DIGITAL PHENOTYPING IN COTTON BREEDING USING GROWTH RATE MODELING BASED ON VISIBLE LIGHT DATA COLLECTED WITH UNMANNED AERIAL SYSTEMS

A Dissertation

by

WILLIAM GEORGE DODGE

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jane Dever
Co-Chair of Committee,	Steve Hague
Committee Members,	Mauricio Ulloa
	Murilo Maeda
	Dana Porter
Head of Department,	David Baltensperger

May 2023

Major Subject: Plant Breeding

ABSTRACT

The need to improve cotton to the benefit of humanity has been the driving factor in scientific cotton breeding for nearly 150 years. The act of cotton breeding itself is not a novel enterprise. What is novel, however, are the tools and technologies breeders utilize in cotton breeding operations. The semi-arid Texas High Plains agricultural area is one of the largest cotton producing regions in the world. This region is supported agriculturally by the ever-shrinking Ogallala aquifer. Identifying cotton varieties that perform or demonstrate yield stability across a range of water limited environments is critical to the long-term sustainability of cotton production on the Texas High Plains. Contemporary remote sensing devices like drones provide a pathway to evaluate cotton growth in a manner that is novel, efficient, and high throughput. The significance of this approach with respect to evaluation of cotton growth on the Texas High Plains is that information can be gathered over many breeding plots at high temporal granularity in a quantified and consistent manner. Observation of this type provides researchers with insight into growth patterns and characteristics not previously evident with manual plot-by-plot field observations. This presents the possibility that growth patterns and characteristics in breeding lines might be identified, when observed over a range of water levels in the semi-arid Texas High Plains, which provide a basis to isolate genotypes that will allow cotton production in this region to remain stable in the coming decades.

DEDICATION

This work is dedicated to my father, George E. Dodge.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Dever, and my committee members, Dr. Ulloa, Dr. Hague, Dr. Maeda, and Dr. Porter, for their guidance throughout the course of this research.

Special thanks to Dr. Paxton Payton and Dr. James Mahan for their willingness to mentor and teach me along the way.

Mr. Andrew Young deserves recognition for his efforts collecting and managing data supporting this project.

I would also like to thank my wife, Martha, for her encouraging words and steadfast support.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Jane Dever, Steve Hague, and Murilo Maeda of the Department of Soil and Crop Sciences; Dr. Dana Porter of the Department of Biological and Agricultural Engineering; and Dr. Muricio Ulloa of the USDA-ARS.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported though a graduate research fellowship from the National Science Foundation Graduate Research Fellowship Program.

NOMENCLATURE

GSD	Ground Sample Distance
GSA	Ground Sample Area
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
ExG	Excessive Greenness Index
TGI	Triangular Greenness Index
VARI	Visible Atmospherically Resistant Index
RGBVI	Red Green Blue Vegetative Index
MGRVI	Modified Red Green Index
GNORM	Green Normalized Layer
CVol	Canopy Volume
CC	Canopy Cover
CV	Coefficient of Variation
VI	Vegetative Index
UAS	Unmanned Aerial Systems
NDVI	Normalized Differential Vegetative Index
DAP	Days After Planting
DOY	Day of Year
RGB	Red Green Blue
THP	Texas High Plains

RCBD	Randomized Complete Block Design
ET_0	Reference Grass Specific Evapotranspiration
ET_C	Crop Specific Evapotranspiration
CORS	Continually Operating Reference Stations
GNSS	Global Navigation Satellite System
HVI	High Volume Instrument

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Background

Forty-one counties comprise the Texas High Plains (THP) region extending from the Oklahoma state boundary in the Northern Texas panhandle Southward along the New Mexico state line and ending at a horizontal line intersecting the Southeastern corner of New Mexico. The region covers an approximate 10,000 km^2 of land surface area (Lascano et al., 2020). The history of agricultural production in the region runs parallel to the development of groundwater-based irrigation beginning in the 1920s. The source of groundwater for THP agricultural production is almost exclusively the Ogallala aquifer. A vast aquifer spanning eight states, the Ogallala is the largest unconfined aquifer in the United States (Texas Water Devlopment Board, 2020). Annual production demands for water in the THP currently exceed the annual average recharge volume. As such, the portion of the aquifer under the THP is a diminishing resource where the estimated average distance to water table has increased about 30 cm per year (McGuire & Fischer, 2001).

Given the status of groundwater resources in the area, most production systems operate with an accumulating irrigation deficit over the growing season. Along with deficit irrigation, the number of non-irrigated production acres and the need for producers to explore irrigated alternatives in the THP are increasing (Chu et al., 2016). The trend away from fully irrigated cotton production to either deficit irrigation or partially irrigated production systems will continue in the coming decades because of the insufficient annual aquifer recharge in the area. Researchers and producers working

within the agricultural sector of the THP region are faced with the reality that cotton production with less water on an annual basis will become requisite to sustainable cotton production. As has been the case with many challenges facing cotton production over the last 150 years, the ability to manage the water-resource challenge facing the THP will come largely through breeding groups producing germplasm well-suited to specific environments capable of responding to continuously improving water and crop management practices.

1.2. Cotton Breeding

Upland cotton has been successfully modified over many decades through global breeding activities that follow the same general process involving a "systematic method of observation, data collection and statistical analysis of plant performance" (Dever, 2012). Examples of improvements include increased lint yield, higher fiber quality, insect resistance, disease resistance, and desirable changes to leaf morphology (Zeng et al., 2018). Such genetic improvements occurred because of continual gain in overall understanding of genetics and gene action in cotton. Early cotton breeding success came during a period when genetic knowledge in breeding was largely limited to Mendelian inheritance of qualitative traits (Bourland, 2019). Even in the early years of scientific cotton breeding and development, electronic tools were designed and utilized to assist breeders with plant phenotyping in the context of breeding (Johnson, 1939). From this early period of scientific cotton breeding there would be considerable advancements

over the next 80 years. The period from 1940 to 2020 would see massive gains in high throughput harvest equipment technologies (Eaton, 2003), fiber evaluation technologies (Kelly et al., 2015), biotechnologies (Hake, 2003), gene manipulation methodologies (Paterson & Smith, 1999), and in digital phenotyping technologies (Chawade et al., 2019; Pabuayon et al., 2019). The level at which cotton breeders have historically been able to advance cotton performance and quality is strongly related to the adequacy of the available technologies supporting such activities.

Observation and data collection have been a fundamental component in breeding for centuries. These observations vary dramatically with respect to specific plant properties or phenotypic characters but the objective, to isolate the most prominent instances of any specific plant trait, remains the same (Dever, 2012). Obtaining reliable plot level data at scale is one of the major aspects of breeding related field work. This essential process has been historically limited by the amount time and labor breeders have at their disposal each growing season.

1.3. Breeding Tools

Tools that provide means to reduce time and labor involved in any of the many coordinated data collection efforts conducted over breeding plots in a single season, are beneficial to breeders and, ultimately, genetic gain. During the first two decades of the 21st century agricultural science in general has been inundated with various proximal and remote sensing technologies (Khanal et al., 2020). Sophisticated sensor technologies

are now available in many forms at reasonable costs to researchers. Of currently available electronic tools, unmanned aerial systems (UAS) have become one of most widely evaluated and utilized remote sensing technologies in agriculture (Tsouros et al., 2019). These devices allow researchers to gather data over large areas of cropland with little time and labor. Although the act of collecting image data itself requires little effort there remains large gaps in both public and private sector agricultural operations regarding analytic output for drone based high throughput evaluations (Yang et al., 2020). Nevertheless, these systems provide a means to generate insight not previously possible with manual plot-by-plot field observations. Examples include non-destructive yield estimation (Ashapure et al., 2020), ground cover (Duan et al., 2016), maturity evaluations (Narayanan et al., 2019), and plant height information (Hu et al., 2018). UAS technologies appear to be establishing a level of necessity in large scale breeding operations. Complete implementation and utilization of UAS technology across the breeding industry is not occurring currently but such an outcome would not be surprising given the current state of research around UAS based remote sensing.

1.4. UAS Technologies

As it stands today, there are several published experiments focused on the use UAS technologies to evaluate crop growth and biomass accumulation over a growing season. These methods for monitoring crop biomass accumulation are conducted via the analysis of multispectral data in the form of some type of vegetative index or through the

use of a three-dimensional representation of the plant derived from LiDAR or through the process of photogrammetric scene reconstruction (Wang et al., 2021). Regarding the use of these processes to monitor the growth of cotton in a breeding context, the number of available published experiments is quite small. Work by (Chu et al., 2016) appears to be the first such published experimentation covering the use of UAS derived imagery to evaluate cotton growth. In this work 35 cotton varieties were evaluated using plant height estimates derived from a series of digital surface models (DSM) using the formula $DSM_i - DSM_0$ where the difference between the bare soil DSM and a DSM collected at some point in the growth period of the crop provided an estimate of plant height. Along with the estimation of plant height values, this work assessed a parameter related to leaf area and biomass called canopy cover. This value is a two-dimensional measure of crop canopy derived from a segmentation process that separated plant pixels from the rest of the scene in an individual plot. These methods for evaluating crop growth with UAS are fundamental and continue to be utilized in both public and private sector crop research.

1.5. UAS Based Growth Modelling

Building on the early drone-based growth modeling research in cotton, further work explored the use of vegetative indices (VI) coupled with kinematics and second derivative parameters of VI derived sigmoid-type plant based growth curves (Yeom et al., 2017). This work is critical in that the focus extends beyond the observed growth and biomass accumulation to the rate curve and specific parameters of that curve. In the end,

the utility in this form of analysis is most evident when the operator can reduce a season's worth of drone observations to a few numerical parameters that provide insight regarding specific genotypes. Further work covering whole-season growth monitoring in cotton focused on comparisons between RGB and multispectral canopy cover (CC) (Ashapure et al., 2019). Canopy cover values provide insight into growth rates during the period of growth that occurs before complete canopy closure between rows. The significance of this work is that it demonstrated strong agreement between RGB derived CC values and those produced via a multispectral sensor. This provides a basis for using RGB sensors to monitor growth via CC over a comparable multispectral device. Given that RGB drone systems are typically much cheaper than a multispectral option, the findings in this work are beneficial to the global agricultural drone community.

1.6. Goals & Objectives

At present there exists a range of opportunities relating to aspects of UAS growth assessments in cotton that remain unexplored; specifically, when curve kinematics are explored relating to rates of biomass accumulation. The research herein focuses on the use of photogrammetric scene reconstruction to estimate biomass via drone derived growth proxies such as canopy volume and canopy cover. The rates of growth for the given varieties in this research were subjected to three water treatments in the semi-arid region of the High Plains of Texas. The focus of the work is to use contemporary methods and novel approaches to evaluate growth rates in cotton over a range of water-

deficit-stress levels with the expectation that these novel approaches will allow for better understanding of cotton growth under water deficit stress variability. Through better understanding of how growth rate values relate to end of season outcomes of interest, better selections can be identified.

The goal of this project is to develop a drone-based method for characterization of cotton growth rates and patterns to further genetic improvement of cotton breeding lines for the Texas High Plains. The components of this study supporting that goal are described in the following phases:

1. Identify a group of genetic material with sufficiently variable growth rates and habits suitable for observation and growth curve derivation. Growth rate variability was identified through an assessment of cotton plant maturity, a phenotypic character described by the ratio of open bolls to total bolls at the point in the season where the check variety assessment is greater than 50 percent.
2. Collect high resolution RGB drone imagery over the trial with a targeted revisit interval of 10 days. The focus is to gather time series image data for both years that is robust enough to allow interpolated sigmoid curve fitting to be representative of actual plant growth within each plot.
3. Both digital surface model and orthomosaic raster layers were produced for all flights using photogrammetric scene reconstruction based on raw RGB image data collected over the breeding trial from each drone flight.

4. Generate a set of plot level extractions for all plots, for each flight, treatment, and year. The individual plot level extractions are compiled for all flights for both years and serve as the base data set from which interpolated growth curves are produced.
5. Each plot for each flight produces both an RGB plot level extraction and a DSM plot level extraction. The RGB raster data is used to generate each of six different visible light spectrum vegetative indices suitable for image segmentation and canopy cover estimates. The canopy volume value was derived from the DSM plot extraction and exported. These numbers serve as the observed data for curve fitting. Specifics of the indices and canopy volume formulas are described in Table 4.
6. Sigmoid curves are interpolated over the observed plot level vegetative indices canopy cover values as well as the canopy volume assessment. The generalized logistic function will be the specific sigmoid equation fit to the observed data. The resultant curve will serve as a proxy for biomass accumulation over time for each plot.
7. From the interpolated sigmoid growth curves, a derivative curve will be produced using the rigorous form of a derivative. The result of this will be a curve that describes biomass accumulation rates as described by each of six canopy cover values as well the volumetric canopy values.
8. Growth rates will be normalized using heat unit accumulation rates. The purpose of this final step will be to produce a single numeric assessment of plot growth from

a season-long drone data time course. This single numeric data point will be a digitally derived phenotypic character called “heat units to maximum growth rate”. The relationship between this novel digitally derived phenotypic character and end of season outcomes like seed cotton yield and fiber quality parameters will be evaluated.

The objectives are to 1) identify, and observe via high-frequency, high-resolution RGB drone data, a group of material with enough growth variability to understand the relationship between the timing of growth rates and yield and quality outcomes, 2) develop a routine to analyze many drone flights over the course of an entire growing season and reduce that data-dense, time-series into singular digitally derived phenotypic assessments and 3) evaluate and understand the relationships between the digitally derived phenotypic assessments, e.g. heat-units-to-maximum-growth-rate, and important yield and fiber quality outcomes as well as the effect of water deficit stress on such relationships.

2. MATERIALS & METHODS

2.1. Entry Selection

Ten cotton (*Gossypium hirsutum*) varieties were selected from breeding germplasm to be imaged via UAS over two seasons at a location in the Texas High Plains (THP) cotton production region. Lines were selected with the intent of generating variability relating to growth characteristics based on maturity ratings. Maturity ratings are traditionally assessed as the number of open bolls to total bolls (Table 1). Commercial, advanced, and intermediate entries were included in the 10 lines chosen for this work.

Agronomic metadata used in the pre-experiment selection process were assessed to identify entries. These data were collected in 2019 over the following three locations spanning the Texas High Plain production region: Halfway, TX; Lubbock, TX; and Lamesa, TX. In the 2019 Lamesa testing location, entries were subjected to both full and limited irrigation treatments. This background information was used in conjunction with other known growth characteristics of these entries to obtain breeding material in this experiment with diverse growth patterns and variability relating to growth characteristics based on maturity ratings. Beyond growth rate characteristics the only other limiting factor used to assemble the group of lines that would be used in this work was seed volume in the existing catalog of material.

2.2. Experimental Design

The trial was conducted at the USDA-ARS Plant Stress and Water Conservation Laboratory in Lubbock, TX, (latitude of 33° 35' 40.85" N and a longitude of 101° 54' 4.99" W) in 2020 and 2021. The soil type at this location is an Amarillo series (fine-loamy, mixed, superactive, thermic Aridic Paleustalfs) which is generally a sandy clay loam for the horizons that comprise the cotton rooting zone. A randomized complete block design was selected with 10 entries and four replications. Planting dates were 12 May 2020, and 06 June 2021. This layout was mirrored over three water treatments: non-irrigated, 50 percent ETc replacement and a fully irrigated treatment which sought to eliminate water deficit stress altogether as measured with canopy temperature observations. Weather data were collected with an onsite weather station.

All plots were planted as two-row plots at 6.096 m (20 ft) in length with 4.572 m (15 ft) of planted row. The seeding rate for all plots was targeted at ~13 plants per meter (4 plants per foot) and planted with a four-row cone planter. In 2020 the target planting depth was 3.81 cm (1.5 in) and in 2021 the target planting depth was 6.35 cm (2.5 in). The difference in planting depths between to two years was a product of moisture levels in the soil at the time of planting. All plots were stripper harvested and samples were pulled by hand from the seed cotton for fiber quality analysis.

The fully irrigated treatment was managed through continuous evaluation of the canopy temperature curve. The biological optimum temperature for cotton has been

shown to be 28 °C (Wanjura et al., 1990). Whenever the canopy temperature was above the biological optimum temperature for a period of 200 minutes or greater over consecutive days, 38.1 mm (1.5 in) of water was applied to the fully irrigated treatment. The 50% irrigation treatment was induced by only matching every other full irrigation treatment in an alternating manner. The water application volume was the same but the time interval between applications was doubled. In both years, in-situ canopy temperature sensors were used while in 2021 both canopy temperature sensors and capacitance-based soil moisture probes were used (Goanna Ag GoField™ Plus). All treatments in both years started with a full soil moisture profile as heavy irrigation was applied until emergence. In 2020, irrigation treatments had greater relevance and produced larger effects regarding yield because the environmental water deficit was greater than that of 2021. Using the method described above, the water treatments for rain-grown, 50% percent, and full irrigation produced irrigation totals of 66.04, 114.3, and 297.18 mm (2.6, 4.5, and 11.7 in) respectively. (The amount for the rain-grown treatment represents the initial early season application given to ensure emergence). In 2020, crop rainfall total during the growing season was 195.58 mm (7.7 in), bringing the respective irrigation totals for treatments to 261.62, 309.88, and 492.76 mm (10.3, 12.3, and 19.4 in). The calculated ET_c values are generated via the GoField Plus interface using NDVI derived K_c values generated by the proprietary GoSat system. This system uses a propriety adaption of the FAO56 Penman-Monteith model (Allen et al., 1998).

The total calculated crop water demand for 2020 was 570.23 mm (22.45 in). The actual percentage of total ET_c for the three treatments in 2020 was 45.8%, 54.8%, and 86.4% respectively. For 2021 the irrigation totals for the respective treatments were 117.35, 157.99, and 216.91 mm (4.62, 6.22, and 8.54 in). The seasonal rainfall in 2021 was 226.57 mm (8.92 in) bringing the treatment totals to 343.92, 384.56, and 443.48 mm (13.54, 15.14, and 17.46 in). The ET_c total for 2021 was 476.25 mm (18.75 in) meaning the respective actual treatment percentages were 72.2%, 80.74%, and 93.12 % of the seasonal crop water demand. The actual treatment percentages for 2021 were greater than 2020 and this outcome is evidenced by nearly doubled plot weights in 2021.

Plots were stripper harvested 05 November 2020, and 15 November 2021 with a two-row cotton harvester and seed cotton weights were recorded. Samples were taken from the harvested material in each plot and ginned for turnout to calculate yield. Fiber traits were measured for all entries in the experiment at Texas Tech Fiber and Biopolymer Research Institute using high volume instrumentation (HVI™) with two micronaire, two color, four length, and four strength sample measurements per protocol.

2.3. Drone Data Collection & Photogrammetry

Drone flights were conducted throughout the 2020 (Table 2) and 2021 (Table 3) growing seasons. The experiment was imaged with a 20-megapixel RGB sensor using a DJI Mavic 2 Pro (DJI Technology Co., Ltd., Shenzhen, China). A flight revisit interval of ten days was targeted for both seasons from planting to just before harvest.

Table 1 Maturity and yield information for selected entries from 2019. Entries were tested across three locations: Lubbock, Halfway, and Lamesa Texas with two water levels: Irrigated and Low Water. This information was used to assemble the group of entries that were included in the field trials for this research in 2020 and 2021. The testing locations span the Texas High Plains region. * Material not included in the 2019 field trials but included based on results from prior testing.

Genotype	*Breeding Stage*	*Halfway, TX % Open 15-Oct Irrigated*	*Halfway, TX Yield Rank Irrigated*	*Lubbock, TX % Open 14-Oct Irrigated*	*Lubbock, TX Yield Rank Irrigated*	*Lamesa, TX % Open 17-Sep Irrigated*	*Lamesa, TX Yield Rank Irrigated*	*Lamesa, TX % Open 12-Sep Low Water*	*Lamesa, TX Yield Rank Low Water*
FiberMax 958	Check variety	81	8	91	1	53	9	80	19
Stoneville 474	Check variety	83	20	86	17	50	1	70	5
Deltapine 491	Check variety	69	12	78	20	48	3	68	8
CA4007 *	Germplasm release	--	--	--	--	--	--	--	--
15-3-115D	Advanced line	78	18	87	23	50	15	73	23
15-3-114D	Advanced line	73	24	80	24	47	22	70	6
16-2-206D	Intermediate line	38	17	65	12	25	5	60	15
16-2-209D	Intermediate line	53	6	66	9	22	3	52	16
16-2-306FQ *	Intermediate line	--	--	--	--	--	--	--	--
16-2-306FQ	Intermediate line	62	11	83	4	68	13	78	8

Environmental conditions, specifically high winds, forced the revisit interval to fluctuate to as much as 25 days, however. Flight missions were conducted in such a manner that the adjacent front and side image overlap was 75% percent. All flights were conducted at an average flight height of 20 m which produced a ground sample distance (GSD) or a per pixel resolution of ~5 mm. The drone data processing workflow adhered to current best practices for RGB image data and followed the method described in (Ashapure et al., 2019) with a minor exception in the order of processing steps relating to the generation of the digital surface model (DSM). Ground control points were fixed objects in the scene. The points were surveyed with the Emlid single-band RS+ GNSS receiver (Emlid Ltd., Budapest, Hungary) and post processed using readily available data from the NOAA Continuously Operating Reference Stations (CORS) Network (NCN), managed by NOAA/National Geodetic Survey.

Agisoft Photoscan Pro (Agisoft LLC, St. Petersburg, Russia) photogrammetry software was used to generate RGB orthomosaics as well as digital surface models (DSM). The photogrammetric processing routine typically follows a workflow that produces a sparse point cloud using a structure from motion (SfM) algorithm. The conventional routine is to then create a dense point cloud, followed by a mesh layer that serves to produce the DSM from which the orthomosaic is derived. In our research it has been observed that a higher quality orthomosaic can be generated if the DSM from which the orthomosaic is produced, is derived from the sparse point cloud rather than the

dense point cloud (Figure 1). This appears to be because the surface utilized to overlay the images to produce the composite orthomosaic is less complex, ultimately reducing "swirling" effects that can be observed in drone derived orthomosaics. This does mean that the DSM must be produced a second time once the dense point cloud is constructed which serves to produce a better three-dimensional representation of the scene, as one might expect.

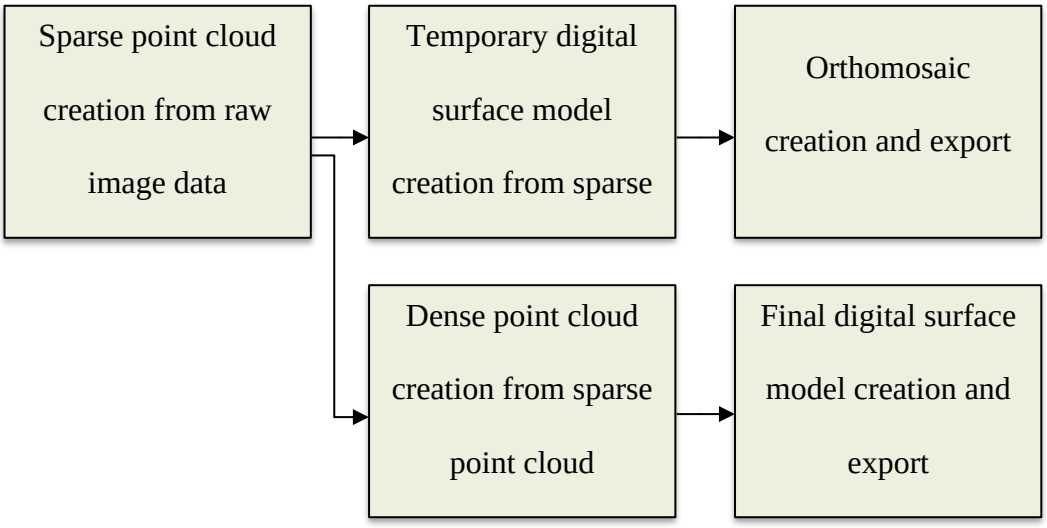

Figure 1 Process diagram of photogrammetric processing steps used in this study which depicts a slight variation in the conventional routine. The orthomosaic is derived only from the sparse point cloud while the digital surface model is derived from the dense point cloud.

Table 2 Flight dates, wind speeds, weather conditions, and resultant orthomosaic and digital surface model (DSM) ground sample distances (GSDs) for all flights in 2020

Flight Date	*Flight Height (meters)*	*Orthomosaic Ground Sample Distance (mm)*	*DSM Ground Sample Distance (mm)*	*Mission Start Time (24 H CDT)*	*Wind Speed (m/s)*
26-May-2020	20	4.98	9.96	12:35	6.70
10-Jun-2020	20	5.65	11.30	10:46	1.50
12-Jun-2020	20	5.65	11.31	10:56	3.60
18-Jun-2020	20	6.19	12.37	11:38	2.10
23-Jun-2020	20	5.42	10.84	13:05	4.60
26-Jun-2020	20	5.71	11.43	9:27	6.58
01-Jul-2020	20	5.15	10.31	10:06	7.70
02-Jul-2020	20	5.37	10.74	9:36	3.59
07-Jul-2020	20	5.35	10.71	14:02	2.06
09-Jul-2020	20	5.37	10.74	10:12	5.70
15-Jul-2020	20	5.15	10.31	13:28	2.10
20-Jul-2020	20	5.34	10.21	13:38	4.10
23-Jul-2020	20	5.11	10.68	13:34	1.65
03-Aug-2020	20	5.30	10.60	13:23	2.01
05-Aug-2020	20	4.85	10.22	15:14	8.20
19-Aug-2020	20	5.05	10.09	11:22	2.10
25-Aug-2020	20	5.36	10.73	14:10	1.50
03-Sep-2020	20	5.39	11.57	13:57	2.96
11-Sep-2020	20	4.77	9.89	14:25	1.92
31-Oct-2020	20	3.35	6.72	12:50	3.60

Table 3 Flight dates, wind speeds, weather conditions, and resultant orthomosaic and digital surface model (DSM) ground sample distances (GSDs) for all flights in 2021

Flight Date	*Flight Height (meters)*	*Orthomosaic Ground Sample Distance (mm)*	*DSM Ground Sample Distance (mm)*	*Mission Start Time (24 H CDT)*	*Wind Speed (m/s)*
17-May-2021	50	13.70	30.87	11:50	7.72
14-Jun-2021	20	4.89	9.89	12:57	4.63
16-Jul-2021	20	4.71	9.25	14:32	5.14
22-Jul-2021	20	4.92	9.53	13:06	3.60
27-Jul-2021	20	4.86	9.46	14:22	2.49
29-Jul-2021	20	4.64	9.08	12:41	0.54
05-Aug-2021	20	4.68	9.15	11:22	4.12
18-Aug-2021	20	4.77	9.75	13:45	2.57
27-Aug-2021	20	4.76	9.68	14:23	3.09
07-Sep-2021	20	4.63	9.58	14:09	3.24
15-Sep-2021	20	4.56	8.96	14:31	1.54
04-Oct-2021	20	4.75	9.20	12:33	3.60
28-Oct-2021	20	5.27	10.27	9:36	5.14
04-Nov-2021	20	4.54	8.91	14:10	2.06

2.4. Plot-level extraction

Once the complete set of time series RGB orthomosaics and DSMs were amassed for each season they were imported into QGIS for analysis. QGIS is an open-source geographic information system well-suited for drone related geospatial analysis (QGIS.org, 2022). With the drone derived raster layers imported into the GIS environment, a vector layer was created representing the experimental design. This vector layer serves to define a set of geometries that allow the metadata associated with individual plots in the experiment to be associated with respective plot-level extractions generated from the drone layers (Figure 2).

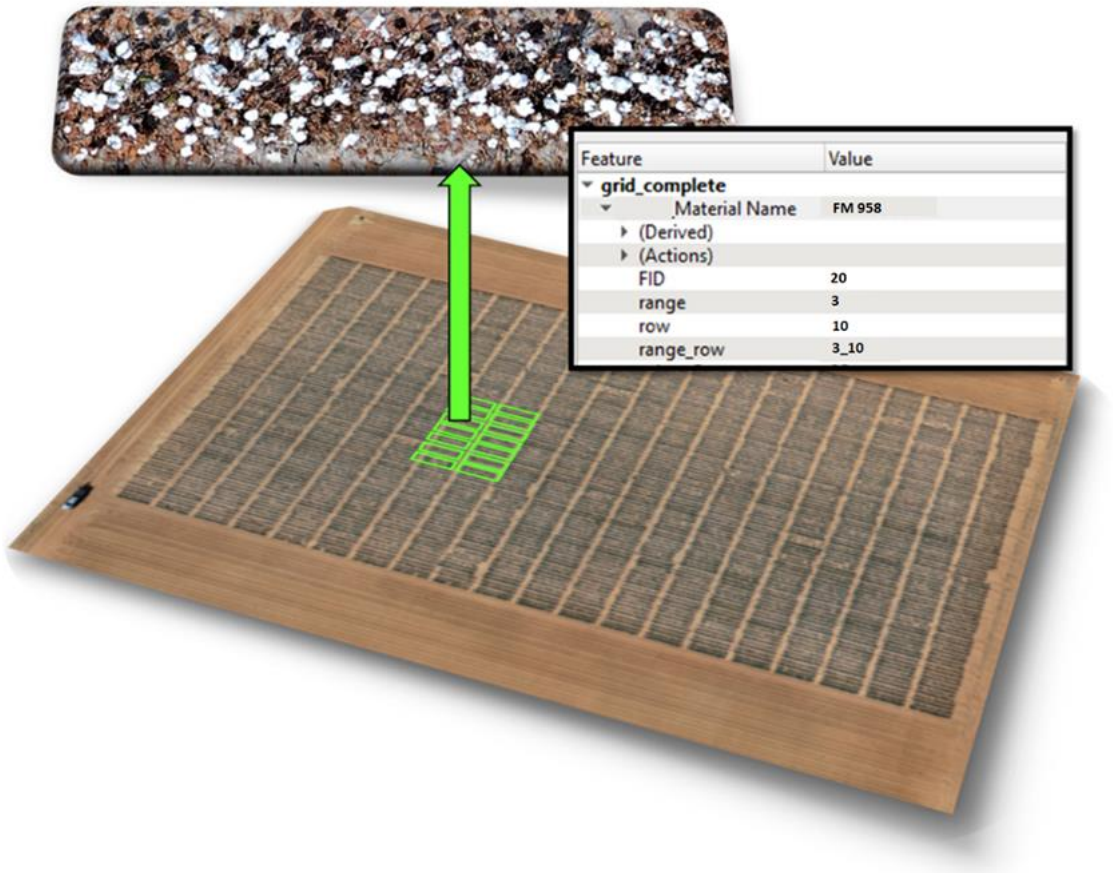

Figure 2 Plot vector creation within a GIS space allows for individual plot level extractions from raster data as well as the association of relevant plot metadata with extracted information

The plot-level extractions, for both RGB and DSM data, serve as the base data for analysis. With a complete set of plot-level extractions for both RGB and DSM data for all flights generated, the growth assessment process can begin in earnest. This begins by calculating canopy cover (CC) values from visible spectrum vegetative indices, and canopy volume (CVol) estimates for each plot in each flight for both years. At this point, the data analysis is reduced to simple multiband geospatial image analysis and was performed utilizing the Python programming language and libraries available through

the general Python ecosystem (Python.org, 2022). The canopy cover analysis is a process of segmentation over a derivative of the plot-level extraction itself. In the work conducted by (Ashapure et al., 2019) four methods were evaluated for generating canopy cover values and it was observed that the RGBVI formula produced results most closely correlated with segmented NDVI images which are widely accepted to accurately represent healthy vegetative matter and canopy cover (Tenreiro et al., 2021).

2.5. Vegetative Index Segmentation & Canopy Volume Derivation

In this work, growth rate observations were generated from six visible light indices: EXG, RGBVI, MGRVI, GNORM, TGI, VARI, and from canopy volume values generated from the DSM. This work includes a value generated from simply normalizing the green band over the sum of all other bands that will be referred to hereafter as GNORM. This process is described in the work by (Ashapure et al., 2019) but the researchers in that case used the normalized green band with red and blue normalized bands to produce an excessive greenness index (ExG) rather than evaluating qualities of the normalized green band exclusively. Our work includes the results of the GNORM thresholding for generating CC values due to its simplicity and efficacy for providing a means to segment cotton cotton canopy in plot-level extractions. Along with RGBVI, GNORM, and ExG we evaluated two other visible spectral indices: triangular greenness index (TGI) as described in (De Ocampo et al., 2019), and visible atmospherically resistant index (VARI) as described in (García-Martínez et al., 2020). The formulas used to calculate the various RGB vegetative indices are recorded in Table 4. All methods require a threshold value to perform the segmentation process and the results of that

process are compared (Figure 3). In all cases the threshold value is identified through an empirical process over a subsample of the image data set. Once the optimal threshold

Table 4 Formulas used to calculate growth observations for each plot-level extraction from both vegetive indices and canopy volume estimates. * The approach described by Ashapure et al, 2020 has been modified herein. † The use of a simple normalized green layer

Name	*Data Product Source*	*Formula*	*Threshold Value*	*Reference*
ExG	*RGB Orthomosaic*	$ExG = 2G_n - R_n - B_n$ $X_n = \frac{X}{R+G+B}$	0.15	(Woebbecke et al., 1995)
MGRVI	*RGB Orthomosaic*	$MGRVI = \frac{G^2 - R^2}{G^2 + R^2}$	0.15	(Bendig et al., 2015)
RGBVI	*RGB Orthomosaic*	$RGBVI = \frac{G^2 - R \times B}{G^2 + R \times B}$	0.15	(Hague et al., 2006)
GNORM	*RGB Orthomosaic*	$G_{norm} = \frac{G}{R+G+B}$	0.35	n/a†
TGI	*RGB Orthomosaic*	$TGI = R_{green} - \alpha R_{red} - \beta R_{blue}$ $\alpha = \frac{2(\lambda_{blue} - \lambda_{green})}{(\lambda_{blue} - \lambda_{green})}$ $\beta = \frac{2(\lambda_{green} - \lambda_{red})}{(\lambda_{blue} - \lambda_{red})}$	0.05 α = 0.59 β = 0.52	(Hunt Jr et al., 2011) (De Ocampo et al., 2019)
VARI	*RGB Orthomosaic*	$VARI = \frac{G - R}{G + R - B}$	0.05	(Gitelson et al., 2002)
CVol	*Digital Surface Model*	$CVol = GSD^2 \times \sum_{i=1}^{w}\sum_{j=1}^{h}(DSM_{ph(i,j)})$ $DSM_{ph} = DSM - \min(DSM)$		(Ashapure et al., 2020) *

value is ascertained, the plot level extractions are subjected to a binary thresholding process and pixel counts generated.

The canopy volume values differ from the canopy cover values in that each pixel value represents a canopy height value. The raw output values in the plot-level extractions from the DSM are in altitude or meters above sea level. Converting the plot-

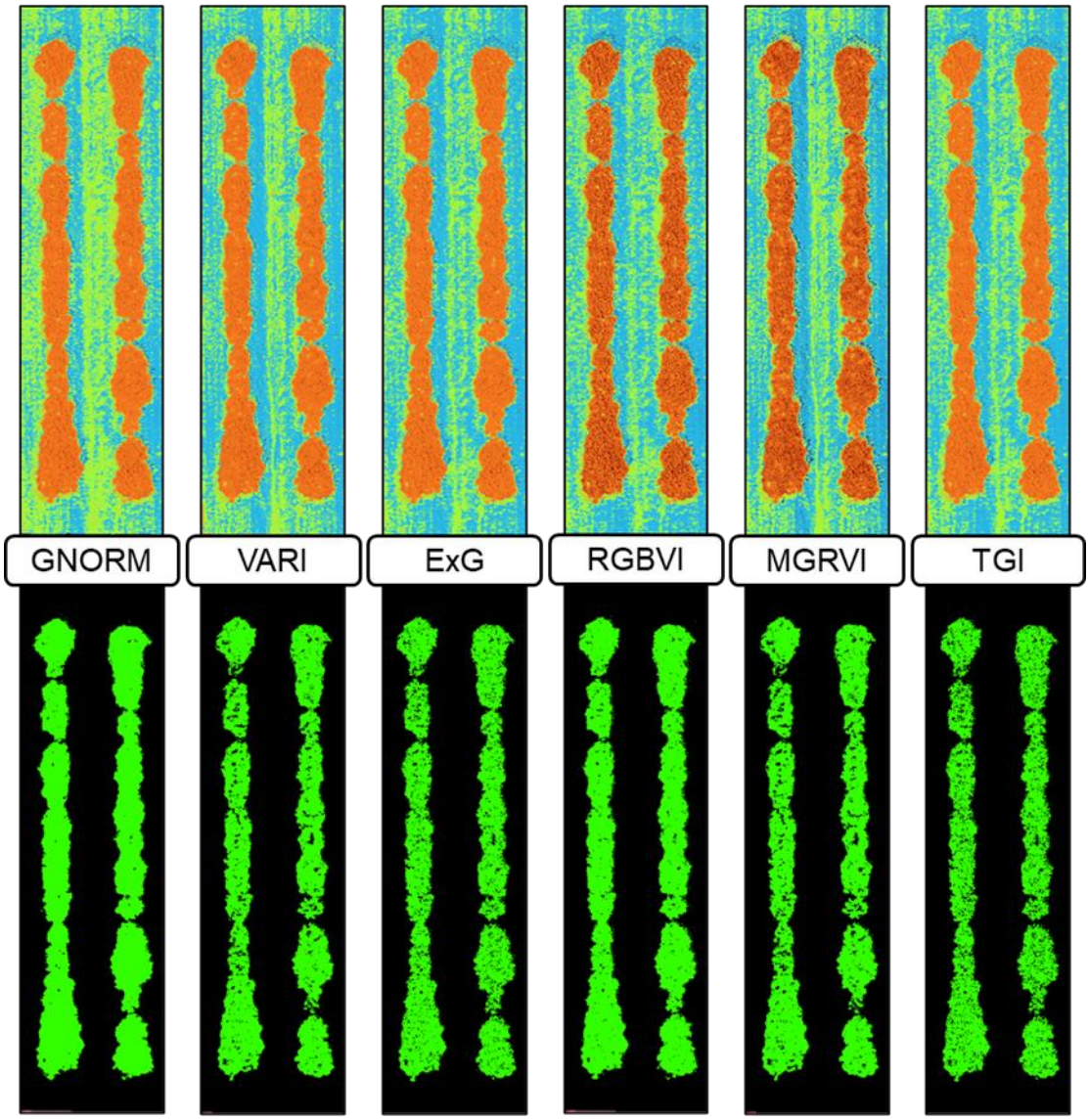

Figure 3 Depiction of RGB based vegetative indices and corresponding segmented plot-level extractions based on threshold values.

level extractions from raw values in meters above sea level to plant height in meters has been done in previous work by taking a bare soil flight and creating a DSM from that flight. The bare soil DSM is subtracted from all successive DSMs to create a raster layer that represents height above bare soil (Ashapure et al., 2020). Our process for converting the DSM values to plant height differed in that the local minimum value in each individual plot-level extraction was used to establish the soil height value. When alleys or bare soil values are present in each plot-level extraction this method for calibration

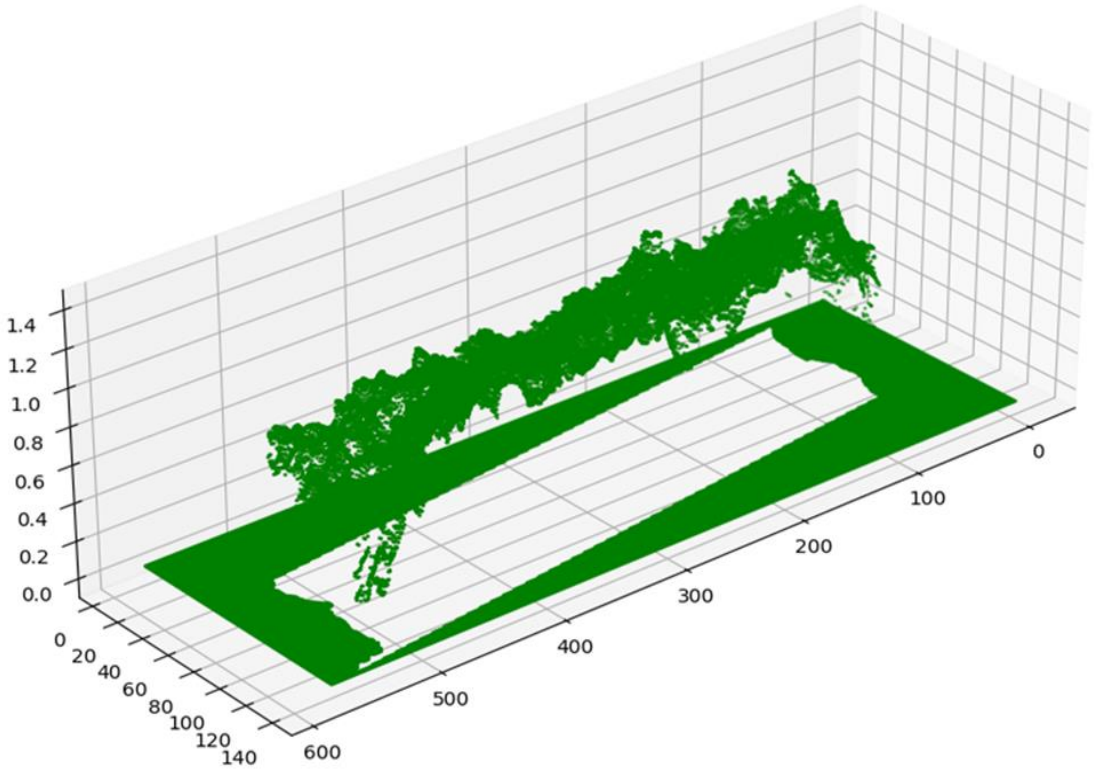

Figure 4 Plot-level DSM extractions are calibrated from raw elevation values to plant height using the local plot minimum value. Outliers are removed using a threshold of 3.5 standard deviations. All pixel values that fall below the 20th percentile are set to 0.

becomes possible. However, in situations where a single plot-level extraction contains only canopy cover and zero bare soil pixels exist, this method would not be feasible. The reason for calibrating DSM values for each plot uniquely is that the distortions that can occur in photogrammetric reconstruction are minimized in this way based on our observations. Another reason for using this approach is that a bare soil flight taken at planting may not be representative of the soil surface two months later. Rainfall, wind and field management practices all change the shape of the field surface at the soil level. Beds change size and furrow depths change. A DSM created at the beginning of the season, which is then used to calibrate all other flights, may not be representative of soil surface conditions mid-season.

The initial step in the process for calibrating plot-level DSM data is to remove outlier values. Photogrammetric reconstruction can create spurious points in the point cloud step that can result in occasional unrealistic values in a single cell in the DSM raster. These were removed by excluding values greater than 3.5 standard deviations from the mean. Cells with outlier values were then replaced with the mean value for the individual plot-level raster. The minimum value at this point is subtracted from all values in the plot-level extraction (Equation 1).

Equation 1

$$DSM_{ph} = DSM - \min(DSM)$$

Equation 2

$$CVol = GSD^2 \times \sum_{i=1}^{w} \sum_{j=1}^{h} \left(DSM_{ph(i,j)} \right)$$

The final step is to remove noise values that fall under 20 percent of the maximum value which removes noise in the data related to the fact the the bed and soil are not uniform surfaces. It is important to be sure that all values used in the canopy volume estimation are plant canopy exclusively, hence the elimination of the values below the 20th percentile. A visual representation of the resultant plot-level plant height raster is depicted in Figure 2. We then find the sum of all pixels in the two-dimensional array are then multiplied by the square of the ground sample distance to create a value that represents the canopy volume (Equation 2).

2.6. Sigmoid curve interpolation

After the canopy cover and canopy volume values are extracted from each plot for all flights in both years, data are used to interpolate a sigmoid like growth curve (Figure 4). The interpolated growth curve over the biomass proxies serves as a means to quantify and characterize aspects of plant growth for a specific variety over water treatments and years. Work conducted by (Koya & Goshu, 2013) and (Szparaga & Kocira, 2018) describes the use of the generalized logistic function (Equation 3) to characterize biological growth.

Equation 3

$$A + \frac{K - A}{(C + e^{-Bt})^{\frac{1}{v}}}$$

Curve fitting was conducted using the Python programming language and the Scipy scientific computing library (Virtanen et al., 2020). Initial parameter guesses are provided prior to processing based on range of values for each RGB vegetative index as well as for the DSMs. Parameter estimation was constrained within reasonable boundaries for respective parameters.

2.7. Curve Derivative Characteristics Analysis

Using the parameters generated through the curve fitting process, the derivative of the sigmoid growth curve was calculated using the rigorous definition of a derivative

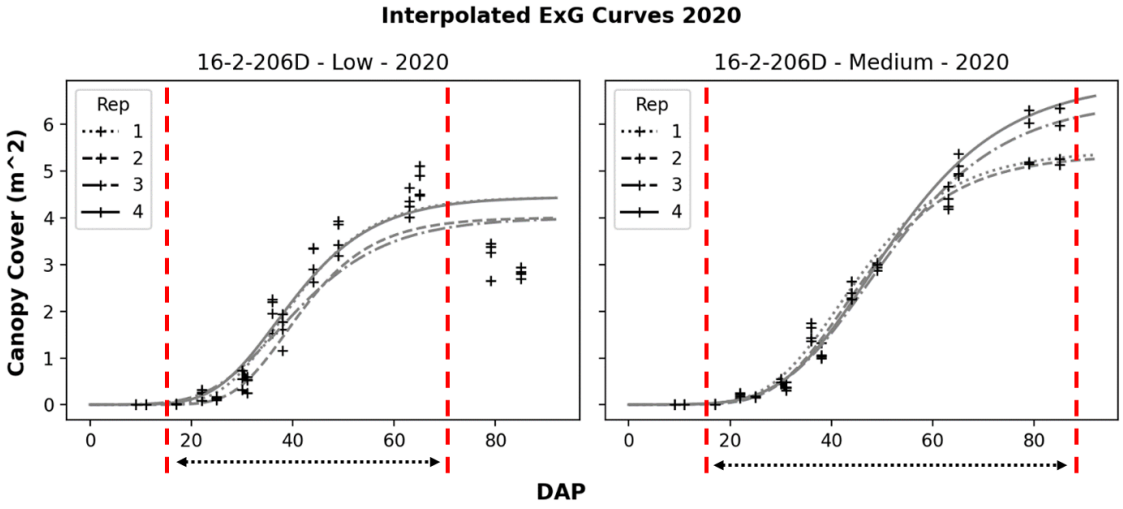

Figure 5 The scientific computing library, SciPy, was used to fit a generalized logistic function to the observed data and interpolate a sigmoid growth curve for all vegetative indices and DSM data.

The scipy.optimize submodule provides python functions facilitating the deployment of the estimated curve parameters. Once the optimization process is complete a series of values representing the days after planting (DAP) of interest can be passed into the curve function and the interpolated curve is produced. When calculating the derivative of the curve, a value of 1e-5 was used for h in Equation 4.

Equation 4

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Once the curve derivatives values were estimated over the series of DAP values representing the period of interest, the max growth rate and the first half-max growth rates were identified and exported. The days after planting values for the max growth

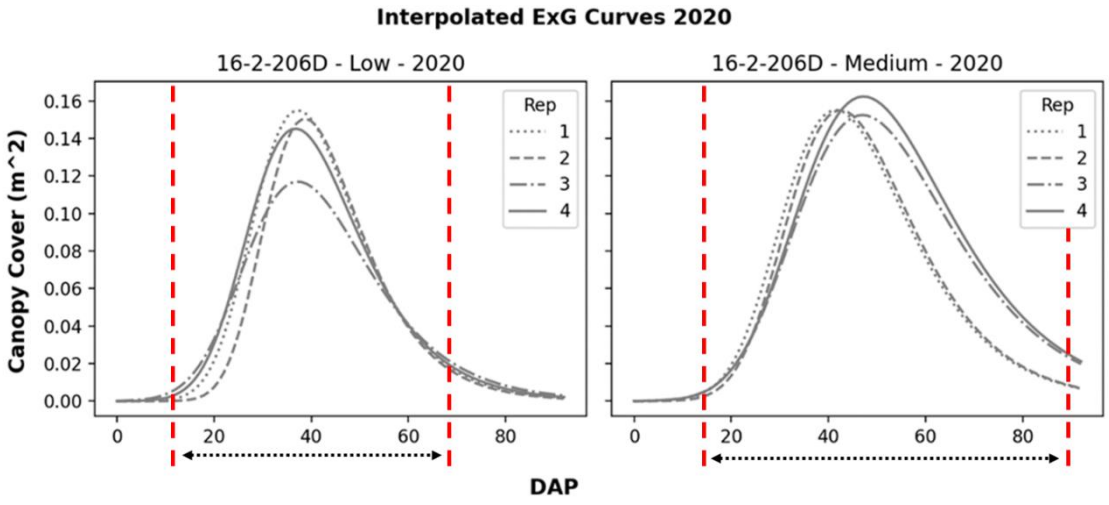

Figure 6 Derivatives of the sigmoid curve were produced to describe growth rate for all RBG indices and DSM plot data.

rate and the first half-max growth rate were exported as well. The outcome of this processing step reduces the season-long growth information into two separate quantifications: a single value for the first half maximum growth rate and a second value for the maximum growth rate (Figure 6). The DAP values for these growth rate events are also exported for analysis.

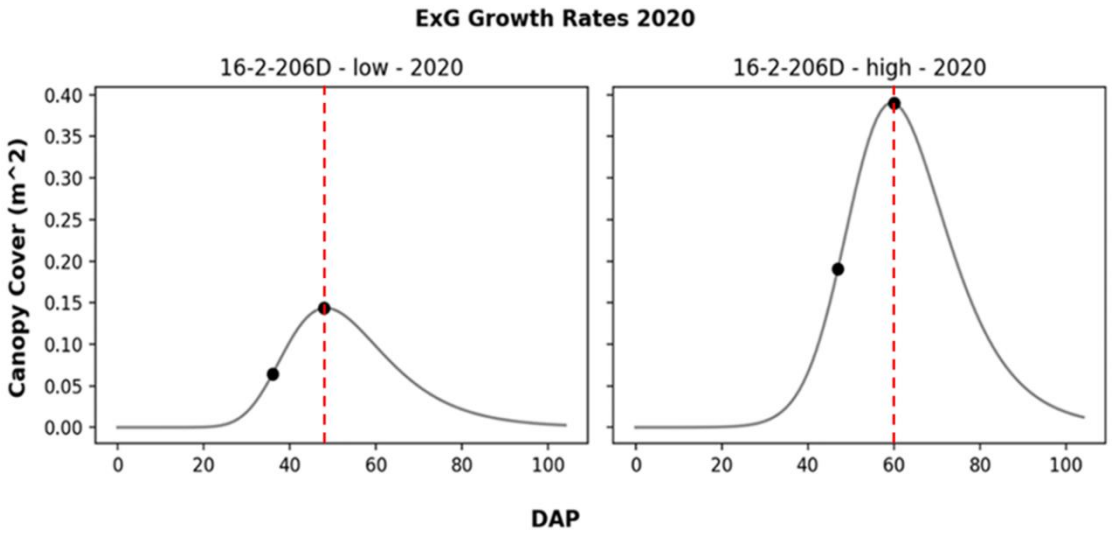

Figure 7 The values for the max growth rate and the first half-max growth rate were exported from the derivative curves. The respective days-after-planting values were also exported.

3. RESULTS

3.1. Water Treatment Effect

In 2020, the ET_c replacement from total irrigation and rain for the three treatments was was 45.8%, 54.8%, and 86.4% respectively. The annual rainfall for 2020 was below average for the Texas High Plains allowing for water treatments effects to materialize. The treatment effect was shown to be significant in 2020 with ANOVA at $\alpha = 0.05$ (Table 5). In 2021 the rainfall total was higher than the previous year resulting in an ET_c replacement for the three treatments of 72.2%, 80.74%, and 93.12 % respectively. Although the separation of entry mean yields between treatments was less distinct in 2021, as a result of abundant rainfall, the treatments were shown to be significant with ANOVA at $\alpha = 0.05$ (Table 6). Figure 6 depicts yield variability across water treatments for specific entries. The separation between water treatments in 2021 was weaker with the largest apparent difference occurring between the high irrigation treatments and the medium and low irrigation treatments combined.

The purpose of an irrigation treatment in this experiment was to create variability in growth to provide insight on both the ability to successfully characterize growth across variable environmental conditions and to identify lines from the ten chosen genotypes that might demonstrate a level of yield and growth stability across variable water inputs. The treatment effect was great enough to generate the desired variability based on the results of the analysis of variance.

Table 5 Analysis of variance for genotype and water treatment effects in 2020. * Indicates significance at alpha 0.05

Source	*DF*	*SS*	*MS*	*F-Value*	*P-value*
Genotype	9	12.6	1.4	2.909	0.00467 *
Water Treatment	2	1020.7	510.3	1064.04	< 0.001 *
Genotype X Water Treatment	18	5.4	0.3	0.621	0.87422
Error	87	41.7	0.5		
Totals	119	1080.4			

Table 6 Analysis of variance for genotype and water treatment effects in 2021. * Indicates significance at alpha 0.05

Source	*DF*	*SS*	*MS*	*F-Value*	*P-value*
Genotype	9	34.3	3.81	0.662	0.740
Water Treatment	2	471.0	235.48	40.904	< 0.001 *
Genotype X Water Treatment	18	98.0	5.44	0.946	0.528
Error	85	489.3	5.76		
Totals	119	1092.6			

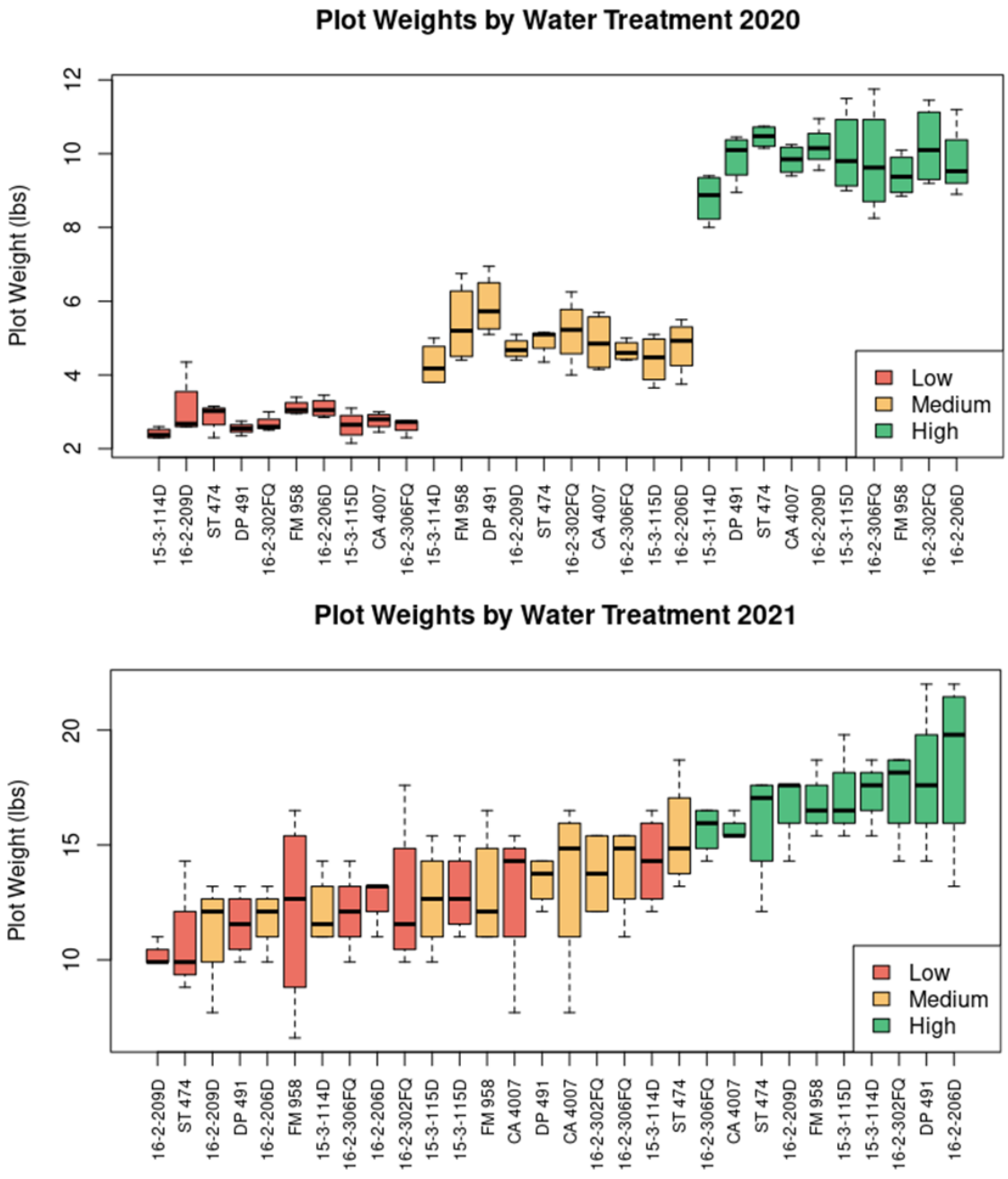

Figure 8 Plot weights by genotype across water treatments for 2020 and 2021. The genotypes are ordered by entry mean plot weight.

3.2. Processing Method & Max Growth Variability

In both years the range in observed DAP value for max growth rate was broader when derived from the DSM processing methodology than in the vegetative index growth curve derivations. A noteworthy aspect of growth curves produced through the DSM data layer versus the RGB orthomosaic data layer is that when applied broadly over many flights the DSM is more error prone. The early volumetric estimations of the plant canopy within a given plot are subject to greater error given that single anomalous features in the DSM will comprise a large portion of the overall volume estimated. The other aspect of the DSM generation is that spurious features in the sparse and dense point clouds tend to create anomalies in the surface model that will ultimately be interpreted as canopy volume. As a result of these nuances with DSM growth curve generation vs canopy cover-based growth curves, the calculated values for DAP to max growth rate were less stable, more variable and demonstrated weaker correlations with end of season traits of interest like yield or specific fiber quality parameters. The digitally derived traits based on the DSM data product were less stable and more error prone (Figure 9). Conversely, the digitally derived phenotypic characters based on canopy cover values demonstrated greater stability and stronger correlations with end of season outcomes over the two years of experimentation.

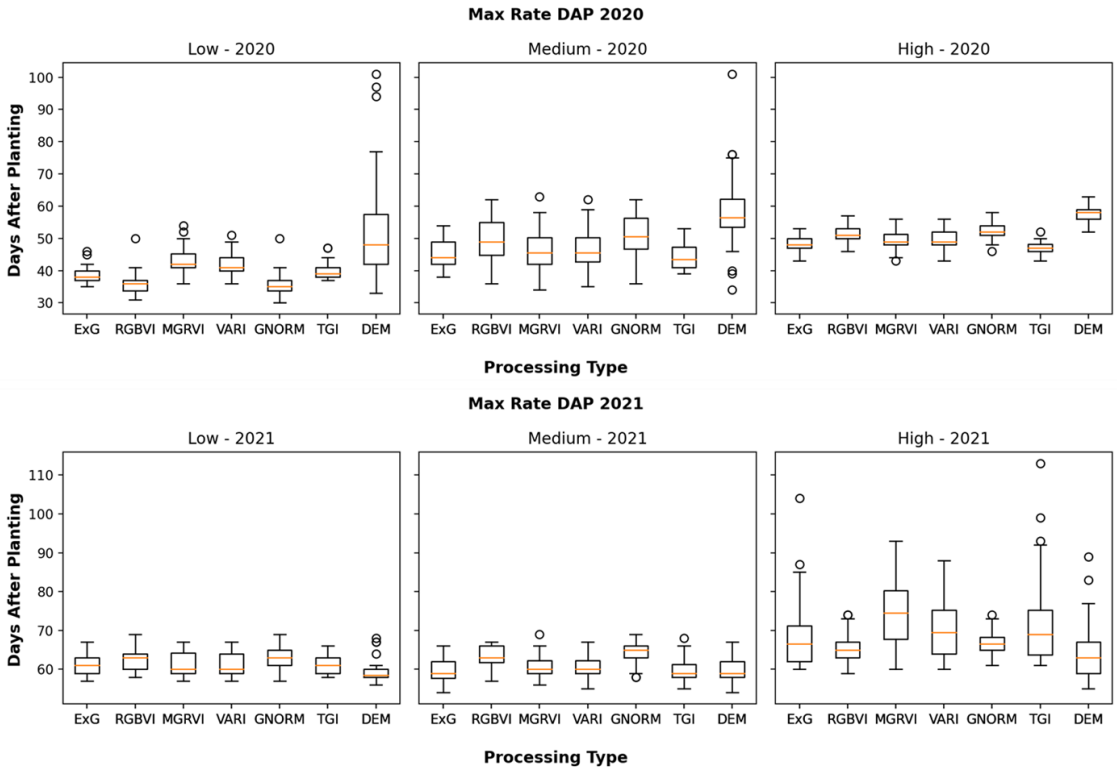

Figure 9 Boxplots for the number of days after planting at which the max growth rate was observed across both processing type and water treatment.

3.3. Normalizing DAP with DD60 Heat Units

Recording number of days after planting is an intuitive and straightforward way to evaluate temporal differences related to max growth rate. When comparing entries in a single year this approach is sufficient. However, when comparing data from multiple years, the count of days after planting to max growth rate is not robust enough to mitigate the reality that a single day in 2020 can be much different than that of 2021 in terms of plant metabolism growth potential. This occurs because seasonal temperature patterns differ year over year (Figure 10).

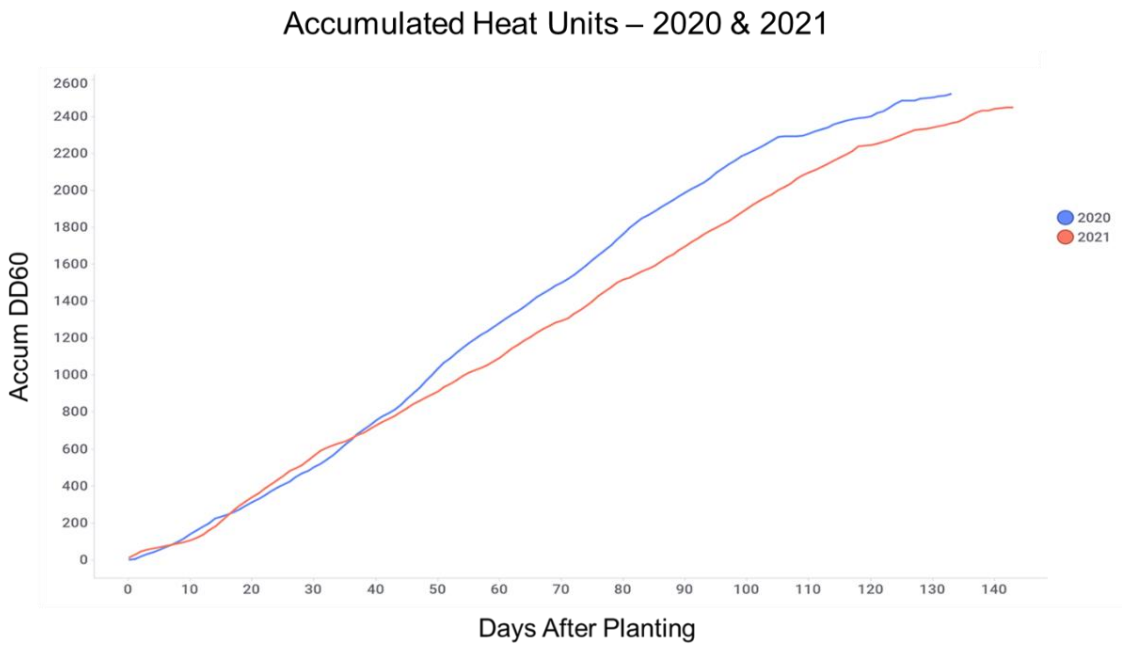

Figure 10: Air temperature differences in 2020 vs. 2021 resulted in distinctly different patterns of heat unit accumulation. Normalizing the time span from planting to landmark growth events with heat units allows for comparisons across time and space.

It makes sense, then, to replace the number of days after planting with a sum of accumulated heat units (DD60s). This is a measure of accumulated daily mean temperature above a threshold, which in this case is 60 °F (Main, 2012). Comparisons of heat units to max growth rate vs. days after planting to max growth rate are more direct and generalizable as environmental heat above the threshold has a large influence on growth.

3.4. Heat Units to Max Rate vs. Fiber Yield

The relationship between heat units to max growth rate and fiber yield was weaker for the the max growth rate values generated from the digital surface model-

based growth curves. When normalized with heat unit accumulation and evaluated over all treatments and years, the relationship between heat units to max growth rate and fiber yield for DSM based growth curves was not evident. This, as described, resulted from the variability in the DSM observations and the greater level of error associated with that data source (Figure 11).

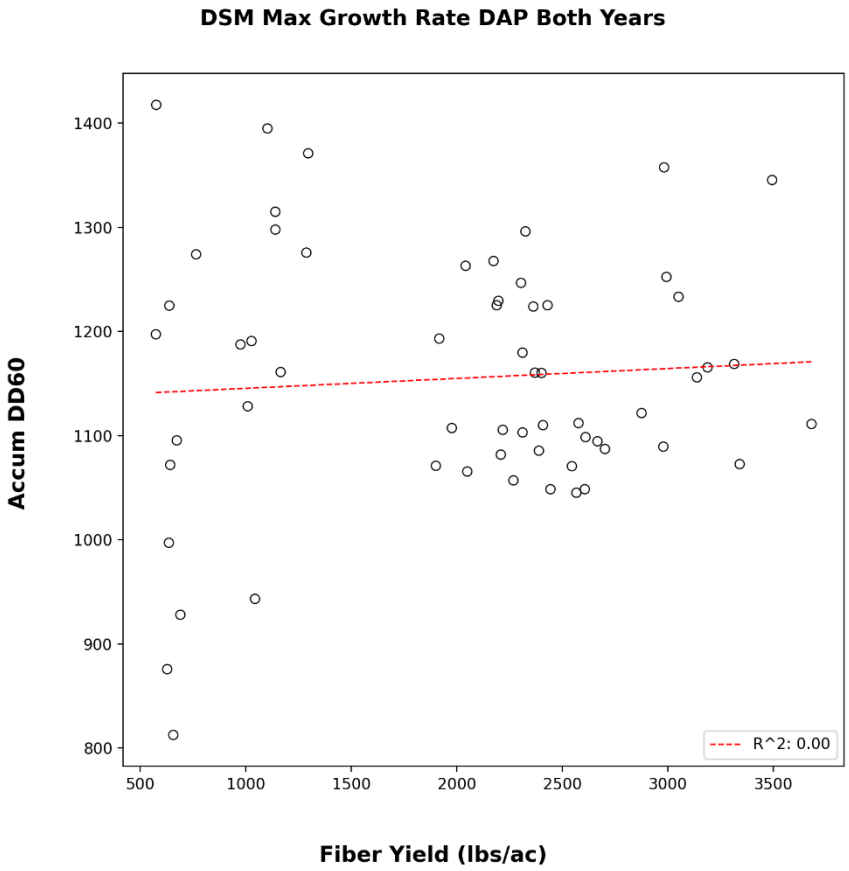

Figure 11: There was not an observed relationship between entry mean heat units to max growth rate values and fiber yield over all years and treatments. The DSM data demonstrated greater variability and was more susceptible to error.

A stronger relationship was observed between the heat units to max growth rate values derived from the canopy-cover-based growth curves and fiber yield. The growth curves interpolated from the observed canopy cover values were more stable and less error prone. The coefficient of determination varied from 0.67 to 0.77 across the six RGB based vegetative indices used to produce the interpolated growth curves and heat unit to max growth rate values (Figure 12).

3.5. Heat Units to Max Rate vs Fiber Quality

Given the lack of relationship between the heat units to max growth rate and fiber yield from the DSM derived growth curves, the entry mean DSM based values were not compared with entry mean fiber quality parameters. The canopy-cover-based growth curve derivative feature values, however, were compared to several fiber quality parameters including micronaire and fiber length. The coefficient of determination for the calculated entry mean heat units to max growth rate value and the entry mean fiber length ranged from 0.11 to 0.34. The vegetative indices that produced the curve derivative features with the strongest correlation to entry mean fiber length were the RGBVI and GNORM indices. The VI curves with the weakest correlation were the TGI and MGRVI indices (Figure 13). For other fiber quality parameters, such as micronaire, no relationship between the calculated entry mean heat units to max growth rate values was observed. All RGB based growth curve derivative features demonstrated this lack of correlation to micronaire.

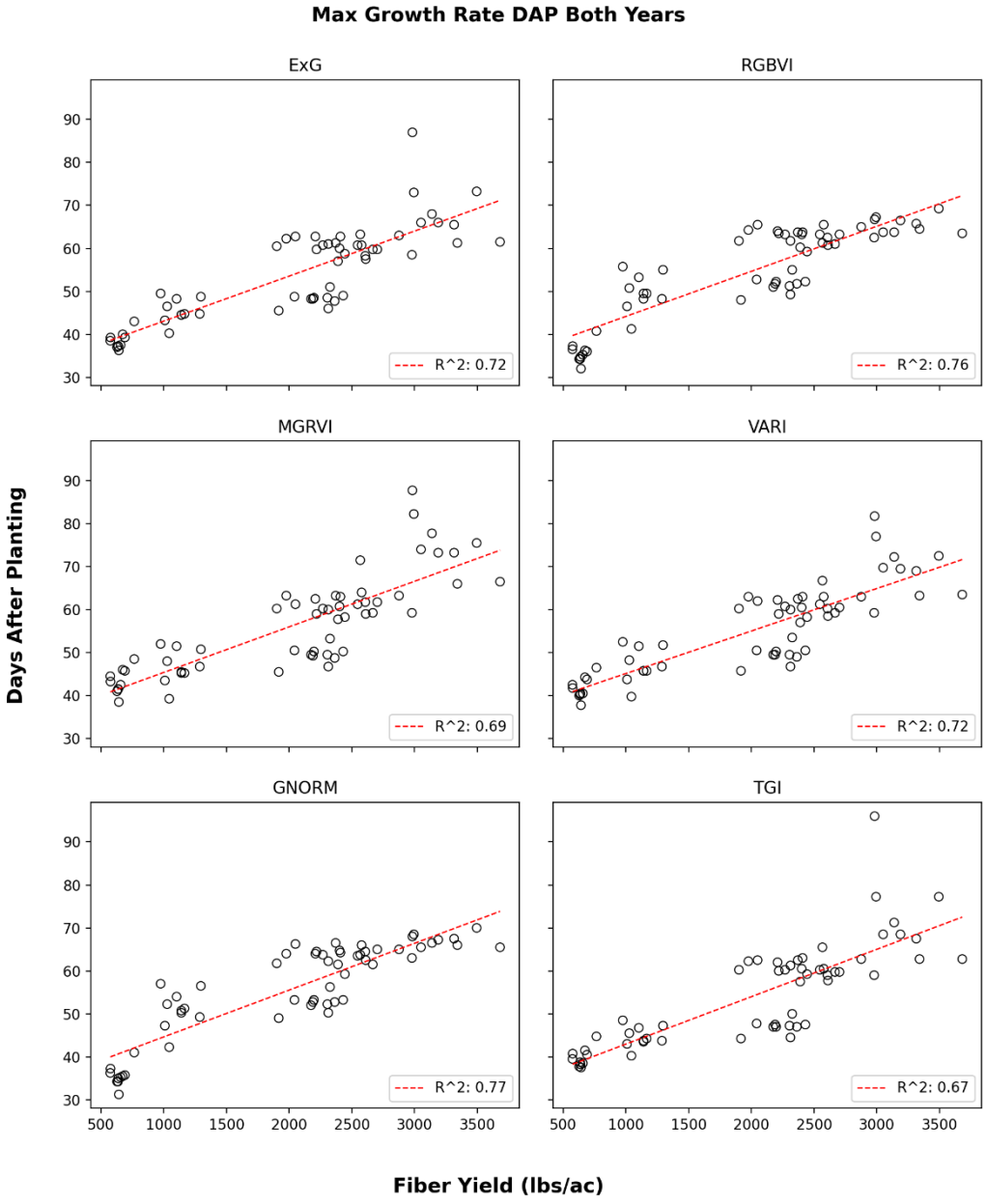

Figure 12: The heat unit to max growth rate values produced from the canopy cover observations based on the six RGB vegetative indices were correlated with entry mean fiber yield when compared across both 2020 and 2021.

3.6. Variance Components and Repeatability

Variance components were analyzed using a mixed model approach. Models were fit in using the R programming language and linear mixed models function, `lmer()`, from the lme4 package (Bates et al., 2014). Variances for the heat units to max growth rate calculations for the two vegetative indices that demonstrated the overall strongest relationship with both fiber yield and fiber length were calculated and are expressed as a proportion of the total variance (Tables 7 & 8).

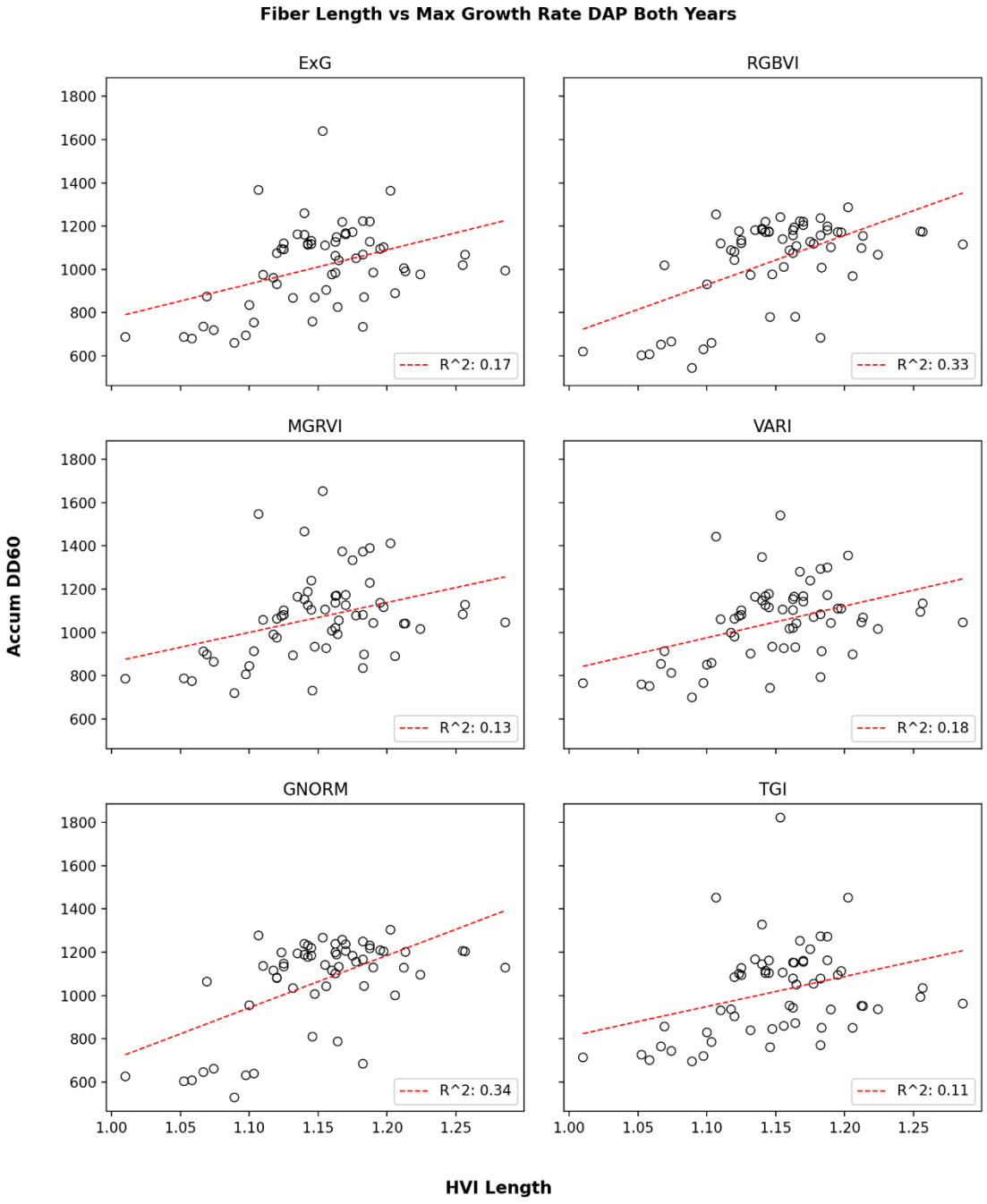

Figure 13: A subtle relationship was observed between the calculated entry mean heat units to max growth rate values and the entry mean fiber length values over both years and all treatments.

Table 7: Variance components as a proportion of total variance from GNORM derived growth curves.

GNORM		
Group	**Value**	**Percentage**
Genotype:Water	0.0	0.0
Genotype:Year	1108.7	1.5
Genotype	835.6	1.1
Water	20871.0	28.2
Year	35784.4	48.3
Residual	15446.3	20.9

Table 8: Variance components as a proportion of total variance from ExG derived growth curves.

ExG		
Group	**Value**	**Percentage**
Genotype:Water	765.0	1.1
Genotype:Year	500.6	0.7
Genotype	1701.2	2.5
Water	10166.3	14.9
Year	44805.1	65.8
Residual	10176.7	14.9

4. DISCUSSION

4.1. Max Growth Rate as a Phenotypic Character

Aspects of the digitally derived phenotypic character described in this research, heat units to max growth rate, reveal meaningful growth trends. Traditional means of characterizing the point in time where such an event might occur are limited in terms of implementation. Measures of LAI can be obtained through several nondestructive methods with devices such as under-canopy light bars and handheld devices like the LI-COR LAI-2200c Plant Canopy Analyzer. These devices have been shown to be accurate but none of the commercially available non-destructive, canopy measurement devices allow for a low-cost, low-labor deployment over many hundreds or thousands of breeding plots.

In plant breeding, phenotyping technologies must be high throughput to be effectively used. The use of drone technology to assess this growth milestone through a high throughput assessment of canopy growth in segmented RGB drone imagery is both scalable and managed with a reasonable level of cost and labor. The significance of growth characterizations produced from such data lies in the relationship with important end of season outcomes like yield and in the manner in which the value can be obtained in a high throughput, scalable and easily managed method.

4.2. Heat Units to Max Growth and Yield

In both years of this research the following relationship between heat units to max growth rate and fiber yield was observed: as the heat units to max growth rate value increases, so does the yield. From a physiological standpoint, this measure of growth

describes the amount of time, as a matter of accumulated heat, a given entry will exploit before reaching the maximum rate of canopy development. The rate of canopy development for cotton plants, and the length of the period of canopy development, are important as this is a direct representation of the upper limit to yield potential in a cotton plant. The reproductive structures on the plant must be supported by photosynthate produced in green leaves. The longer the vegetative head start a given cotton plant enjoys in a season, the greater potential for that plant to support fruiting structure production and development. This source-sink relationship is a fundamental component of yield potential in most crops.

Another aspect of the source-sink relationship in cotton and the ability to characterize the maximum rate of source (canopy) development, is that before the initiation of fruiting structures on the plant, the primary carbohydrate sink on the plant, other than new canopy growth, is the roots (Ritchie et al., 2007). For areas that experience environmental water deficit stress such as the Texas High Plains, this early source-sink relationship and the timing of the maximum canopy development rate may provide important information to the breeder as to the length of the period where the plant's roots can exclusively consume carbohydrates from productive leaves. Max growth rate timing and its relationship with canopy development supporting carbohydrate production used in root growth may also contribute to the positive correlation between heat units to max growth rate and fiber yield.

4.3. Heat Units to Max Growth and Fiber Length

A subtle correlation was observed between the heat units to max growth rate values and fiber length. The maximum coefficient of determination for this relationship was .34 but the average R^2 score was just under 0.2. First flower development is generally accepted to occur in a cotton crop around 900 DD60s or roughly 60 days after planting. Soon after opening, the flower is self-pollinated and roughly 12 hours after pollination, fertilization will occur (Stewart, 1986). At this point, the boll begins to develop where the initial portion of the the boll filling phase is entirely devoted to the elongation of the fiber cells. As bolls develop most of the assimilate received by the boll is generated in leaves on the same sympodial branch (Schubert et al., 1986). The physiological basis for a positive correlation with fiber length and heat units to maximum growth rate is likely similar to that of fiber yield as a whole. The longer the plant can develop canopy the greater the support for boll development and carbohydrate intensive developmental processes that occur within the boll such as fiber elongation.

The fact that a relationship to fiber length and not micronaire was observed may be a product of the micronaire evaluation itself with the High Volume Instrument (HVI) system. Different fiber configurations with regard to secondary cell wall thickness and fiber diameter, also known as fiber maturity, can produce similar micronaire values and therefore might result in poor correlations to milestone growth events. HVI is widely used for its speed and cost but debate is ongoing with respect to its use in plant breeding (Kelly et al., 2012). Future work could explore the relationship with growth milestones and fiber maturity measured with devices that more precisely measure secondary cell

wall development as a proportion of overall fiber diameter using the Advanced Fiber Information System (AFIS).

4.4. Genetic vs Environmental Effects on Observation

The fact that heat units to maximum growth rate is a useful measure of crop growth is clear. However, when analyzing variance components in this work we see that this measure of growth is heavily influenced by the environment. In breeding studies, the focus of nearly all effort is to select entries with some form of genetic superiority. In the case of this work, with the assessment of growth as a measure of heat units to maximum growth rate, it appears that little of the variability for this value is a direct result of genetic differences. The experimental design was such that the major environmental variability was introduced in the form of irrigation treatments because of the initial focus to evaluate lines that might demonstrate a level of yield stability and growth rate stability in the semi-arid region of the Texas High Plains.

The percentage of genetic variability from the total variability reported in Tables 7 and 8 for heat units to maximum growth rate for the GNORM and ExG based growth curves was quite small with values of 1.1 and 2.5 percent respectively. This outcome highlights one of the major challenges associated with conducting breeding experiments, specifically for a species native to tropical regions like cotton, in a water deficit stress environment or under water deficit stress treatments. One might posit that cotton breeding operations should only be conducted in areas like the Mississippi Delta, where rain is abundant and floodplain soils are nutrient rich. Phenotypic variability in these

environments is much more likely to be a result of genetics, rather than environment, given the ideal growth conditions for all breeding plots.

4.5. Screening Early Generation Material

In this research, previously tested lines at an "intermediate" or above level of development were selected. Intermediate in this case means uniform breeding lines in the second year of multiple location testing. Advanced lines are in the third year of multiple location testing. The observed maturity variability from previous years was used as a means to identify germplasm to include in this experiment that might possess growth rate variability. Even after intentionally introducing growth variability based on maturity, the variability within a treatment was small and the intrinsic genetic variability with respect to all other variability was minimal. In nurseries and progeny row trials the innate genetic variability between lines is typically far greater than that of lines that have been selected and advanced over several years as the act of selection itself reduces genetic variability over time.

It would be a tremendously labor intensive and potentially cost prohibitive exercise to generate a heat unit to maximum growth rate characterization for entries in a large cotton breeding nursery. Conversely, early generation and nursery material would likely be the segment of the breeding pipeline that would contain the greatest proportion of intrinsic genetic variability. It follows then that the method for assessing maximum growth rate as function of heat unit accumulation developed in this research would be well suited to early generation and nursery material, primarily because there is not currently another practical method to quantitatively derive such information over many

thousands of plots. Future efforts might include further method development and implementation over large cotton breeding nurseries and early generation trials.

4.6. Instability of DSM Derived Growth Curves

The evaluation of the growth rate values derived from the digital surface models indicated early in the analytic process that this data source contained variability that was too great to be useful in the heat units to maximum growth rate assessments. Three-dimensional point cloud generation from RGB imagery is a complex process that can be affected by plant morphologies as well as by the environmental conditions at data capture. Time of day, specifically in water deficit stress environments, can alter the overall position of the leaves in the canopy of a cotton plant. In the early hours of the day, before wilting occurs, the position of the leaves is different than what might be observed in flight data captured at solar noon or during the hottest portion of the day in the afternoon. The point cloud is based on feature matching and detection which would be influenced by canopy wilting.

With the RGB based vegetative indices, the primary observed value from which the sigmoid curves are interpolated is a segmented measure of ground cover. From the nadir aerial view of the drone, the breadth of the canopy in all directions would change in instances of severe wilting but this variation is smaller than a three-dimensional assessment of a wilted canopy as produced in the volumetric value derived from the digital surface model. Another factor contributing to the high level of variability and poor relationship with yield and fiber quality parameters for the digital-surface-model-based heat units to maximum growth rate values is the fact that the soil surface itself is

highly variable and changes over time. As mentioned, previous methods for determining plant growth based on digital surface models used a bare soil flight collected before planting occurs. The subsequent digital surface models are then calibrated throughout the season by subtracting the values of the DSM_0. What remains is a surface model that describes a plant canopy in terms of volume above the initial soil surface. In principle this is a straightforward process, but the surface soil is fluid throughout the season and is affected by rain, wind and crop management practices in the field. The early season representation of the soil is certainly much different than what exists at season's end after many weather events and passes through the field with heavy equipment. In the approach used in this research the soil surface was calibrated using the bare soil surface observed in each flight in the alley between plots and adjacent to each plot where this value was assessed.

Beyond dealing with soil surface variability, this method also mitigates the between flight variability that can be observed between successive digital surface models even when a sufficient number of ground control points are captured during flight and used during the photogrammetric stitching process. Nevertheless, efforts to manage variability in the point cloud derived three-dimensional digital surface model generated in this research may not be sufficient to generate growth curve and growth rate values that have relationships to end of season outcomes for individual plots.

4.7. Selected Lines

Based on the heat units to maximum growth rate, yield and fiber quality values across water treatments, two lines were identified through this research as candidate

lines potentially well-suited for cultivation in the Texas High Plains regions where water availability and environmental impacts are becoming increasingly variable as weather patterns continually shift. These lines are 15-3-114 and 16-2-306.

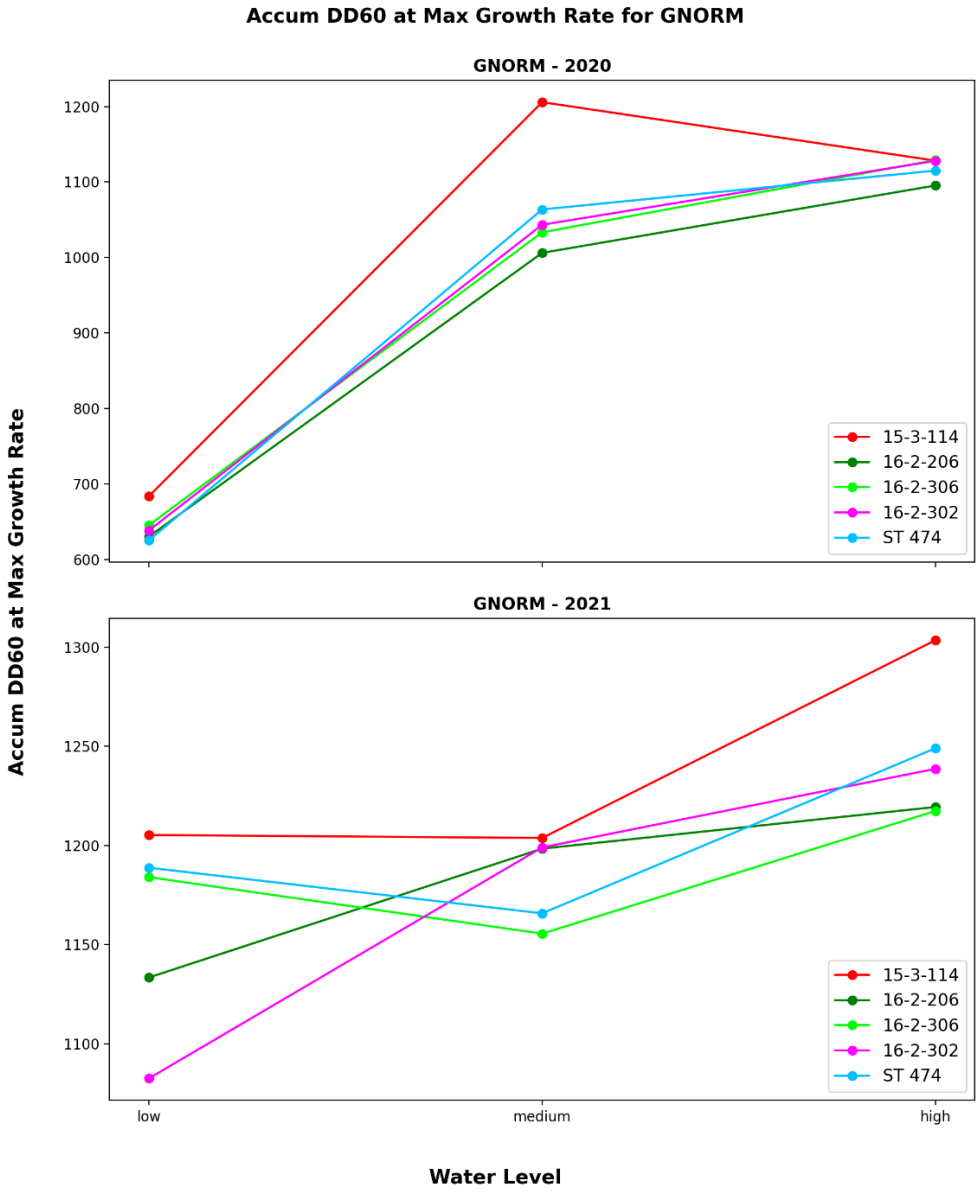

Figure 14: Fiber yield values for selected varieties across both treatments and years. From these, two varieties were selected: 15-3-114 and 16-2-306.

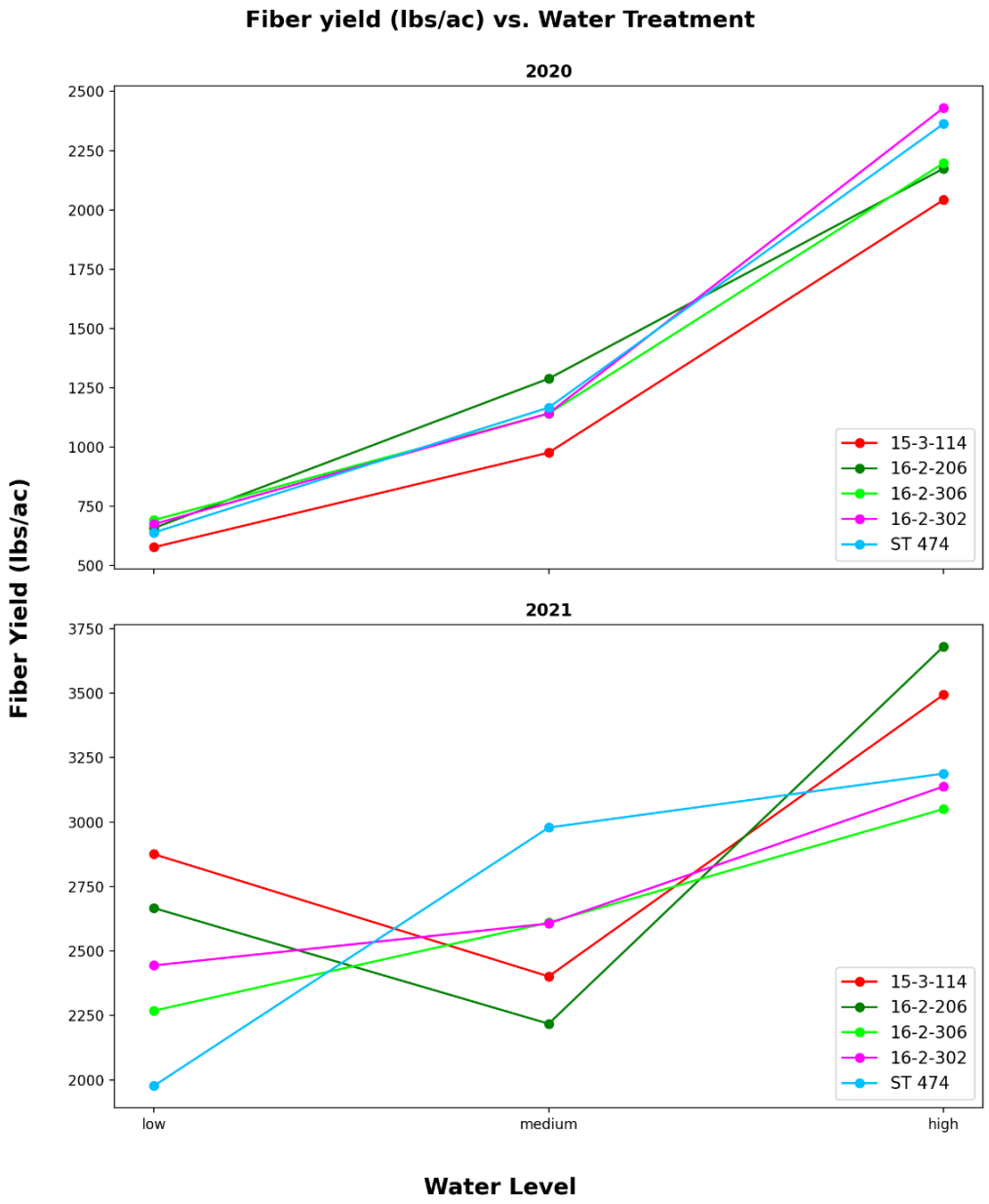

Figure 15: Heat units to maximum growth rate values for selected varieties across both treatments and years. From these, two varieties were selected: 15-3-114 and 16-2-306.

In 2020, which was the most discriminating of the two years in terms of water deficit stress across treatments, 15-3-114 demonstrated a long period of canopy development

with a high value for heat units to max growth rate value. In 2021, the wetter of the two years, this line again had a long canopy development period and later max growth rate. 15-3-114 had average yield across treatments in 2020 and above average yield in 2021 in the low and high-water treatments. 16-2-306 was average for growth rate timing and canopy development in 2020 and above average for canopy development in 2021. For yield 16-2-306 was one of the better performing lines in the low water treatment of 2020 and just above average in the low and high-water treatments of 2021.

5. CONCLUSION

5.1. Growth Rate Selection in Breeding

The yield of a cotton crop is related to the ability of the crop to produce carbohydrates from photosynthesis to support fruit production. Plants that can develop canopy and increase leaf area have a greater maximum potential for photosynthesis and a greater ability to produce carbohydrates. Therefore, increased assimilate production is a sensible breeding objective. Identifying entries in a breeding trial that possess a longer canopy development phase and a superior ability to maximize leaf development would be a challenging proposition with traditional methods and techniques. Point-in-time plot level observations for canopy development might be gathered with a reasonable level of labor and time but when one considers gathering such information across a large breeding trial 10 or 12 times a season, the workload becomes unmanageable.

The use of high throughput digital technologies like drones makes the challenge of collecting plot level data many times throughout the season possible with the level of resources one might typically find in a breeding program. Drone technologies have become common in plant science over the last decade and are becoming an integral part of most breeding programs. The potential for drone technologies to gather data with high temporal frequency that can be reduced to single numeric values, as is the case with heat units to max growth rate, is particularly useful to the plant breeder. Rather than attempting to makes sense of periodic changes or full sigmoid growth curves, the method of analysis in this research allows the breeder to make straightforward comparisons on growth rate across years and regions. Using the accumulated DD60

value mitigates the differences that occur when comparing days after planting from one year to the next or from one region to the next.

For regions like the Texas High Plains where economic productivity and long-term stability is dependent on healthy cotton production systems and germplasm, it will be necessary to augment traditional breeding practices in the region with technologies and practices that provide more information for selection. Digitally derived phenotypic characters such heat units to max growth rate allow breeders to select for traits not previously quantifiable. As the weather patterns of the THP regions change and natural groundwater resources decline, breeders and growers will face the challenge of producing greater yields with fixed or reduced inputs. With quantifiable growth rate data breeders will be able make selections with greater information about environmental stability and explore aspects of germplasm in their programs not previously investigated.

REFERENCES

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). FAO Irrigation and drainage paper No. 56. *Rome: Food and Agriculture Organization of the United Nations*, *56*(97), e156.

Ashapure, A., Jung, J., Chang, A., Oh, S., Maeda, M., & Landivar, J. (2019). A comparative study of RGB and multispectral sensor-based cotton canopy cover modelling using multi-temporal UAS data. *Remote Sensing*, *11*(23), 2757.

Ashapure, A., Jung, J., Chang, A., Oh, S., Yeom, J., Maeda, M., Maeda, A., Dube, N., Landivar, J., & Hague, S. (2020). Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, 180-194.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 79-87.

Bourland, F. M. (2019). Jacob Osborne Ware, an Early Cotton Breeding Giant. *The Journal of Cotton Science*, *23*(3), 239-245.

Chawade, A., van Ham, J., Blomquist, H., Bagge, O., Alexandersson, E., & Ortiz, R. (2019). High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy*, *9*(5), 258.

Chu, T., Chen, R., Landivar, J., Maeda, M., Yang, C., & Starek, M. (2016). Cotton growth modeling and assessment using unmanned aircraft system visual-band imagery. *Journal of Applied Remote Sensing*, *10*(3), 036018. https://doi.org/10.1117/1.JRS.10.036018

De Ocampo, A. L. P., Bandala, A. A., & Dadios, E. P. (2019). Estimation of Triangular Greenness Index for Unknown PeakWavelength Sensitivity of CMOS-acquired Crop Images. 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM),

Dever, J. (2012). Cotton breeding and agro-technology. In *Handbook of natural fibres* (pp. 469-507). Elsevier.

Duan, T., Zheng, B., Guo, W., Ninomiya, S., Guo, Y., & Chapman, S. C. (2016). Comparison of ground cover estimates from experiment plots in cotton, sorghum and sugarcane based on images and ortho-mosaics captured by UAV. *Functional Plant Biology*, *44*(1), 169-183.

Eaton, F. (2003). Measuring yields from small plots: A modified cotton picker. International Conference on Crop Harvesting and Processing,

García-Martínez, H., Flores-Magdaleno, H., Ascencio-Hernández, R., Khalil-Gardezi, A., Tijerina-Chávez, L., Mancilla-Villa, O. R., & Vázquez-Peña, M. A. (2020). Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and RGB images acquired with unmanned aerial vehicles. *Agriculture*, *10*(7), 277.

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote sensing of Environment*, *80*(1), 76-87.

Hague, T., Tillett, N., & Wheeler, H. (2006). Automated crop and weed monitoring in widely spaced cereals. *Precision Agriculture*, *7*(1), 21-32.

Hake, K. (2003). Cotton biotechnology: beyond Bt and herbicide tolerance. Proceedings of the world cotton research conference-3: cotton production for the new millennium. Cape Town, South Africa,

Hu, P., Chapman, S. C., Wang, X., Potgieter, A., Duan, T., Jordan, D., Guo, Y., & Zheng, B. (2018). Estimation of plant height using a high throughput phenotyping platform based on unmanned aerial vehicle and self-calibration: example for sorghum breeding. *European Journal of Agronomy*, *95*, 24-32.

Hunt Jr, E. R., Daughtry, C., Eitel, J. U., & Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal*, *103*(4), 1090-1099.

Johnson, B. P. (1939). Apparatus for measurements of lengths of cotton fibers.

Kelly, B., Abidi, N., Ethridge, D., & Hequet, E. F. (2015). Fiber to fabric. *Cotton*, *57*, 665-744.

Kelly, C. M., Hequet, E. F., & Dever, J. K. (2012). Interpretation of AFIS and HVI fiber property measurements in breeding for cotton fiber quality improvement. *J Cotton Sci*, *16*, 1-16.

Khanal, S., Kc, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing*, *12*(22), 3783.

Koya, P. R., & Goshu, A. T. (2013). Generalized mathematical model for biological growths. *Open Journal of Modelling and Simulation*, *2013*.

Lascano, R. J., Leiker, G. R., Goebel, T. S., Mauget, S. A., & Gitz III, D. C. (2020). Water Balance of Two Major Soil Types of the Texas High Plains: Implications for Dryland Crop Production. *Open Journal of Soil Science*, *10*(07), 274.

Main, C. L. (2012). Cotton Growth and Development. *University of Tennessee Extension Bulletin W*, *287*, 12-0108.

McGuire, V. L., & Fischer, B. C. (2001). *Water-level changes in the High Plains aquifer, 1980 to 1999*. US Department of the Interior, US Geological Survey.

Narayanan, B., Floyd, B., Tu, K., Ries, L., & Hausmann, N. (2019). Improving soybean breeding using UAS measurements of physiological maturity. Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV,

Pabuayon, I. L. B., Sun, Y., Guo, W., & Ritchie, G. L. (2019). High-throughput phenotyping in cotton: a review. *Journal of Cotton Research*, *2*(1), 18. https://doi.org/10.1186/s42397-019-0035-0

Paterson, A. H., & Smith, R. H. (1999). Future horizons: biotechnology for cotton improvement. *Cotton: origin, history, technology, and production*, *4*, 415.

Python.org. (2022). Python 3.9. In.

QGIS.org. (2022). QGIS Geographic Information System. In.

Ritchie, G. L., Bednarz, C. W., Jost, P. H., & Brown, S. M. (2007). Cotton growth and development.

Schubert, A. M., Benedict, C., & Kohel, R. (1986). Carbohydrate distribution in bolls. *JR Mauney and J. McD. Stewart (eds.). Cotton Physiology. The Cotton Foundation, Memphis, Tenn*, 311-324.

Stewart, J. (1986). Integrated events in the flower and fruit. *Cotton physiology*, *1*, 261-300.

Szparaga, A., & Kocira, S. (2018). Generalized logistic functions in modelling emergence of Brassica napus L. *PLoS One*, *13*(8), e0201980.

Tenreiro, T. R., García-Vila, M., Gómez, J. A., Jiménez-Berni, J. A., & Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture*, *182*, 106038.

Texas Water Devlopment Board. (2020). *Ogallala Aquifer*. Retrieved 05/02/2022 from http://www.twdb.texas.gov/groundwater/aquifer/majors/ogallala.asp

Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A review on UAV-based applications for precision agriculture. *Information*, *10*(11), 349.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272.

Wang, T., Liu, Y., Wang, M., Fan, Q., Tian, H., Qiao, X., & Li, Y. (2021). Applications of UAS in crop biomass monitoring: A review. *Frontiers in Plant Science*, *12*, 616689.

Wanjura, D., Upchurch, D., & Mahan, J. (1990). Evaluating decision criteria for irrigation scheduling of cotton. *Transactions of the ASAE*, *33*(2), 512-0518.

Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, *38*(1), 259-269.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., Xiong, L., & Yan, J. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular Plant*, *13*(2), 187-214.

Yeom, J., Jung, J., Chang, A., Maeda, M., & Landivar, J. (2017). Cotton growth modeling using unmanned aerial vehicle vegetation indices. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS),

Zeng, L., Stetina, S. R., Erpelding, J. E., Bechere, E., Turley, R. B., & Scheffler, J. (2018). History and current research in the USDA-ARS cotton breeding program at Stoneville, MS. *Journal of Cotton Science*, *22*, 24-35.