weedone	3	3	6	12	0	76	76%
Sensitivity	82%	76%	69%	78%	89%	79%	79%[†]

Hyperparameters: filt2.filter.nfeat= 255.4232; k= 3.141905
[a]Abbreviations: No 2,4-D, non-treated control; Unison, 2,4-D acid; Weedar, 2,4-D dimethylamine salt; Weedone, 2,4-D ethylhexyl ester; Enlist 1, 2,4-D choline; Enlist 2, 2,4-D/glyphosate premixture; PPV, positive predictive value
[b]Transformed spectral data were scaled to a mean of zero and standard deviation of 1, and smoothed and derived using the Savitzky-Golay filter
[†]Overall accuracy of the classification model

Figures

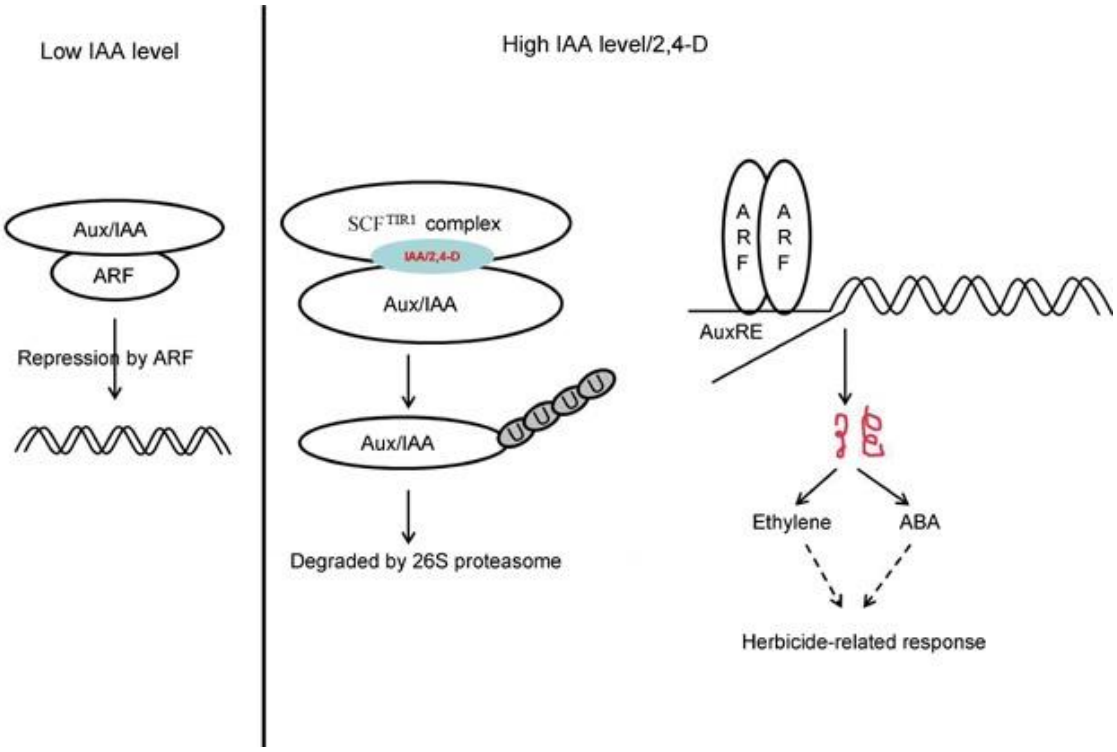

Figure 3.1 Proposed molecular mechanism of 2,4-D.[a,b]

[a]Abbreviations: IAA, Indole-3-acetic acid; Aux/IAA, Auxin/IAA active repressor protein complex; ARF, Auxin Response Factor; SCF^{TIR1}, Skp-Cullin,-F-box TRANSPORT INHIBITOR RESISTANT1 complex

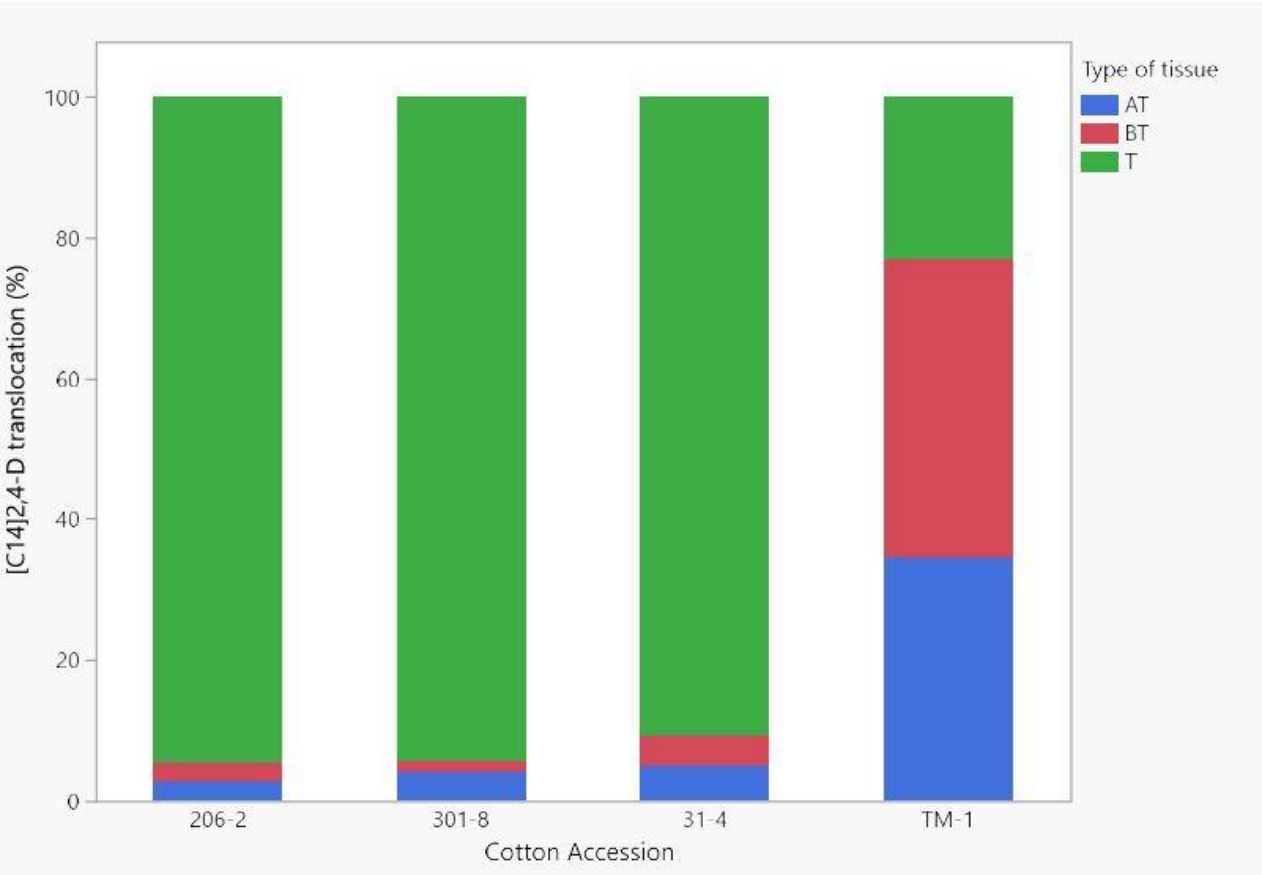

Figure 3.2 Translocation of ^{14}C-2,4-D on cotton seedlings of CS-B15sh (31-4), CS-T07 (206-2), CS-T04-15 (301-8), and TM-1 at 24 hours after treatment

Type of tissue: AT, above the treated leaf; BT, below the treated leaf; T, treated leaf.

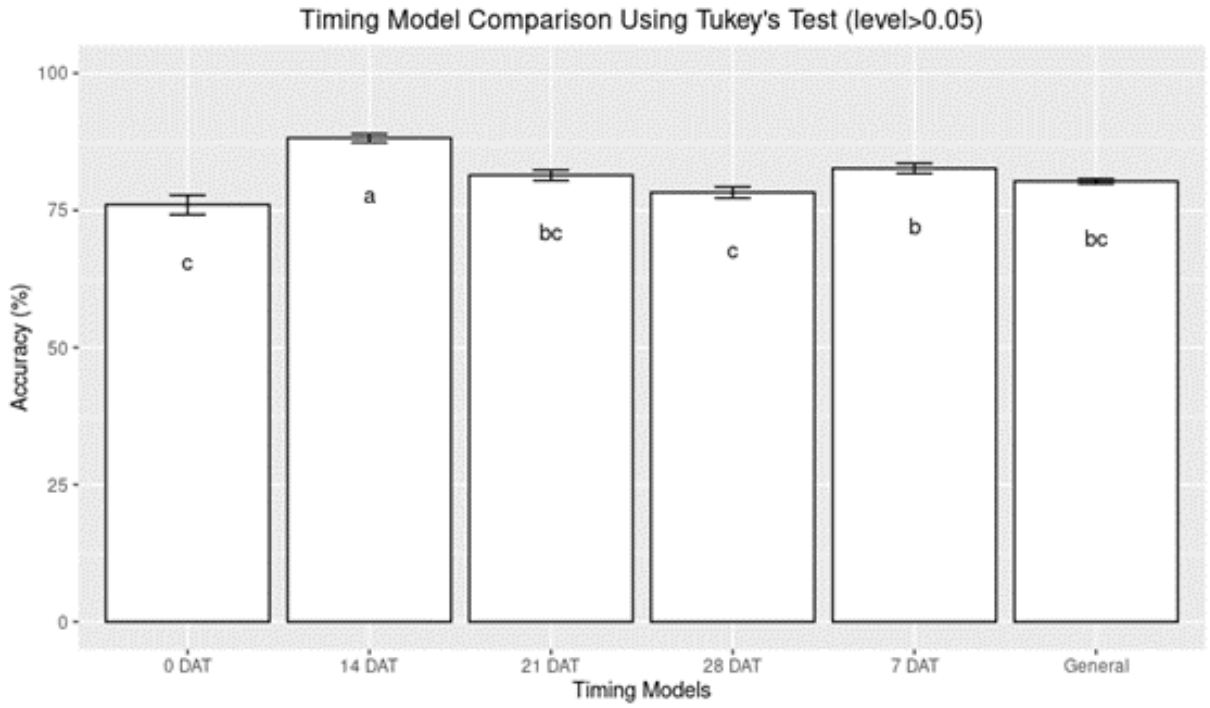

Figure 3.3 Comparison of supervised machine learning models utilizing Tukey's honest significance difference test with transformed cotton spectra pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

[a]Abbreviations: DAT; days after treatment; 0 DAT; classification model built using tissue samples collected 0 DAT; 7 DAT, classification model built using tissue samples collected 7 DAT; 14 DAT, classification model built using tissue samples collected 14 DAT; 21 DAT, classification model built using tissue samples collected 21 DAT; 28 DAT, classification model built using tissue samples collected 28 DAT; General, classification model built using all tissue samples collected.

[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

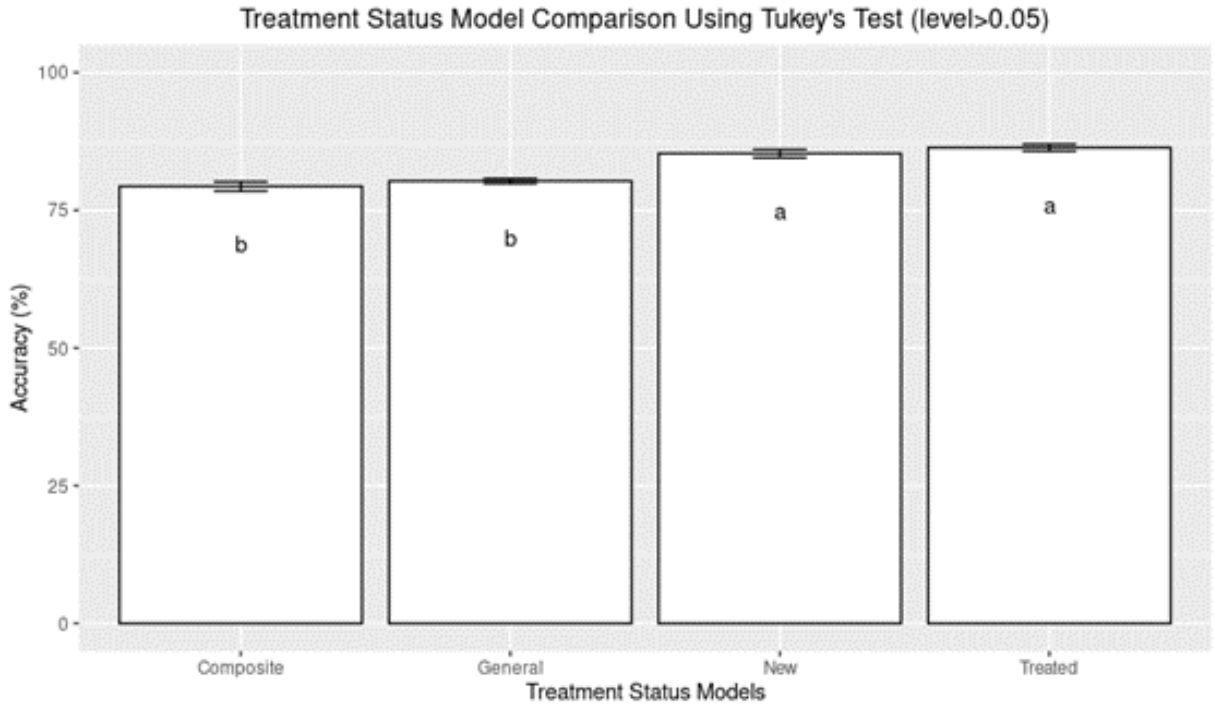

Figure 3.4 Comparison of supervised machine learning models utilizing Tukey's honest significance difference test with transformed cotton spectra pooled over treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: Composite; classification model built using tissue samples both directly treated with and which grew post-treatment of 2,4-D application; New, classification model built using tissue samples which grew post-treatment of 2,4-D application; Treated, classification model built using tissue samples directly treated with 2,4-D applications; General, classification model built using all tissue samples collected.
[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

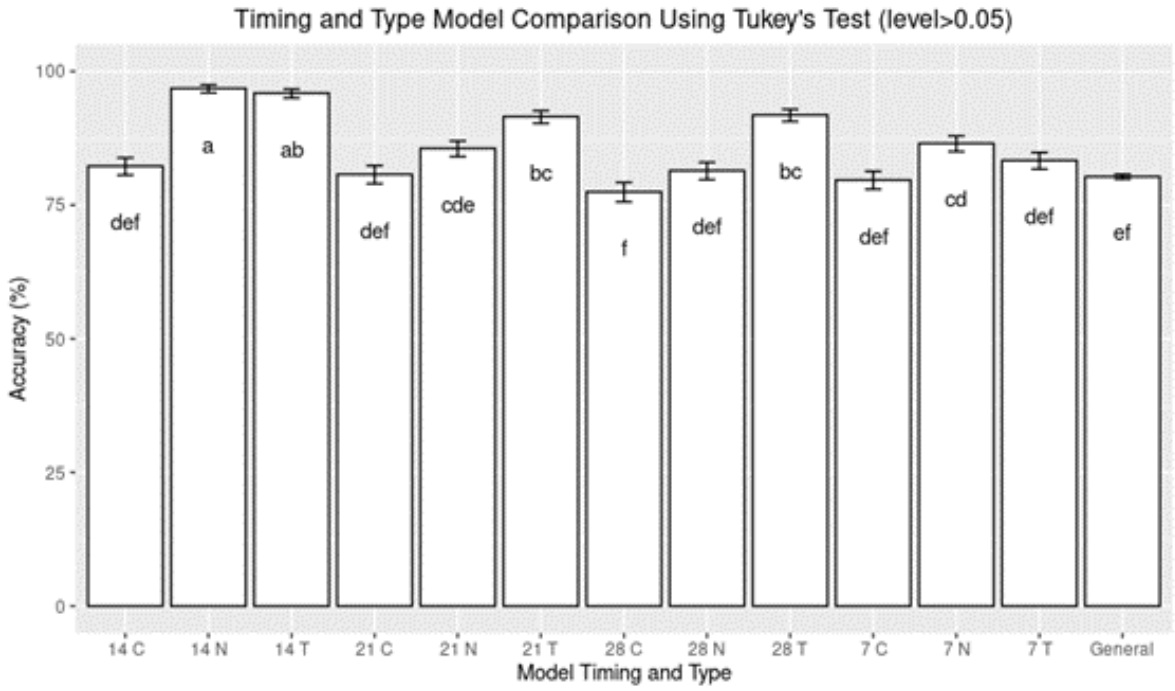

Figure 3.5 Comparison of supervised machine learning models utilizing Tukey's honest significance difference test with transformed cotton spectra pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: 14 C; 14 composite classification model; 14 N, 14 DAT New Growth classification model; 14 T, 14 DAT Treated classification model; 21 C, 21 DAT Composite classification model; 21 N, 21 DAT New Growth classification model; 21 T, 21 DAT Treated classification model; 28 C, 28 DAT Composite classification model; 28 N, 28 New Growth classification model; 28 T, 28 DAT Treated classification model; 7 C, 7 DAT Composite classification model; 7 N, 7 DAT New Growth classification model; 7 T; 7 DAT Treated classification model; General; classification model built using all tissue samples collected.
[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

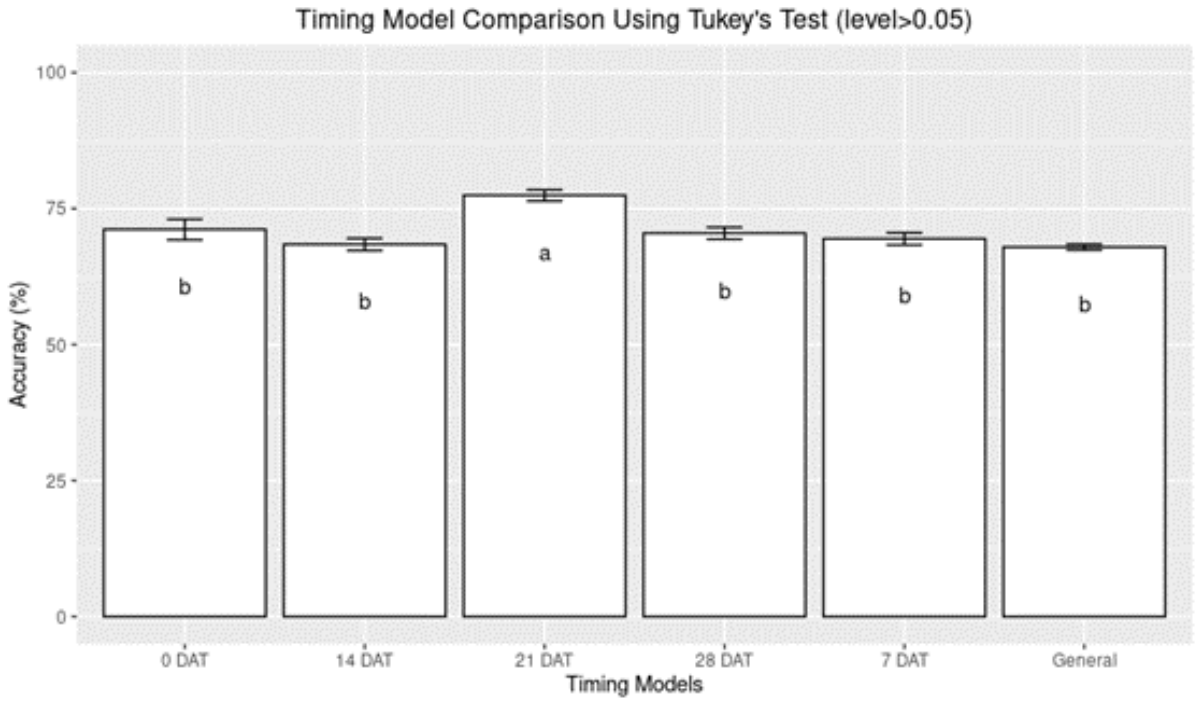

Figure 3.6 Comparison of supervised machine learning models utilizing Tukey's honest significance difference test with transformed soybean spectra pooled over sampling timing (0, 7, 14, 21, and 28 DAT).[a,b]

[a]Abbreviations: DAT; days after treatment; 0 DAT; classification model built using tissue samples collected 0 DAT; 7 DAT, classification model built using tissue samples collected 7 DAT; 14 DAT, classification model built using tissue samples collected 14 DAT; 21 DAT, classification model built using tissue samples collected 21 DAT; 28 DAT, classification model built using tissue samples collected 28 DAT; General, classification model built using all tissue samples collected.

[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

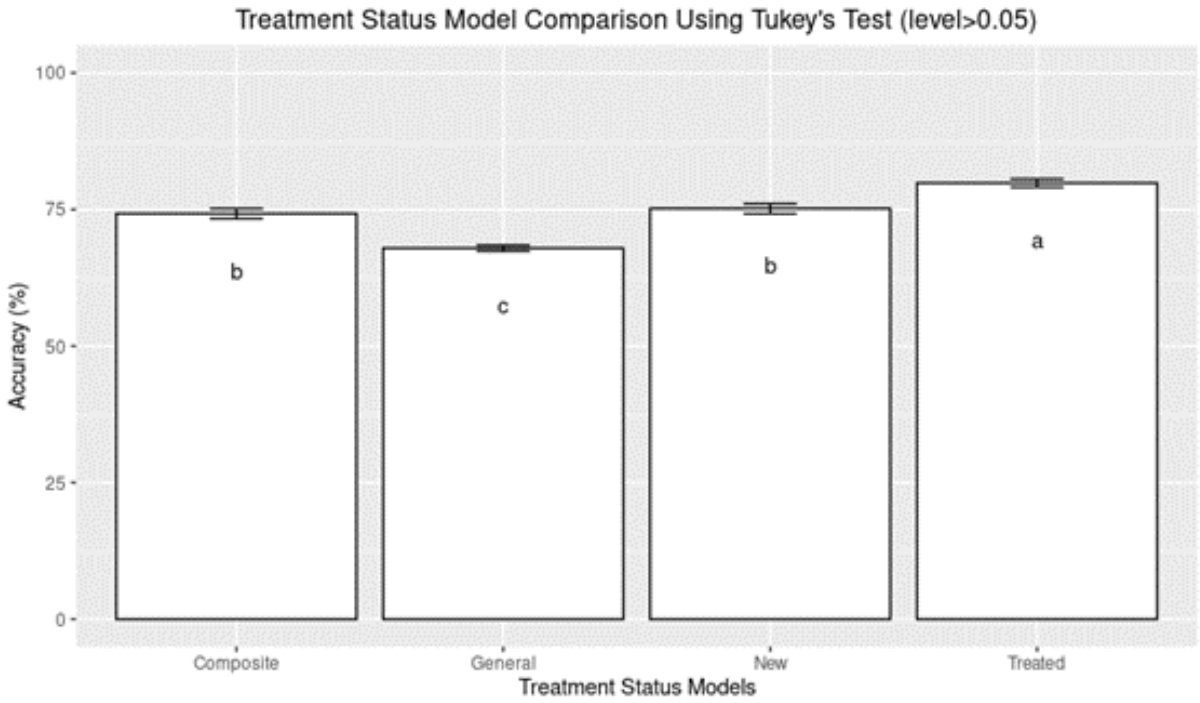

Figure 3.7 Comparison of supervised machine learning models utilizing Tukey's honest significance difference test with transformed soybean spectra pooled over treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: Composite; classification model built using tissue samples both directly treated with and which grew post-treatment of 2,4-D application; New, classification model built using tissue samples which grew post-treatment of 2,4-D application; Treated, classification model built using tissue samples directly treated with 2,4-D applications; General, classification model built using all tissue samples collected.
[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

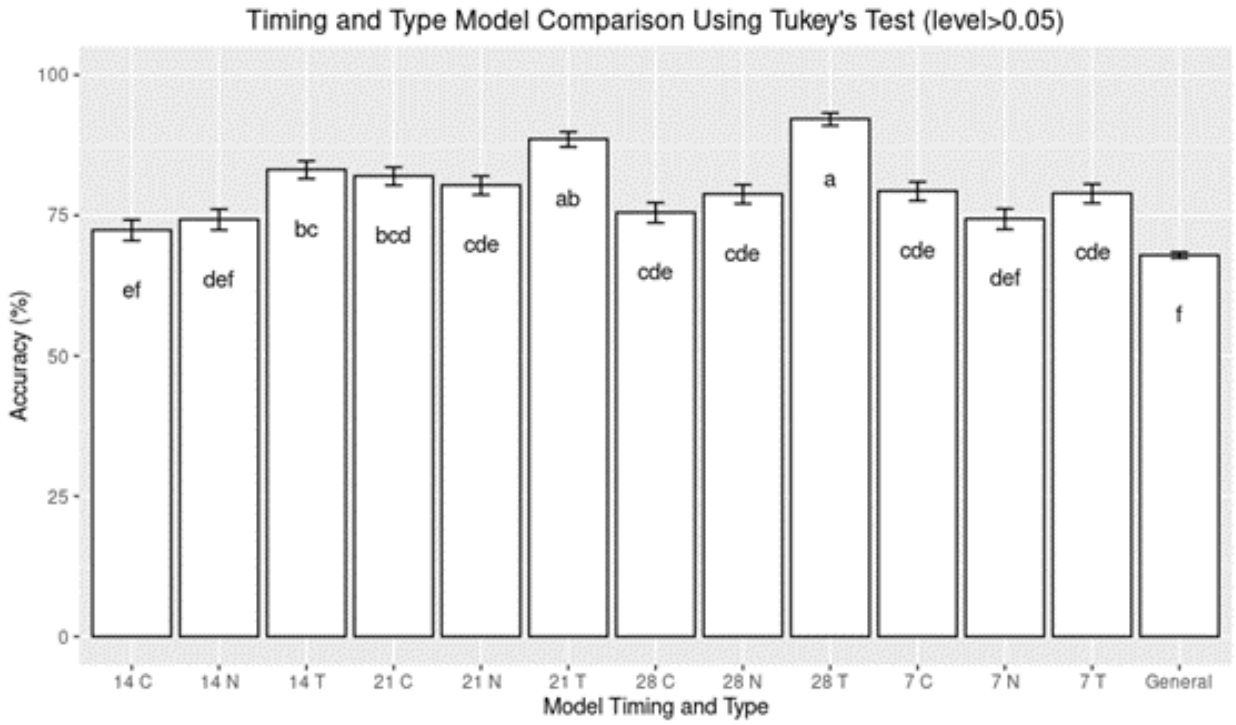

Figure 3.8 Comparison of supervised machine leanring models utilizing Tukey's honest significance difference test with transformed soybean spectra pooled over sampling timing (0, 7, 14, 21, and 28 DAT) and treatment status (Treated, New Growth, and Composite).[a,b]

[a]Abbreviations: 14 C; 14 composite classification model; 14 N, 14 DAT New Growth classification model; 14 T, 14 DAT Treated classification model; 21 C, 21 DAT Composite classification model; 21 N, 21 DAT New Growth classification model; 21 T, 21 DAT Treated classification model; 28 C, 28 DAT Composite classification model; 28 N, 28 New Growth classification model; 28 T, 28 DAT Treated classification model; 7 C, 7 DAT Composite classification model; 7 N, 7 DAT New Growth classification model; 7 T; 7 DAT Treated classification model; General; classification model built using all tissue samples collected.
[b] Lowercase letters indicate significant differences between treatments based on Tukey's honest significant difference ($p < 0.05$)

References

2,4-D Technical Fact Sheet. (n.d.). Retrieved April 9, 2020, from http://npic.orst.edu/factsheets/archive/2,4-DTech.html

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encyclopedia of Research Design*, *3*(1), 1–5.

Alazem, M., & Lin, N. S. (2017). Antiviral roles of abscisic acid in plants. *Frontiers in Plant Science*, *8*, 1–10. https://doi.org/10.3389/FPLS.2017.01760/BIBTEX

Andersen, S. M., Clay, S. A., Wrage, L. J., & Matthees, D. (2004). Soybean Foliage Residues of Dicamba and 2,4-D and Correlation to Application Rates and Yield. *Agronomy Journal*, *96*(3), 750–760. https://doi.org/10.2134/AGRONJ2004.0750

Brown, C., Meyers, S., Rose, M. A., & Doohan, D. (n.d.). *An Overview of Dicamba and 2,4-D Drift Issues*.

Buol, J. (2019). Stewarding 2,4-D- and dicamba- based weed control technologies in cotton and soybean production systems [Mississippi State University]. In *Theses and Dissertations*. https://scholarsjunction.msstate.edu/td/5049

Buol, J. T., Reynolds, D. B., Dodds, D. M., Mills, J. A., Nichols, R. L., Bond, J. A., Jenkins, J. N., & Dubien, J. L. (2019). The effect of cotton growth stage on response to a sublethal concentration of 2,4-D. *Weed Technology*, *33*(2), 321–328. https://doi.org/10.1017/WET.2019.9

Chang, C. (2016). Q and A: How do plants respond to ethylene and what is its importance? *BMC Biology*, *14*(1), 1–7. https://doi.org/10.1186/S12915-016-0230-0/FIGURES/4

Egan, J. F., Barlow, K. M., & Mortensen, D. A. (2014). A Meta-Analysis on the Effects of 2,4-D and Dicamba Drift on Soybean and Cotton. *Weed Science*, *62*(1), 193–206. https://doi.org/10.1614/WS-D-13-00025.1

Everitt, J. D., & Keeling, J. W. (2009). Cotton Growth and Yield Response to Simulated 2,4-D and Dicamba Drift. *Technology*, *23*(4), 503–506. https://doi.org/10.1614/WT-08-061.1

Fuente, R. D. la, Cancino, J., & Acuña, E. (2021). Comparison of machine learning methods for dry biomass estimation based on green logging residues chips. *International Journal of Forest Engineering*, *32*(2), 174–184. https://doi.org/10.1080/14942119.2021.1892415

Ghanashyam, C., & Jain, M. (2009). Role of auxin-responsive genes in biotic stress responses. *Plant Signaling & Behavior*, *4*(9), 846. https://doi.org/10.4161/PSB.4.9.9376

Grossmann, K. (2010). Auxin herbicides: current status of mechanism and mode of action. *Pest Management Science*, *66*(2), 113–120. https://doi.org/10.1002/PS.1860

Marple, M. E., Al-Khatib, K., & Peterson, D. E. (2008a). Cotton Injury and Yield as Affected by Simulated Drift of 2,4-D and Dicamba. *Weed Technology*, *22*(4), 609–614. https://doi.org/10.1614/WT-07-095.1

Marple, M. E., Al-Khatib, K., & Peterson, D. E. (2008b). Cotton Injury and Yield as Affected by Simulated Drift of 2,4-D and Dicamba. *Weed Technology*, *22*(4), 609–614. https://doi.org/10.1614/WT-07-095.1

Mockaitis, K., & Estelle, M. (2008). Auxin Receptors and Plant Development: A New Signaling Paradigm. *Annual Review of Cell and Developmental Biology*, *24*, 55–80. https://doi.org/10.1146/ANNUREV.CELLBIO.23.090506.123214

Mohseni-Moghadam, M., Wolfe, S., Dami, I., & Doohan, D. (2016). Response of Wine Grape Cultivars to Simulated Drift Rates of 2,4-D, Dicamba, and Glyphosate, and 2,4-D or Dicamba Plus Glyphosate. *Weed Technology*, *30*(3), 807–814. https://doi.org/10.1614/WT-D-15-00106.1

Oakley, G. R. (2021). *Soybean (Glycine max) response to multiple, sublethal exposures Soybean (Glycine max) response to multiple, sublethal exposures of 2,4-D and dicamba from vegetative through reproductive of 2,4-D and dicamba from vegetative through reproductive growth growth* [Mississippi State University]. https://scholarsjunction.msstate.edu/td

Pazmiño, D. M., Rodríguez-Serrano, M., Romero-Puertas, M. C., Archilla-Ruiz, A., del Río, L. A., & Sandalio, L. M. (2011). Differential response of young and adult leaves to herbicide 2,4-dichlorophenoxyacetic acid in pea plants: role of reactive oxygen species. *Plant, Cell & Environment*, *34*(11), 1874–1889. https://doi.org/10.1111/J.1365-3040.2011.02383.X

Perez, L. M., Yue, Z., Saha, S., Dean, J. F. D., Jenkins, J. N., Stelly, D. M., & Tseng, T. M. (2021). Absorption and Translocation of [14c]2,4-Dichlorophenoxyacetic Scid in Herbicide-tolerant Chromosome Substitution Lines of Gossypium Hirsutum L. *Preprints*. https://doi.org/10.20944/PREPRINTS202109.0395.V1

Peterson, G. E. (1967). The Discovery and Development of 2,4-D. In *History* (Vol. 41, Issue 3). https://about.jstor.org/terms

Peterson, M. A., Mcmaster, S. A., Riechers, D. E., Skelton, J., & Stahlman, P. W. (2016). 2,4-D Past, Present, and Future: A Review. *Weed Technology*, *30*, 303–345. https://doi.org/10.1614/WT-D-15-00131.1

Reid, C. (2017). Monitoring Aspergillus Flavus Progression and Aflatoxin Accumulation in Inoculated Maize (Zea Mays L.) Hybrids [Mississippi State University]. In *Theses and Dissertations*. https://scholarsjunction.msstate.edu/td/3195

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Russell V. Lenth (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.2. https://CRAN.R-project.org/package=emmeans

Skelton, J. J., Simpson, D. M., Peterson, M. A., & Riechers, D. E. (2017). Biokinetic Analysis and Metabolic Fate of 2,4-D in 2,4-D-Resistant Soybean (Glycine max). *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.7b00796

Song, Y. (2014). Insight into the mode of action of 2,4-dichlorophenoxyacetic acid (2,4-D) as an herbicide. *Journal of Integrative Plant Biology*, *56*(2), 106–113. https://doi.org/10.1111/JIPB.12131

Tan, X., Calderon-Villalobos, L. I. A., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M., & Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*, *446*(7136), 640–645. https://doi.org/10.1038/nature05731

Tiwari, S. B., Wang, X.-J., Hagen, G., & Guilfoyle, T. J. (2001). AUX/IAA Proteins Are Active Repressors, and Their Stability and Activity Are Modulated by Auxin. *The Plant Cell*, *13*(12), 2809. https://doi.org/10.1105/TPC.010289

Tu, M., Hurd, C., Randall, J. M., & Rice, B. (2001). Weed Control Methods Handbook: Tools and Techniques for Use in Natural Areas. *The Nature Conservancy*, 87–96.

Zhang, Yonghong, Ge, T., Tian, W., & Liou, Y. A. (2019). Debris Flow Susceptibility Mapping Using Machine-Learning Techniques in Shigatse Area, China. *Remote Sensing*, *11*(23), 2801. https://doi.org/10.3390/RS1123280