5-2022

Modeling Leaf Area Index and Canopy Height Using Growing Degree Days

Ellie J. Kuhn
University of Arkansas, Fayetteville

Modeling Leaf Area Index and Canopy Height Using Growing Degree Days

Ellie Kuhn

Biological Engineering Program

Biological and Agricultural Engineering Department

College of Engineering

University of Arkansas

Undergraduate Honors Thesis

This thesis has been approved by the Biological and Agricultural Engineering Department for submission to the College of Engineering and Honors College at the University of Arkansas.

____

(Signature and Date)

BENG Honors Advisor – Dr. Benjamin Runkle

____

(Signature and Date)

Honors Committee Member and Departmental Honors Coordinator – Dr. Brian Haggard

____

(Signature and Date)

Honors Committee Member – Dr. Kusum Naithani

____

(Signature and Date)

Honors Committee Member – Dr. Beatriz Moreno-García

Abstract

As the global population increases and food security is recognized as a critical issue, crop growth prediction models help ensure the sustainability of reliable food sources. Using a prediction model based on temperature and simple, measurable field parameters, e.g., Leaf Area Index (LAI) or Canopy Height (H_{can}), may allow farmers and others to intervene mid-season with fertilizer, irrigation, or other inputs to obtain a better harvest.

This study aims to create a general model that could predict LAI and H_{can} values for numerous rice varieties using Growing Degree Days (GDD) as the time scale. The models use data collected during the 2018-2020 growing seasons for 16 fields in east-central Arkansas. After comparing model performance indicators (coefficient of determination (R^2), root mean square error (RMSE), percent bias (pbias), percent difference, and Akaike Information Criterion values (AIC) of quadratic and sigmoid regression forms, a sigmoid regression with GDD as its time scale was chosen as the best functional form for the datasets provided. The sigmoid with GDD was chosen due to its higher R^2 values and lower AIC values (LAI: R^2= 0.82, AIC= 14.97; H_{can}: R^2= 0.88, AIC= 83.01), compared to the other models. The data was then divided into calibration and validation datasets, accounting for field and rice variety differences. The calibration dataset created a generalizable model, and the validation dataset ensured the model could be applied successfully over varying field conditions (LAI: R^2= 0.78, RMSE= 1.15 m^2m^{-2}; H_{can}: R^2= 0.85, RMSE= 13.7 cm).

Three cultivar-specific models for the CL-XL745, XP753, and Gemini214 CL cultivars were created and compared to the general model. Overall, there were only minor differences between each model, with the statistics values remaining within a tight range between the

general and cultivar-specific models. Further work is being pursued on the benefits of dividing the data based on field cultivar. The uncertainties due to less representative calibration datasets within the cultivar-specific models make the general model the preferred choice for a future wide-scale application for farmers to make field management decisions concerning improving yield and general field management practices within Arkansas.

1. Introduction

Crop modeling has been used for years by farmers and agronomy researchers to predict possible outcomes of current and upcoming growing seasons and prepare for issues that may be experienced during these seasons (Hammer, 1997; Matthews et al., 2013; Mehdi et al., 2018). More importantly, as food scarcity becomes a pressing issue due to the growing world population, diets shift, and biofuel production competes with crops, crop modeling and estimation will be vital for maintaining global food security (Waldner et al., 2019). There have been many approaches to creating crop models, specifically through phenological plant characteristics, like Leaf Area Index (LAI) and Canopy Height (H_{can}).

One crop of global interest is rice (*Oryza sativa*), a primary source of sustenance for approximately 3.5 billion people globally and a primary source of income for hundreds of millions of people throughout many developing countries (Muthayya et al., 2014). Although the United States accounts for less than 2% of global rice production, it contributes about 6% of global rice exports (Childs, 2021). Within the United States, Arkansas (the location of this study) is the top producer of rice. With rice being such an important food source, creating a rice crop model could assist many farmers in ensuring a productive and successful growing season. Additionally, a crop model can help farmers make pre-season decisions regarding cultivar choice, fertilizer and pesticide applications, and potential yield when applying any new techniques to their fields (Mehdi et al., 2018).

When developing a mathematic crop model, it is crucial to designate a reliable time series to build the model. Many studies use days after planting (DAP) or the day of the year (DOY) since they are simple to calculate and are consistent regardless of external factors such as cultivar and soil type (Campos-Taberner et al., 2016; Gilardelli et al., 2019). However, temperature and photoperiod have also proved to be primary factors that affect crop development (Sharifi et al., 2017). Using a temperature-based time series like Growing Degree Days (GDD) considers the varying temperatures throughout multiple growing seasons, as well as differences in planting dates. Therefore, it could create a more accurate crop model than a model using DAP. GDD is based on thermal time accumulated between a base temperature and a cut-off temperature, and it is used to describe the timing of each phenological stage (Boschetti et al., 2018). Since temperature influences Leaf Area Index (LAI) and canopy height (H_{can}), these factors should be considered when creating a mathematic crop model (Shah et al., 2011).

Leaf Area Index (LAI) has proved to be a primary contributor to mathematical model building as an integrative plant physiological variable that helps govern vegetated surfaces' mass and energy balances (Colaizzi et al., 2017). LAI is defined as the total one-sided area of photosynthetic tissue per unit surface area. It is one of the essential variables in climatic, ecological, and agronomical research studies (Stroppiana et al., 2006). However, measuring LAI can be both labor-intensive and destructive to the crops. The use of LAI and H_{can} as outcomes for a crop model may eliminate the need to visit numerous locations throughout a field, or several fields, to obtain readings. Farmers could measure H_{can} in their fields, and while they may not be able to measure LAI easily, they could instead use satellite-derived products to estimate it. Obtaining these field values would allow for comparisons with the model and gauge possible stress. There has also been limited cultivar-only modeling work within rice fields, particularly in Arkansas and other parts of the US Mid-South. Relationships between a general model including all the cultivars could be compared to individual models for each cultivar to determine whether creating cultivar-specific models is necessary for a successful model.

Although it has limited accuracy in predicting yield, maximum LAI is a crucial factor to observe when implementing weather variables into a crop prediction model (Waldner et al., 2019). For example, there is some evidence that maximum LAI can be used to predict yield; however, little work has been done to test this claim on rice crops, and the literature review could not uncover evidence that this test has been applied to rice crops in Arkansas. Comparing yield and Peak LAI will assist in determining if LAI has the capability of substituting for yield for farmer and modeler use in future models.

Since there has been little work on creating a rice crop prediction using a temperature-dependent time series, this work provides new insight into the extent to which GDD applies to different types of regression curves. The objectives of this study were to (1) compare models using polynomial and sigmoid-curve regressions to determine the best-fitting model across GDD and DAP time scales; (2) once the best fitting model is selected, the general and cultivar-specific models are calibrated and validation, and then compared; and (3) compare peak LAI to yield to test its predictive power.

2. Methods

2.1 Study Area

Over the 2018, 2019, and 2020 growing seasons, the study area consisted of 16 field sites, which differed for each growing season. These sites were located throughout the central-eastern region of Arkansas (Figure 1).

Rice was grown following the typical practice in Arkansas, where it was either dry, drill-seeded, and flooded at the five-leaf stage, or where the pre-germinated seed was broadcast seeded into a wet field ("water seeding"). Table 1 displays the seeding method for each field, with 27% of the fields being water-seeded and the rest drill-seeded. This table also shows each field's agronomic and soil details, including soil type, planting/harvest date, water regime used, rice cultivar, and yield values. The study includes eight different rice cultivars and two irrigation regimes: Alternate Wetting and Drying (AWD) and Continuous Flooding. Continuous Flooding occurs from the 5-leaf stage until approximately two weeks before harvest, when the fields typically have at least 10-15 cm of pooled water present, depending on the field type (i.e., zero-grade, multiple inlet irrigation (MIRI), precision grade, etc.) (Henry et al., 2021). On the other hand, AWD is a practice in which the field is allowed to dry down before reapplying irrigation water (Belder et al., 2004; Lampayan et al., 2004; Lampayan et al., 2015), creating significant water savings and mitigating greenhouse gas emissions, like methane (Chidthaisong et al., 2018; LaHue et al., 2016; Nalley et al., 2015; Runkle et al., 2019). Yield values (t/ha) were shared with the researchers by the farmers from yield monitors or on-farm estimations. These yield values were corrected to a 13% grain moisture basis for consistent yield comparisons. Soil type information was collected from the SSURGO database (USDA-NRCS).

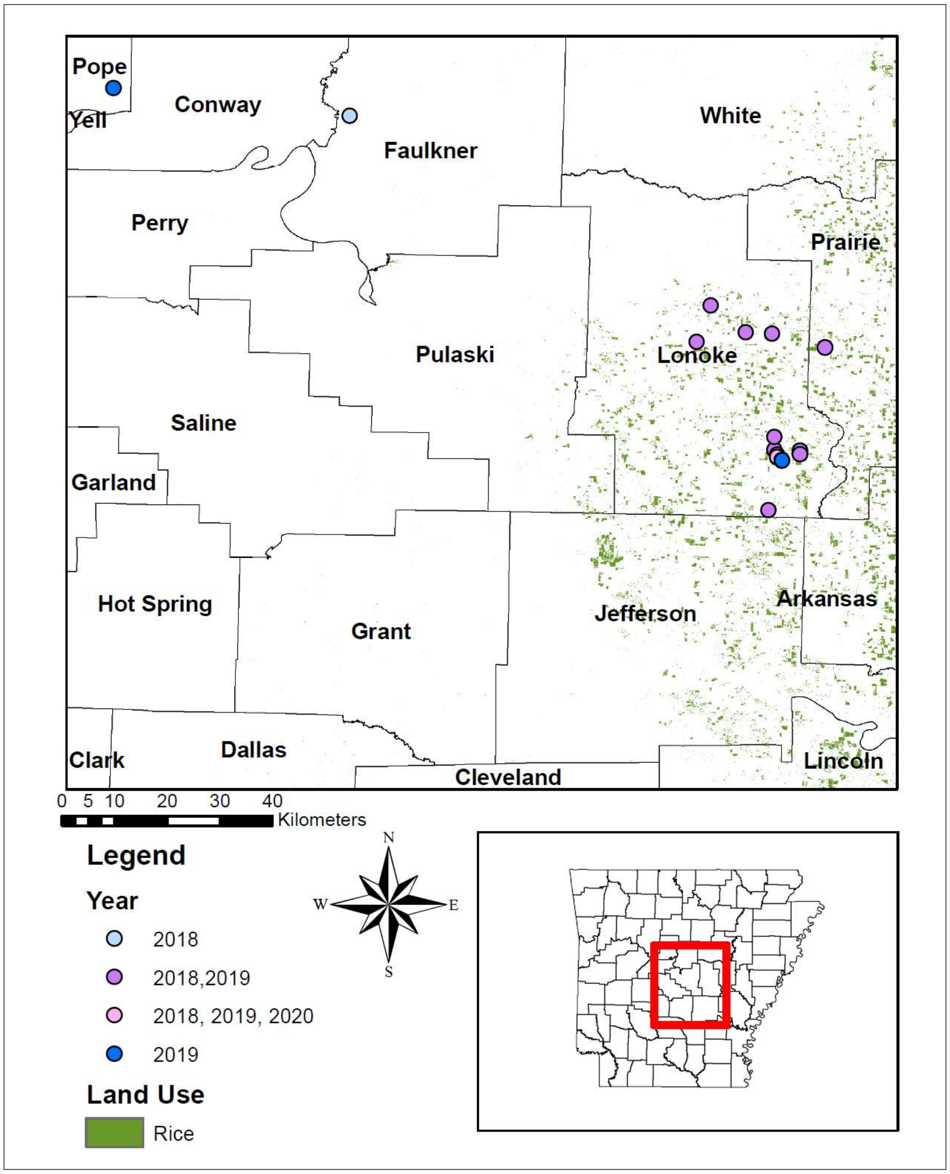

Figure 1: Map of Study Sites in Arkansas map with selected counties labeled; Land Use Data was provided by the Arkansas GIS Office (2013).

Table 1: Research site agronomic and soil details separated into their respective calibration and validation datasets

Field Name	Growing Season Year	Planting Date	Harvest Date	Soil Type	Seeding Method	Irrigation Regime	Rice Cultivar	Yield Corrected to 13% moisture (t/ha)
Calibration Dataset								
W3	2018	Apr 30	Sep 15	Perry silty clay	Drill	AWD*	CL-XL745	7.1
CS	2018	May 03	Sep 11	Calhoun silt loam	Drill	AWD	CL-XL745	8.3
R2	2018	May 01	Sep 04	Stuttgart silt loam	Drill	AWD	XP753	10.7
R7	2018	May 11	Oct 07	Perry clay	Drill	CF**	Gemini214-CL	10.4
R4	2018	May 02	Sep 16	Perry silty clay	Drill	CF	RT7311-CL	9.5
R10	2019	Jun 05	Oct 08	Perry silty clay	Water	CF	CL-XP4534	5.0
R5	2018	Apr 12	Sep 15	Stuttgart silt loam	Drill	AWD	XP753	8.1
B50	2019	Jun 10	Oct 25	Perry silty clay	Water	AWD	Gemini214 + XL745[1]	5.9
R8	2019	May 14	Sep 28	Calhoun silt loam	Drill	AWD	Gemini214-CL	10.6
R3	2019	Apr 24	Sep 22	Dewitt silt loam	Drill	CF	XP753	10.7
W3	2020	Apr 02	Aug 18	Perry silty clay	Water	AWD	CL-XL745	10.5
R9	2019	Apr 17	Sep 02	Hebert silt loam	Drill	AWD	XP760	10.4
W4	2019	May 13	Sep 12	Perry silty clay	Water	AWD	CL-XL745	8.4
CN	2019	May 16	Sep 14	Calloway silt loam	Drill	AWD	Gemini214-CL	9.0
R4	2019	May 13	Sep 18	Perry silty clay	Water	CF	CL-XL745	8.6
R10	2018	May 04	Sep 05	Perry silty clay	Drill	AWD	CL-XL745	8.8
Validation Dataset								
W4	2018	Apr 30	Aug 31	Perry silty clay	Drill	AWD	CL-XL745	9.3
CN	2018	May 03	Sep 11	Calloway silt loam	Drill	AWD	CL-XL745	8.6
R3	2018	May 06	Nov 07	Dewitt silt loam	Drill	AWD	XP753	10.3
R8	2018	Jun 04	Sep 20	Calhoun silt loam	Drill	AWD	Gemini214-CL	10.6
R9	2018	Mar 23	Aug 15	Hebert silt loam	Drill	CF	RT7311-CL	10.2
R1	2018	Apr 25	Sep 05	Calhoun silt loam	Drill	AWD	CL-XP4534	8.6
R1	2019	Apr 02	Sep 02	Calhoun silt loam	Drill	AWD	XP753	9.1
B30	2019	Jun 04	Oct 10	Perry silty clay	Drill	CF	Gemini214-CL	6.4
R2	2019	Apr 03	Sep 09	Stuttgart silt loam	Drill	CF	Gemini214-CL	10.3
R5	2019	Apr 09	Sep 03	Stuttgart silt loam	Drill	AWD	XP753	9.6
W4	2020	Apr 02	Aug 19	Perry silty clay	Water	AWD	CL-XL745	10.4
CS	2019	Jun 02	Oct 03	Calhoun silt loam	Drill	AWD	Gemini214-CL	8.4
R6	2019	May 15	Oct 17	Roellen clay	Water	AWD	Provisia	9.0
W3	2019	May 13	Sep 12	Perry silty clay	Water	AWD	CL-XL745	9.1

*AWD = alternate wetting and drying
**CF = continuous flooded from 5-leaf stage
[1] The farmer mixed the seeds from two different varieties before planting.

2.2 Field Data Collection

2.2.1 LAI Data Collection

Leaf Area Index (LAI) and canopy height (H_{can}) were measured approximately every 1-2 weeks during 2018, 2019, and 2020 growing seasons. The team obtained at least 8-15 sampling dates for each field for all three growing seasons. However, the median number of sampling dates across all fields was 11 dates, with the most consistent measurements taken in 2019. Both the LAI-2000 and LAI-2200C instruments measured LAI in 2018. In 2019 and 2020, all the measurements were performed with the LAI-2200C. It should be noted that the LAI-2000 and

LAI-2200C technically measured plant area index (PAI), but for convenience, it will be referred to as LAI throughout this study, as the instrument reports these values as LAI. The PAI includes other material that blocks the light from reaching the sensor, like stems and grain, and not only the leaf material. The LAI-2200C measures any objects blocking sunlight from reaching the sensor; therefore, the instrument measures PAI (LI-COR Biosciences, 2013).

Measurements were taken each season differently due to annual improvements in the protocol. In 2018, the measurements at each field were taken in one location arbitrarily selected within a relatively small and consistent area of the field. In 2019, two flagged locations with an area of 1 m^2 each within every field marked the measurement locations (Figure 2A). Similarly, in 2020, measurements at two flagged locations were taken, but the measurement area was increased to 4 m^2 within a 4m x 1m area (Figure 2B). The average of these two measured LAI values on each date was used for modeling. With methods for obtaining LAI measurements improving each growing season, it is possible for the models to also improve with respect to accuracy. For more information regarding LAI data readings and scattering corrections, refer to Appendix A.

A) B)

Figure 2: (A) Flagged LAI measurement location in Field R2 during the 2019 growing season, and (B) LAI measurements were taken by a research group member in the 2020 growing season at field W3; photos by B. Moreno-García.

2.2.2 Canopy Height Data Collection

Canopy height was measured in 5 locations in each field every time LAI was measured, typically within the same measurement area as LAI. The average of the five measurements was used for modeling purposes. H_{can} was measured from the soil surface to the line of the upper leaves and the top of the panicle after heading. For each growing season, at least eight and no more than 17 H_{can} sampling dates were taken at each field; most of the fields had an average of 12 height measurements.

2.3 Weather Data and Conversion to Time Series

The weather data used for GDD calculations were collected from a database supported by the PRISM Climate Group. This group gathers climate observations from a range of monitoring networks, applies quality control measures, and develops spatial climate datasets to reveal short- and long-term climate patterns (PRISM Climate Group, 2019). The PRISM datasets provide estimates of 6 basic climate terms: precipitation (ppt), minimum temperature (T_{min}), maximum temperature (T_{max}), mean dew point temperature (T_{dmean}), minimum vapor pressure deficit (vpd_{min}), and maximum vapor pressure deficit (vpd_{max}) (PRISM Climate Group, 2019). Of these basic climate terms, we used only the daily T_{min} and T_{max} to calculate GDD. For each growing season, climate data were obtained by entering the coordinates and specifying the growing season dates into the PRISM application for each field individually. The base equation for calculating GDD is:

$$GDD = \left(\frac{\min{(T_{max}, 30)} + T_{min}}{2}\right) - T_{base} \tag{1}$$

where T_{base} is 10°C. When the T_{max} is higher than 30°C, the value of T_{max} must be set to 30°C (Liu et al., 2016). This temperature range was chosen based on the desired temperatures for optimal rice growth (Sharifi et al., 2017). The base equation is applied to the PRISM temperature data but includes a conditional statement setting GDD to 0 when $\frac{(T_{max}+T_{min})}{2} < T_{base}$ (McMaster and Wilhelm, 1997). For model testing, we calculated cumulative GDD values for each field and season.

2.4 Relationships between LAI and Canopy Height to DAP and GDD

2.4.1 Polynomial Regression Models

First, quadratic polynomial regression models were created to test the relationships between LAI and H_{can} with different time series (either GDD or DAP). The general polynomial regression equation used for both models is as follows:

$$f(x) = ax^2 + bx + c \tag{2}$$

where *x* is GDD or DAP (within units of °C·day or day, respectively), and *a*, *b*, and *c* are all parameters fit using the ordinary least squares method, modeled in Excel.

2.4.2 Sigmoid-Curve Regression Models

Second, we tested a sigmoid regression form using non-linear fitting in Excel. Like the polynomial regression approach, data was inserted into an Excel file for each field for each growing season. The models were based on the general sigmoid-curve equation:

$$\mathrm{f(x)} = \mathrm{base} \cdot \left(1 + \exp\left[\frac{-(\mathrm{x}-\mathrm{inf}\,)}{sprd}\right]\right)^{-1} \cdot \left(1 - (1 + \exp\left[\frac{-(\mathrm{x}-\mathrm{opt})}{opt\_shp}\right])^{-1}\right) \qquad (3)$$

where, x is GDD or DAP, and the five parameters of the sigmoid regression are: *base* is the base of the function, *infl* is the inflection point of sigmoid curve, *sprd* is the spread of data, *opt* is Optimum GDD or DAP, and *opt_shp* is a curve shape parameter.

Each parameter had a different effect on the general curve (Figure 3). Alteration of the base parameter affected the magnitude of the curve over the y-axis. When we changed the inflection point, the part of the curve that increased the most rapidly shifted to the left or right. The spread parameter shifted the upper and lower curves of the "S," producing a more or less defined curve. Changing the optimum GDD or DAP parameter resulted in the curve's final LAI point increasing or decreasing along the y-axis. When the optimum shape parameter was altered, the uppermost peak of the curve moved upwards or downwards.

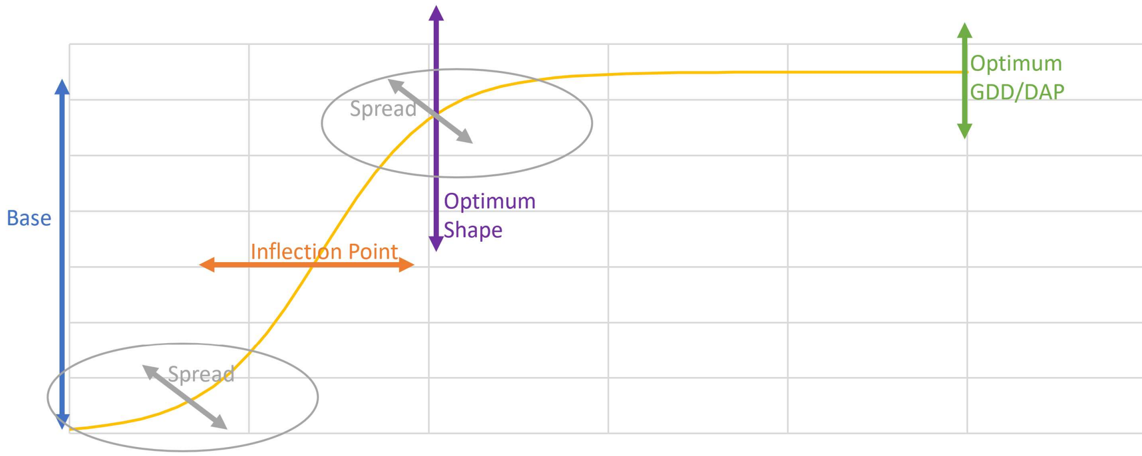

Figure 3: Schematic of how each sigmoid parameter affects the overall curve

Four graphs were created for each field year's sigmoid-curve regression model, comparing GDD and DAP to H_{can} or LAI. Initial parameter values were estimated by considering what each parameter represents within the equation and approximating values that best fit the dataset before applying the solver function. The solver function then minimized the sum of the squared errors for each regression by changing the five estimated parameters. Table 2 displays the projected initial values used for the general sigmoid models.

Table 2: Sigmoid regression estimated values for initial modeling tests

Model	**Base** (m^2m^{-2}, cm)	**Infl** (days, °C•day)	**Sprd** (days, °C•day)	**Opt** (days, °C•day)	**Opt_shp** (m^2m^{-2}, cm)
LAI v. GDD	8	750	150	3000	3500
H_{can} v. GDD	400	1700	600	1700	270
LAI v. DAP	10	50	8	1970	4600
H_{can} v. DAP	100	70	13	5000	3000

2.4.3 Testing Model Performance

For both the polynomial and sigmoid models, we analyzed the coefficient of determination (R^2), root mean square error (RMSE), percent bias (pbias), and percent difference (% diff) of the mean to test the data across a range of model performance metrics (eq. 4-7).

$$R^2 = \left[\frac{\sum_{i=1}^{n}(OBS_i - \overline{OBS})(EST_i - \overline{EST})}{\sqrt{\sum_{i=1}^{n}(OB\ \ _i - \overline{OBS})^2}\sqrt{\sum_{i=1}^{n}(EST_i - \overline{ES\ \ })^2}} \right]^2 \qquad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(OB\ \ _i - EST_i)^2)}{n}} \qquad (5)$$

$$PBIAS = \frac{\sum_{i=1}^{n}(OBS_i - EST_i)}{\sum_{i=1}^{n}(OBS_i)}\ x\ 100\% \qquad (6)$$

$$\%\ Difference = \frac{\overline{OBS} - \overline{EST}}{\frac{\overline{OBS} + \overline{EST}}{2}}\ x\ 100\% \qquad (7)$$

where n is the number of observations; OBS_i is either measured LAI or H_{can}; EST_i is either estimated LAI or H_{can} from the models; *i* = 1,2,3…n; and $\overline{OBS}\ and\ \overline{EST}$ are the mean observed and estimated values, respectively. An ANOVA test was then completed on the data to determine the p-value to test the significance of the regressions. The ANOVA was run on both the polynomial and sigmoid functions. Any models with a p-value below 0.05 for a dataset were considered statistically significant.

Akaike's Information Criterion (AIC) values also were calculated to observe the models' complexity and robustness. AIC considers the number of inputs into the model, which helps weigh whether the improvement from a polynomial to a sigmoid model justifies the use of two or more parameters. AIC is calculated according to the equation:

$$AIC = n \cdot \log(MSE) + 2T \qquad (8)$$

where n is the number of data points in the model, MSE is the mean squared error of the regression, and T is the number of parameters and input variables into the model. AIC values help to represent model complexity, where these values lie in the range of -∞ to +∞, with an optimum value of -∞ (Akaike, 1974).

2.5 Calibration and Validation of a General Model to Predict LAI and H_{can}

Creating a general model (i.e., across all cultivar and field conditions) was approached by dividing the field data into calibration and validation datasets. The datasets were first divided into pairs based on rice cultivar, growing season year, and soil type, and then we separated each pair into the calibration or validation dataset (Table 1). These categories provided an even distribution of specific field conditions throughout the calibration and validation processes. After examining the initial polynomial and sigmoid regression models, we decided it was best to create the general model with a sigmoid curve and GDD as its time series. Once a general model was calibrated, we applied the validation dataset to test whether the model could be used over various field conditions.

2.6 Calibration and Validation of Cultivar-Specific Models to Predict LAI and H_{can}

The calibration and validation process was also performed per cultivar to test whether cultivar-specific models improve the general model. The calibration and validation datasets remained the same (Table 1); however, we divided the datasets into smaller groups based on cultivar. Only three cultivars were considered for the field variety tests: CL XL745, XP753, and Gemini214 CL. For the CL XL745 cultivar, six field-years were included in the calibration dataset, while the validation dataset was composed of 4 field-years. For the XP753 cultivar, three field-years were included in each calibration and validation. For Gemini214 CL, the calibration and validation datasets each contained data from 4 field-years. For model comparison, both the cultivar-specific model and the general model were applied to the validation datasets to establish a comparison and detect whether these cultivar-specific models improve the prediction of LAI or H_{can}.

2.7 Evaluating the Effectiveness of Peak LAI as a Yield Substitute

Additional models were examined to predict yield values for the growing seasons observed. To create these models, the Peak LAI value from each field was determined either as the maximum observed or estimated in the sigmoid model using GDD as the time series (Waldner et al., 2019). Due to the sparse time series data with LAI measurements taken every 1-2 weeks, using simple metrics to create a yield prediction model can help predict harvest conditions and amounts.

3. Results

3.1 Data Collection and Model Selection

3.1.1 Data Collection

The average differences and standard deviations of the LAI and H_{can} were calculated to test the accuracy of the field measurements. The two LAI measurements for individual fields differed by 0.20-1.03 m^2m^{-2}, with an average standard deviation of 0.364 m^2m^{-2}. The average differences in height between the five initial H_{can} measurements were typically between 6-10 cm at each field for each sampling date. The average standard deviation between these measurements was 2.97 cm.

3.1.2 GDD and DAP Models

The relationships between LAI and H_{can} as a function of GDD and DAP were examined for the 2018, 2019, and 2020 growing seasons using polynomial and sigmoid-curve regressions (Figure 4, Appendix B, C, D, and E). Table 3 displays the final regression equations for both of these models. Overall, GDD was the preferred time series choice for polynomial and sigmoid regression models, with higher R^2 values and lower RMSE and pbias values reported for the GDD models (Table 4). However, for some of the individual fields' models, the R^2 values remained the same, leading to inconclusive results concerning choosing the best fit model (Appendix B and C). Cases where this occurred were most likely due to fewer data points available for use within the creation of the regression. Overall, the 2019 datasets performed best due to more consistent field measurements taken when compared to values measured in the 2018 and 2020 growing seasons. This finding shows how vital data collection frequency is for creating a prediction model.

There was an average increase of 11.3% in the R^2 value and a decrease of 14.3% in the RMSE value when observing the polynomial LAI graphs for GDD (R^2 = 0.75, RMSE = 1.17 m^2m^{-2}) compared to the model using DAP (R^2 = 0.67, RMSE = 1.35 m^2m^{-2}). Similarly, there was an average R^2 increase of 8.6% and a decrease of about 20.5% in the RMSE value when observing the polynomial H_{can} graphs (GDD: R^2 = 0.85, RMSE = 13.6 cm, DAP: R^2 = 0.78, RMSE = 16.7 cm). This improvement in the models confirms that applying a temperature-dependent time series to models increases the overall model accuracy regarding LAI and Hcan prediction for polynomial and sigmoid regression models.

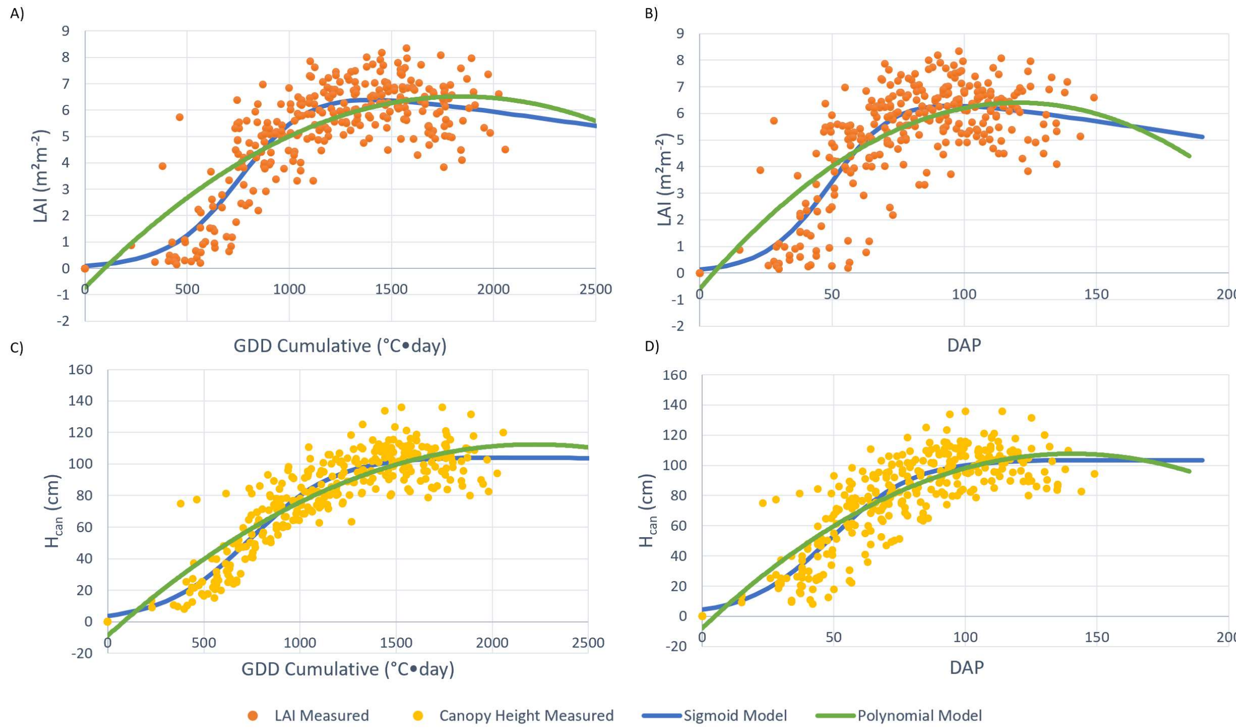

Figure 4: For the 2018, 2019, and 2020 growing seasons, the following relationships were fit to both polynomial and sigmoid-curve regressions: (A) Leaf Area Index (LAI) v. Growing Degree Days (GDD), (B) LAI v. Days after Planting (DAP), (C) Canopy Height (H_{can}) v. GDD, and (D) H_{can} v. DAP. These comparisons were tested using two different time series: Growing Degree Days (A, C) and Days after Planting (B, D).

Table 3: Fitted equations from the polynomial and sigmoid regressions for GDD and DAP with respect to LAI and H_{can} for the datasets, including information from every field

Polynomial Regression	
Model	*Equation*
LAI v. GDD	$f(x) = -2x10^{-6}x^2 + 0.0078x - 0.7525$
H_{can} v. GDD	$f(x) = -2x10^{-5}x^2 + 0.1095x - 8.9341$
LAI v. DAP	$f(x) = -0.0005x^2 + 0.1125x - 0.5386$
H_{can} v. DAP	$f(x) = -0.0058x^2 + 1.622x - 7.3904$
Sigmoid Regressions	
Model	*Equation*
LAI v. GDD	$f(x) = 10.3 \cdot \left(1 + \exp\left[\frac{-(x - 792.8)}{181.6}\right]\right)^{-1} \cdot \left(1 - (1 + \exp\left[\frac{-(x - 2415.1)}{1776.2}\right])^{-1}\right)$
H_{can} v. GDD	$f(x) = 285.7 \cdot \left(1 + \exp\left[\frac{-(x - 777.6)}{232.6}\right]\right)^{-1} \cdot \left(1 - (1 + \exp\left[\frac{-(x + 4233.6)}{9664.2}\right])^{-1}\right)$
LAI v. DAP	$f(x) = 9.42 \cdot \left(1 + \exp\left[\frac{-(x - 50.5)}{12.33}\right]\right)^{-1} \cdot \left(1 - (1 + \exp\left[\frac{-(x - 217.3)}{156.4}\right])^{-1}\right)$
H_{can} v. DAP	$f(x) = 122.7 \cdot \left(1 + \exp\left[\frac{-(x - 48.8)}{15.55}\right]\right)^{-1} \cdot \left(1 - (1 + \exp\left[\frac{-(x - 4728.1)}{2929}\right])^{-1}\right)$

3.1.3 Polynomial and Sigmoid Regression Models

Since the two models differed in the number of initial inputs, AIC values were calculated for the polynomial and sigmoid regressions to select the model that best fits the data while accounting for the cost of increasing parameters. On average, there was a 112.4% decrease in the AIC value for the LAI v. GDD model using a sigmoid regression (AIC = 14.97), compared to the polynomial regression (AIC = 53.35) (Table 4). However, there was only a 1.64% decrease in the sigmoid AIC value (AIC = 826.6) compared to the polynomial AIC value (AIC = 840.3) for the H_{can} v. GDD model. The sigmoid model was thus the preferred regression since it decreased AIC values for both LAI and H_{can} models.

Table 4: Descriptive statistics for polynomial and sigmoid regression models, comparing GDD and DAP to LAI and H_{can}

	Polynomial Regression				**Sigmoid-Curve Regression**				
Model	*R^2*	*RMSE* (m^2m^{-2}/cm)	*pbias*	*AIC*	*R^2*	*RMSE* (m^2m^{-2}/cm)	*pbias*	*% Diff.*	*AIC*
LAI v. GDD	0.75	1.17	2.2×10^{-14}	53.35	0.82	1.01	0.13	2.70	14.97
LAI v. DAP	0.67	1.35	3.9×10^{-14}	840.3	0.71	1.27	-1.87	-0.14	826.6
H_{can} v. GDD	0.85	13.6	1.4×10^{-13}	94.53	0.88	12.7	4.22	2.91	83.01
H_{can} v. DAP	0.78	16.7	3.8×10^{-14}	908.7	0.80	16.0	0.28	-1.20	900.33

*Note: All p-values for these regressions were < 0.001 and were therefore not included in the table. The % difference values for the polynomial regression had values that were << 0.001 and were also not included in the table.

Additionally, the sigmoid models were preferred by the R^2 and RMSE metrics. The R^2 increased by approximately 8.92% and the RMSE decreased by 14.7% when applying the sigmoid regression to compare LAI and GDD (Polynomial: R^2 = 0.75, RMSE = 1.17 m^2m^{-2}; Sigmoid: R^2 = 0.82, RMSE = 1.01 m^2m^{-2}). Similarly, for the H_{can} and GDD comparison, the R^2 increased by 3.47% and the RMSE decreased by about 6.8% (Polynomial: R^2 = 0.85, RMSE = 13.6 cm; Sigmoid: R^2 = 0.88, RMSE = 12.7 cm). The slight increases in the coefficients of determination and decreases in the RMSE values for both LAI and H_{can} models demonstrate the improvements when applying a sigmoid regression model to the datasets. The descriptive statistics for each field's polynomial models are shown in Appendix B and C for the 2018 and 2019 growing seasons. Appendix D and E contain the same information for the sigmoid models.

3.2 Calibration and Validation of Models

3.2.1 Calibration and Validation of the General Models

After initial model selection tests had been completed, the final model form was a sigmoid function with GDD as its time series variable. The sigmoid regressions were fit to the calibration datasets and then applied to the validation sets. The final regression equations are shown in Figure 5, with the final calibrated parameter values located in Appendix F. Between the calibration and validation fits, there was only a drop in the R^2 value of approximately 10% for the LAI models and 7% for the H_{can} models. Thus, the model performed well, and the corresponding modeling procedure was considered good and contained well-selected calibration datasets (Table 5).

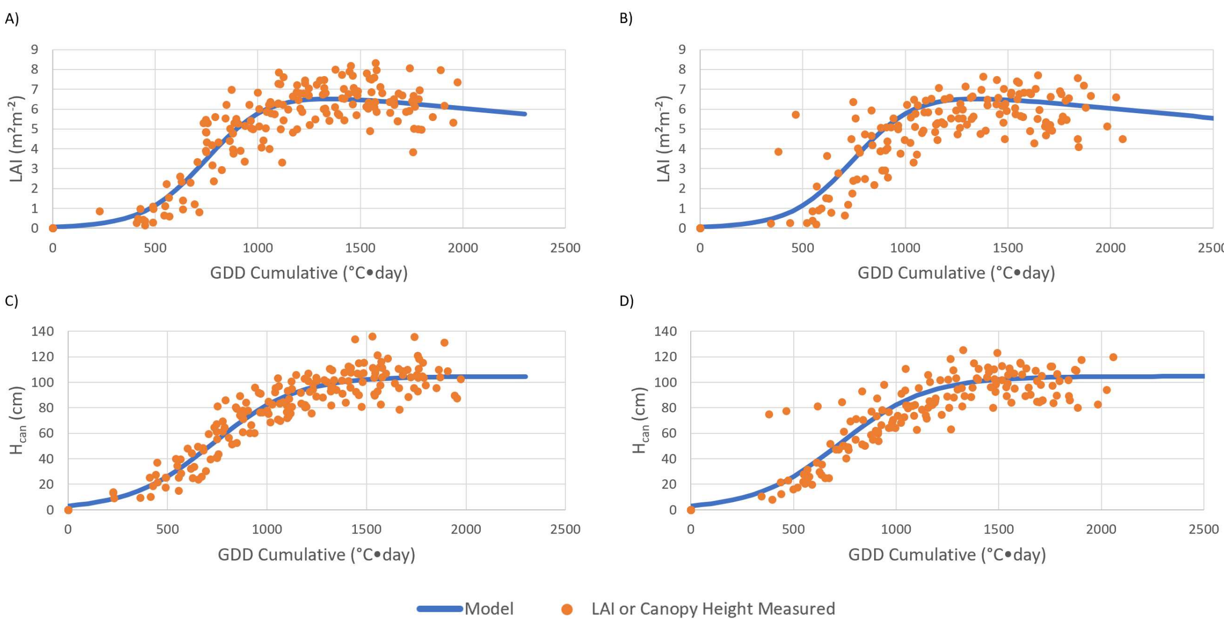

Figure 5: (A) Comparing LAI and GDD for the model of the calibration dataset; (B) Comparing LAI and GDD for the model applied over the validation dataset; (C) Comparing H_{can} and GDD for the model of the calibration dataset; (D) Comparing H_{can} and GDD for the model over the validation dataset

Similarly, the RMSE values for the calibration datasets were lower than those from the validation datasets. For the calibration of LAI, the RMSE value was 0.92 m^2m^{-2}, while it was 1.15 m^2m^{-2} for LAI for the validation datasets (Table 5). The RMSE value for the calibration comparison of H_{can} v. GDD was 10.9 cm, compared to a value of 13.7 cm for the validation of H_{can} (Table 5). As expected, the validation datasets' pbias and percent difference values are also higher than those for the calibration datasets.

Table 5: Descriptive statistics for general calibration and validation datasets for both LAI v. GDD and H_{can} v. GDD

Calibration Datasets				
Model	*R^2*	*RMSE* (m^2m^{-2}/cm)	*pbias*	*% Diff.*
LAI v. GDD	0.86	0.92	0.10	2.41
H_{can} v. GDD	0.91	10.9	0.19	-1.42
Validation Datasets				
Model	*R^2*	*RMSE* (m^2m^{-2}/cm)	*pbias*	*% Diff.*
LAI v. GDD	0.78	1.15	-8.52	-5.49
H_{can} v. GDD	0.85	13.7	-3.97	-5.20

3.2.2 Calibration and Validation of Cultivar-Specific Models

After the general models were calibrated and validated, we tested whether the rice cultivar-specific model formulations improved model accuracy. The three main cultivars observed were CL XL745, XP753, and Gemini214 CL. Models were fit to the calibration datasets using the general model for each cultivar, and a new model was calibrated using only values from the cultivar-specific dataset (Figure 6). With most models having an inflection point around a GDD value of 700-750 °C•day, most cultivars had the most phenological change within this short period. The final calibrated parameters for these models are located in Appendix G.

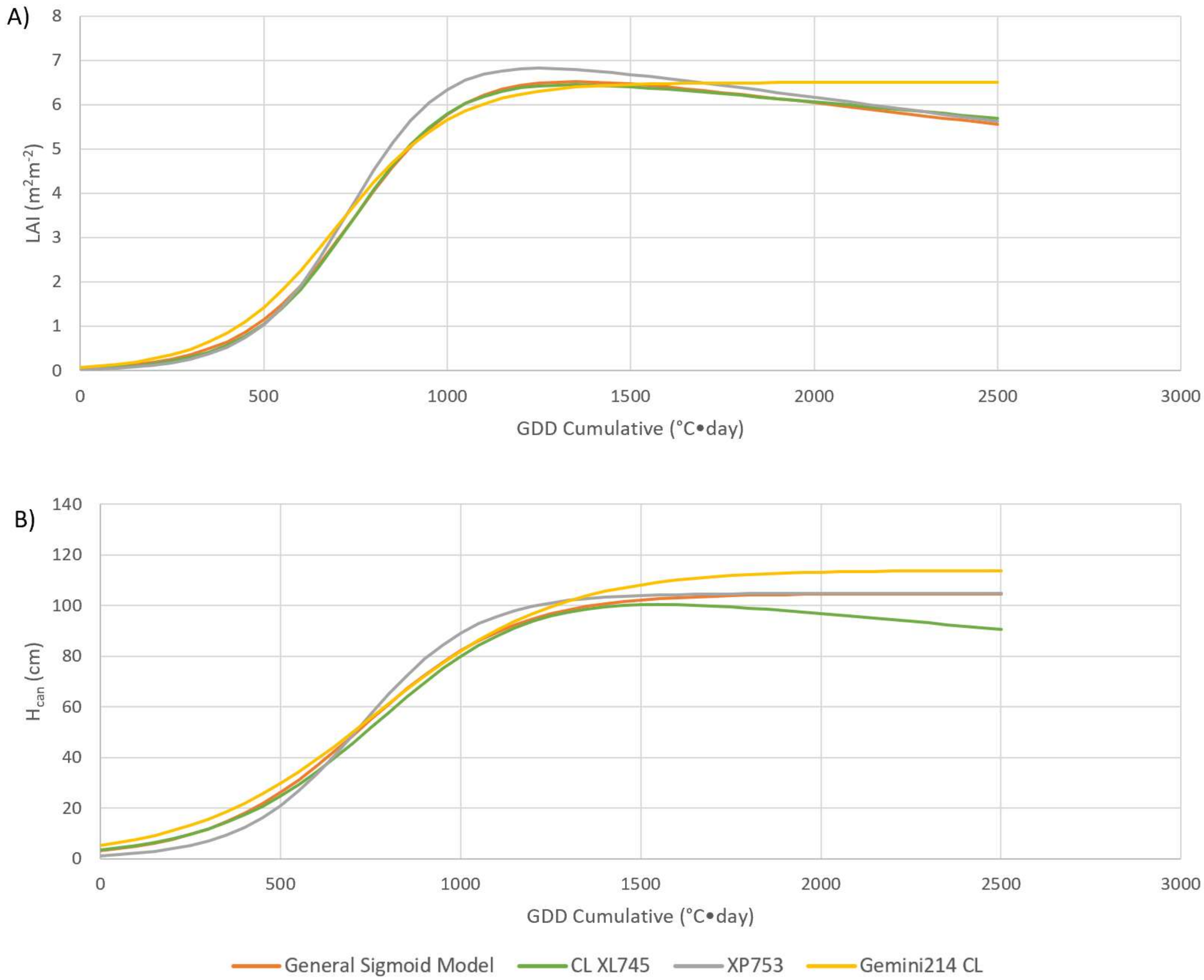

Figure 6: (A) Comparison of LAI and GDD for the general polynomial model (blue), general sigmoid model (orange), the CL XL745 cultivar-specific model (green), the XP753 cultivar-specific model (gray), and the Gemini214 CL cultivar-specific model (yellow); (B) Comparison of H_{can} and GDD for the general polynomial model (blue), general sigmoid model (orange), the CL XL745 cultivar-specific model (green), the XP753 cultivar-specific model (gray), and the Gemini214 CL cultivar-specific model (yellow).

The calibration equations were then applied to the validation datasets to test the accuracy of each of the models. For LAI models, the cultivar-specific model performed only slightly better for the Gemini214 cultivar (LAI: R^2= 0.73, RMSE= 1.21 m^2m^{-2}), compared to the general model (LAI: R^2= 0.72, RMSE= 1.22 m^2m^{-2}) (Table 6). For the XL745 cultivar, the cultivar-specific and general models had almost identical R^2 values (Cultivar-specific and General: R^2 = 0.82), but the cultivar-specific model had a slightly lower RMSE value (Cultivar-Specific: RMSE =

0.99 m^2m^{-2}; General: RMSE = 1.00 m^2m^{-2}) (Table 5). The general model performed better for the XP753 cultivar (LAI: R^2= 0.84, RMSE = 1.28 m^2m^{-2}) compared to the cultivar-specific model (LAI: R^2= 0.83, RMSE = 1.50 m^2m^{-2}) (Table 6). However, the model for the XP753 cultivar did appear to overpredict the model for the validation dataset, which could be a result of a poor representative dataset chosen for this calibration.

Similarly, for the H_{can} models, the cultivar-specific model provided slight model improvement for the Gemini214 cultivar (H_{can}: R^2= 0.80, RMSE= 16.2 cm), while the general model for Gemini214 did not perform as well containing lower R^2 and higher RMSE values (H_{can}: R^2= 0.79, RMSE= 16.6 cm) (Table 5). The cultivar-specific model performed similar to its general model for the XP753 cultivar (Cultivar-Specific: R^2= 0.90, RMSE = 14.2 cm; General: R^2= 0.90, RMSE = 12.3 cm) (Table 5). The general model also reported slightly better results that for the cultivar-specific model for the CL XL745 cultivar (General: R^2= 0.89, RMSE= 14.2 cm; Cultivar-specific: R^2= 0.86, RMSE= 13.5 cm) (Table 5). Since the R^2 and RMSE values are within a tight range, it is not clear if the general or cultivar-specific model is preferred due to the mixed statistical results from the CL XL745 and XP753 cultivars, which contained higher R^2 values for the general models but lower RMSE values for the cultivar-specific models.

Table 6: Descriptive statistics for validation datasets for specific field cultivar model and all fields model for both LAI v. GDD and H_{can} v. GDD

LAI v. GDD								
	Cultivar-Specific Model				**General Model**			
Cultivar	R^2	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*	R^2	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*
XL745	0.82	0.99	-5.46	-2.56	0.82	1.00	-6.14	-3.22
XP753	0.83	1.50	-28.7	-24.0	0.84	1.28	-22.6	-19.5
Gemini214	0.73	1.21	-4.42	-1.28	0.72	1.22	-2.82	0.89
H_{can} v. GDD								
	Cultivar-Specific Model				**General Model**			
Cultivar	R^2	*RMSE* (cm)	*pbias*	*% Diff.*	R^2	*RMSE* (cm)	*pbias*	*% Diff.*
XL745	0.91	10.59	-1.86	-3.29	0.91	11.0	-4.98	-6.66
XP753	0.90	13.1	-8.55	-10.5	0.90	12.3	-5.97	-8.29
Gemini214	0.80	16.2	-0.06	-0.51	0.79	16.6	4.26	4.37

*Note: All p-values for these regressions were < 0.001 and were therefore not included in the table

3.3 Evaluating the Effectiveness of Peak LAI as a Yield Substitute

The relationships between the two peak LAI values (measured and modeled) and yield (Figure 7) show that it is not viable only to use peak LAI to estimate yield since the coefficients of determination for these tests are less than 0.01. Additionally, an ANOVA test was performed to determine if the linear regressions were significant. Each test revealed an insignificant p-test value, which indicated that this model does not allow for an accurate prediction of yield. The poor relationship between the peak LAI and yield could be due to the grain-filling period after the LAI has peaked.

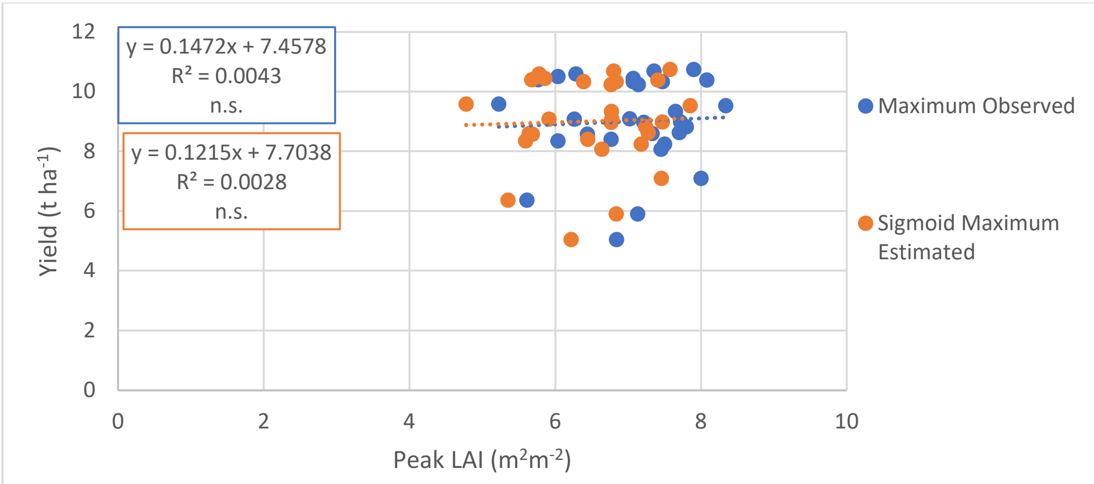

Figure 7: Comparison of Yield to observed and estimated Peak LAI

4. Discussion

4.1 Model Selection

4.1.1 Comparing GDD Models and DAP Models

Sigmoid models created with GDD as the time series consistently had higher coefficients of determination and lower RMSE and pbias values than models using DAP (Table 4). However, this pattern was not present when comparing some of the individual fields' GDD and DAP models.

GDD allows for the consideration of temperature within the growing season, which creates a more accessible application of these models to different regions worldwide. Rice is a temperature-dependent crop, so including climatic parameters within the prediction model assists in better explaining other components of the model, i.e., early-season yield predictions and effects from extreme weather events. However, in other studies (Waldner et al., 2019), including weather parameters helped with models containing fewer parameters but did not have any beneficial effect otherwise. Additionally, there have been challenges identified with empirical models not being easily applied to more generalized scales. It is more difficult to use them in areas with specific crop cultivars, crop growth stages, geographical regions, etc. (Waldner et al., 2019). However, the results of this study display generalizability of LAI and H_{can} models in Arkansas.

4.1.2 Comparing Polynomial and Sigmoid Regression Models

The sigmoid regression functions tended to fit more accurately to the datasets than when using polynomial regression. The models consistently recorded higher coefficients of determination and lower RMSE values when using a sigmoid regression (Table 4). However, the polynomial regressions did report lower pbias values, which may be due to more significant differences between the measured and estimated LAI or H_{can} values for the general sigmoid model (Table 4). When considering the models for individual fields, the Excel solver had more difficulties creating an accurate model. These difficulties could result from the model requiring more accurate initial guesses for the parameters, likely due to fewer data points available than the general models (Appendix D and E). Boschetti et al., 2018 also encountered the challenge of creating precise, individual models, especially when different crops and varieties are present in a small area. Although the polynomial regression was easier to apply in Excel, the datasets recurringly behaved in an s-curve type manner, making the sigmoid regression a better choice for creating a representative model when including the entire data series. Although the sigmoid function did not always fit well for the individual fields' models, the sigmoid model did consistently fit better when including all the data.

Although the pbias in the sigmoid model are still relatively low (approximately <10%), the lower pbias from the polynomial regressions than the sigmoid regressions warrant further examination. With limited variability within the pbias and percent difference values for the polynomial regression, having fewer initial parameters into a model could help create minor variation and bias throughout the datasets when the model is applied. However, the decreases in AIC values when comparing sigmoid and polynomial regressions suggest it is better to include more parameters. Future work could simplify the sigmoid-curve regression equation to better represent the spread of the data throughout the growing season.

4.2 Calibration and Validation of the Models

A better understanding of how each individual cultivar behaves within the growing season was developed by dividing the data into calibration and validation datasets based on field cultivar. Comparing the calibrated parameter values for the different cultivar-specific models helped conclude the individual cultivars and their relationships with the other parameters. Most models showed an inflection point around a GDD value of 700-750 °C•day; perhaps, more frequent measurements could be taken within the beginning to the middle of the growing season to better predict the rapid changes the cultivars experience in future work.

The best performing cultivar models for LAI v. GDD and H_{can} v. GDD were the Gemini214 CL models, which improved the R^2 values by approximately 1.38% and 1.26% for the LAI and H_{can} models, respectively. However, the models continue to perform similarly for the other two cultivars, with only slight differences between the R2 and RMSE values for the cultivar-specific and general models.

Creating models for specific cultivars can aid in reducing uncertainty through the consideration of field information particular to each cultivar, i.e., time spent in each rice growth stage and ideal growing temperatures. Although their model used remotely-sensed LAI data, Gilardelli et al. (2019) found that separating the fields into sub-sections assisted in creating a better representative model. However, they noted that more cultivar-specific calibrations for assimilating LAI data were still needed. Recalibrating our data based on cultivar conditions made it difficult to observe whether cultivar-specific models are needed or if more research of initial model parameters may be required to predict LAI and Hcan for specific cultivars better. Overall, the general model performed well for the combined dataset and had similar results to the cultivar-specific models for some cases. These results suggest a general model can simplify LAI and H_{can} predictions and ultimately allow for one model to be applied across Arkansas over various cultivars.

4.3 Evaluating the Effectiveness of Peak LAI as a Yield Substitute

In a study on wheat crops in Australia, the integration of peak LAI into yield prediction models has proven best for early season grain predictions (Waldner et al. 2019). However, there was no significant relationship when completing our peak LAI tests to yield. This lack of correlation was also apparent in a study conducted in Northern Italy, where researchers could not define any direct relationships between maximum LAI and final yield for almost all the observed varieties (Gilardelli et al., 2019). Additionally, another study observing correlations between crop yields and satellite-obtained LAI data indicated no significant relationship between LAI and rice yields (Johnson, 2016). More parameters may need to be considered within the peak LAI model to use it as a viable yield prediction option.

4.4 Extension of Research

To represent the climate more accurately within Arkansas, the GDD values could be recalculated by changing the value where the maximum temperature is capped from 30°C to approximately 34°C. By changing this temperature, GDD values will better represent field conditions and more accurately depict the higher temperatures present in Arkansas (Hardke and University of Arkansas (System). Cooperative Extension Service., 2018). In addition, high temperatures can result in lower yields due to heat stress (Mendez et al., 2021). For example, nighttime temperature fluctuations have been reported to drop rice yield by 10% for each increase in 1°C (Peng et al., 2004). GDD does not account for the penalization of high temperatures and their negative effect on crop growth, limiting the use of GDD units within models.

Improving the datasets by removing outliers could also improve model fit. We will go back through each field dataset to identify if any data points caused the model to fit the dataset poorly. If any consistent responses are observed throughout the data, these points will be eliminated from the model, allowing for a more accurate and precise general model. Additionally, implementing satellite data, i.e., NDVI, can assist this modeling process. NDVI and LAI have been found to have a close relationship, which could allow for gaps within data to be filled. By considering NDVI, it could be possible to obtain more accurate parameters by which the data can be calibrated and even help with the calibration of the regression parameters. LAI estimates can also be retrieved from satellite data (Campos-Taberner et al., 2016; Waldner et al., 2019). . The satellite LAI data could lead to less reliance on manual field measurements and allow for a greater number of LAI data points to be used for model calibration purposes.

5. Conclusion

With higher R^2 values and lower RMSE values present for the GDD-based models, GDD proved to be a more accurate time series to use when creating a general prediction model compared to the models using DAP as its time series. Considering the recurring S-curve behavior of the datasets when completing initial tests helped create a more precise prediction model due to the increase in R^2 values, lower RMSE values, and an overall decrease in the AIC values. The sigmoid model created with the calibration dataset was applied to the validation dataset, which proved to effectively model both LAI and H_{can}, with R^2 values of 0.78 and 0.85 and RMSE values of 1.15 m^2m^{-2} and 13.7 cm, respectively. Overall, the data suggest a general model can be applied to predict LAI or H_{can}. However, more work still needs to be completed to see the benefit of having cultivar-specific models.

Further work on simplifying the general sigmoid model still needs to be completed by either eliminating outliers or simplifying the overall model by reducing the number of parameters used within the sigmoid equation. Incorporating satellite data can also help apply these general models more easily to other rice fields in a precise and accurate manner. The variables modeled in this work can be used as inputs for other more extensive and sophisticated models, such as yield prediction models. As food scarcity becomes a more significant issue, the use of prediction models based on temperature and simple measurement parameters similar to the ones created in this study will be beneficial for ensuring a more sustainable and reliable food source as the global population increases.

6. Acknowledgments

Ground data were collected with support from Unilever, the Knorr Sustainability Partnership Fund (Project Water Sustainability in Arkansas rice fields), and the USDA-NRCS (under the Cooperative Agreement 68-7103-17-119). We are grateful to the rice farmers for hosting our experiments. We also thank Colby Reavis, Micaela Wine, Anthony Zadoorian, Juan Arguijo, Sam Carroll, Bennett Barr, and Connor Pearson for their help in field data collection.

I would like to thank Dr. Benjamin Runkle and Dr. Beatriz Moreno-García for the countless hours of work and assistance throughout this entire modeling and writing process. I would not have completed any of this work without their help. I would also like to thank Dr. Brian Haggard and Dr. Kusum Naithani for their work and time on my honors thesis defense committee. I would also like to acknowledge the Honors College for their financial support of this work through an Honors College research grant.

7. References

2013. gSSURGO MapUnit FY 2013. Arkansas GIS Office: U.S. Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control* 19(6):716-723.

Belder, P., B. Bouman, R. Cabangon, L. Guoan, E. Quilang, L. Yuanhua, J. Spiertz, and T. Tuong. 2004. Effect of water-saving irrigation on rice yield and water use in typical lowland conditions in Asia. *Agricultural Water Management* 65(3):193-210.

Biosciences, L.-C. 2013. *LAI-2200C: Plant Canopy Analyzer Instruction Manual*. LI-COR, Inc., Lincoln, NE.

Boschetti, M., L. Busetto, L. Ranghetti, J. G. Haro, M. Campos-Taberner, and R. Confalonieri. 2018. Testing Multi-Sensors Time Series of Lai Estimates to Monitor Rice Phenology: Preliminary Results. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE.

Campos-Taberner, M., F. J. García-Haro, G. Camps-Valls, G. Grau-Muedra, F. Nutini, A. Crema, and M. Boschetti. 2016. Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring. *Remote Sensing of Environment* 187:102-118.

Chidthaisong, A., N. Cha-un, B. Rossopa, C. Buddaboon, C. Kunuthai, P. Sriphirom, S. Towprayoon, T. Tokida, A. T. Padre, and K. Minamikawa. 2018. Evaluating the effects of alternate wetting and drying (AWD) on methane and nitrous oxide emissions from a paddy field in Thailand. *Soil Science and Plant Nutrition* 64(1):31-38.

Childs, N. 2021. Rice Sector at a Glance. USDA. Available at: https://www.ers.usda.gov/topics/crops/rice/rice-sector-at-a-glance/.

Colaizzi, P. D., S. R. Evett, D. K. Brauer, T. A. Howell, J. A. Tolk, and K. S. Copeland. 2017. Allometric Method to Estimate Leaf Area Index for Row Crops. *Agronomy Journal* 109(3):883-894.

Gilardelli, C., T. Stella, R. Confalonieri, L. Ranghetti, M. Campos-Taberner, F. J. García-Haro, and M. Boschetti. 2019. Downscaling rice yield simulation at sub-field scale using remotely sensed LAI data. *European journal of agronomy* 103:108-116.

Group, P. C. 2019. PRISM Climate Data. Oregon State University.

Hammer, G. 1997. Crop modelling: current status and opportunities to advance. *II Modelling Plant Growth, Environmental Control and Farm Management in Protected Cultivation 456*:27-36.

Hardke, J. T., and University of Arkansas (System). Cooperative Extension Service. 2018. *Arkansas Rice Production Handbook*. Mp No. 192. Cooperative Extension Service, University of Arkansas, Little Rock, Ark.

Henry, C., M. Daniels, M. Hamilton, and J. Hardke. 2021. Water Management. In *Rice Production Handbook*, 103-123. J. T. Hardke, ed. Little Rock, AR: University of Arkansas Division of Agriculture.

Johnson, D. M. 2016. A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. *International Journal of Applied Earth Observation and Geoinformation* 52:65-81.

Kobayashi, H., Y. Ryu, D. D. Baldocchi, J. M. Welles, and J. M. Norman. 2013. On the correct estimation of gap fraction: How to remove scattered radiation in gap fraction measurements? *Agricultural and Forest Meteorology*:170-183.

LaHue, G. T., R. L. Chaney, M. A. Adviento-Borbe, and B. A. Linquist. 2016. Alternate wetting and drying in high yielding direct-seeded rice systems accomplishes multiple environmental and agronomic objectives. *Agriculture, ecosystems & environment* 229:30-39.

Lampayan, R., B. Bouman, J. De Dios, A. Lactaoen, A. Espiritu, T. Norte, E. Ouilang, D. Tabbal, L. Llorca, and J. Soriano. 2004. Adoption of Water Saving Technologies in Rice Production in the Phillipines. *Food and Fertilizer Technology Center*.

Lampayan, R. M., R. M. Rejesus, G. R. Singleton, and B. A. Bouman. 2015. Adoption and economics of alternate wetting and drying water management for irrigated lowland rice. *Field Crops Research* 170:95-108.

Liu, Z., G. Zhang, and P. Yang. 2016. Geographical Variation of Climate Change Impact on Rice Yield in the Rice-Cropping Areas of Northeast China during 1980–2008. *Sustainability* 8(7):670.

Matthews, R. B., M. Rivington, S. Muhammed, A. C. Newton, and P. D. Hallett. 2013. Adapting crops and cropping systems to future climates to ensure food security: The role of crop modelling. *Global Food Security* 2(1):24-28.

McMaster, G. S., and W. Wilhelm. 1997. Growing degree-days: one equation, two interpretations.

Mehdi, B., B. Lehner, and R. Ludwig. 2018. Modelling crop land use change derived from influencing factors selected and ranked by farmers in North temperate agricultural regions. *Science of the Total Environment* 631:407-420.

Mendez, K., M. A. Adviento-Borbe, A. Lorence, and H. Walia. 2021. Significant Shift of Ambient Night-Time Air Temperature during Rice Growing Season in Major US Rice States. *American Journal of Climate Change* 10(1):134-151.

Muthayya, S., J. D. Sugimoto, S. Montgomery, and G. F. Maberly. 2014. An overview of global rice production, supply, trade, and consumption. *Annals of the new york Academy of Sciences* 1324(1):7-14.

Nalley, L., B. Linquist, K. Kovacs, and M. Anders. 2015. The economic viability of alternative wetting and drying irrigation in Arkansas rice production. *Agronomy Journal* 107(2):579-587.

Peng, S., J. Huang, J. E. Sheehy, R. C. Laza, R. M. Visperas, X. Zhong, G. S. Centeno, G. S. Khush, and K. G. Cassman. 2004. Rice yields decline with higher night temperature from global warming. *Proceedings of the National Academy of Sciences* 101(27):9971-9975.

Runkle, B., E. Ward, L. Windham-Myers, Y. Ryu, M. Kang, S. Bansal, R. B. Jackson, G. McNicol, S. H. Knox, W. J. Riley, A. Lohila, K. V. Schafer, R. Vargas, and M. L. Reba. 2019. How drying and wetting events impact landscape methane fluxes.

Shah, F., J. Huang, K. Cui, L. Nie, T. Shah, C. Chen, and K. Wang. 2011. Impact of high-temperature stress on rice plant and its traits related to tolerance. *The Journal of Agricultural Science* 149(5):545-556.

Sharifi, H., R. J. Hijmans, J. E. Hill, and B. A. Linquist. 2017. Using Stage-Dependent Temperature Parameters to Improve Phenological Model Prediction Accuracy in Rice Models. *Crop Science* 57(1):444-453.

Stroppiana, D., M. Boschetti, R. Confalonieri, S. Bocchi, and P. A. Brivio. 2006. Evaluation of LAI-2000 for leaf area index monitoring in paddy rice. *Field Crops Research* 99(2-3):167-170.

USDA-NRCS. SSURGO Database. National Resources Conservation Service.

Waldner, F., H. Horan, Y. Chen, and Z. Hochman. 2019. High temporal resolution of leaf area data improves empirical estimation of grain yield. *Scientific reports* 9(1):1-14.

8. Appendices

A. LAI Scattering Corrections

For each location, different readings above and below the canopy were measured following a sequence already programmed within the LAI-2000 or LAI-2200C devices. The sequence was:

4A 4B 1A 3B for the LAI 2000

4A 4B 1A 4B 1A 4B 1A for the LAI-2200C,

where A is a measurement above the canopy, and B is a measurement below the canopy, i.e., close to the ground in case there is no ponding water, or above the water level in case there is ponding water. The sequence includes extra sky radiation measurements that will be used by the software FV2200 to perform scattering corrections. We used the 4A sequence using the white diffuser cap (LI-COR (Biosciences, 2013). For the normal above and below readings, the 90° cap was used to hide the operator from the sensor.

Figure 10: 90° cap used on LAI-2200C to hide the operator from the sensor

LAI files were transferred to the computer in order to perform the scattering corrections using the software FV2200. Files from the LAI-2000 were first converted to LAI-2200C format before the scattering correction process. One of the traditional underlying assumptions of the LAI-2000 and LAI-2200C has been that foliage absorbs all the radiation in the wave and seen by the sensor (320-490 nm). Starting with version 2.0, FV2200 allows this assumption to be set aside and provides a mechanism (Kobayashi et al., 2013), for correcting measurements for the radiation reflected and transmitted by the foliage (LI-COR (Biosciences, 2013).

B. Comparing the Polynomial Regression Equations and R² Values when using GDD Cumulative and Days After Planting for 2018:

	LAI v. GDD					**LAI v. DAP**				
Field Name	*R^2*	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*	*P-value*	*R^2*	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*	*P-value*
R1	0.89	0.60	-2.74×10^{-13}	-3.72×10^{-14}	< 0.001	0.88	0.61	4.66×10^{-15}	0	< 0.001
R2	0.89	0.72	4.48×10^{-15}	0	< 0.001	0.88	0.74	-1.65×10^{-14}	-1.52×10^{-14}	< 0.001
R3	0.79	0.99	7.69×10^{-14}	7.09×10^{-14}	0.009	0.78	1.02	-2.36×10^{-14}	-1.77×10^{-14}	0.099
R4	0.92	0.65	1.78×10^{-14}	2.84×10^{-14}	0.004	0.92	0.67	1.42×10^{-14}	4.26×10^{-14}	0.004
R5	0.84	0.75	3.16×10^{-14}	1.65×10^{-14}	0.031	0.82	0.79	4.12×10^{-15}	1.65×10^{-14}	0.018
R7	0.94	0.53	0	0	< 0.001	0.94	0.52	3.24×10^{-14}	3.01×10^{-14}	< 0.001
R8	0.75	0.81	-9.48×10^{-15}	0	0.013	0.75	0.81	4.74×10^{-15}	0	0.015
R9	0.89	0.75	-2.10×10^{-15}	0	< 0.001	0.88	0.79	4.20×10^{-15}	0	< 0.001
R10	0.95	0.49	1.04×10^{-14}	0	0.007	0.94	0.50	5.96×10^{-15}	0	< 0.001
W3	0.84	0.90	2.02×10^{-14}	1.56×10^{-14}	< 0.001	0.83	0.93	-2.41×10^{-14}	-3.12×10^{-14}	< 0.001
W4	0.81	0.87	-4.66×10^{-14}	-5.12×10^{-14}	< 0.001	0.80	0.90	-2.21×10^{-14}	-3.41×10^{-14}	0.003
CN	0.71	1.08	2.45×10^{-14}	0	0.005	0.70	1.10	3.84×10^{-14}	3.68×10^{-14}	0.005
CS	0.78	1.00	4.37×10^{-14}	5.08×10^{-14}	0.005	0.77	1.03	-4.93×10^{-15}	0	0.005
All Fields	0.79	0.98	-3.54×10^{-14}	-3.31×10^{-14}	< 0.001	0.77	1.02	-9.04×10^{-15}	-4.97×10^{-14}	< 0.001

	H_{can} v. GDD					**H_{can} v. DAP**				
Field Name	*R^2*	*RMSE* (cm)	*pbias*	*% Diff.*	*P-value*	*R^2*	*RMSE* (cm)	*pbias*	*% Diff.*	*P-value*
R1	0.95	5.30	5.56×10^{-15}	-2.04×10^{-14}	< 0.001	0.95	5.40	2.69×10^{-14}	2.04×10^{-14}	< 0.001
R2	0.95	6.74	-2.12×10^{-14}	-1.54×10^{-14}	< 0.001	0.95	6.45	-1.93×10^{-15}	-1.54×10^{-14}	< 0.001
R3	0.91	11.3	7.69×10^{-14}	7.09×10^{-14}	0.0087	0.90	11.9	3.19×10^{-14}	3.19×10^{-14}	< 0.001
R4	0.96	6.66	7.94×10^{-15}	1.59×10^{-14}	< 0.001	0.96	6.70	-5.56×10^{-15}	0	< 0.001
R5	0.91	8.16	2.44×10^{-14}	1.89×10^{-14}	< 0.001	0.89	9.34	3.14×10^{-15}	0	< 0.001
R7	0.97	7.49	1.02×10^{-14}	0	< 0.001	0.96	7.79	2.63×10^{-15}	0	< 0.001
R8	0.94	7.47	-2.50×10^{-15}	0	< 0.001	0.94	7.26	-1.07×10^{-14}	1.60×10^{-14}	< 0.001
R9	0.96	6.18	-7.03×10^{-15}	0	< 0.001	0.96	6.81	7.57×10^{-15}	0	< 0.001
R10	0.95	6.80	3.43×10^{-14}	3.27×10^{-14}	< 0.001	0.95	6.78	1.31×10^{-14}	1.64×10^{-14}	< 0.001
W3	0.91	11.4	7.27×10^{-16}	2.04×10^{-14}	< 0.001	0.90	12.0	-1.82×10^{-14}	-2.04×10^{-14}	< 0.001
W4	0.89	12.2	-1.06×10^{-14}	0	< 0.001	0.87	12.9	-6.81×10^{-15}	0	< 0.001
CN	0.96	6.74	-3.08×10^{-15}	1.97×10^{-14}	< 0.001	0.96	6.72	-1.03×10^{-14}	0	< 0.001
CS	0.95	6.82	-1.70×10^{-15}	0	< 0.001	0.95	6.72	-6.37×10^{-15}	-2.04×10^{-14}	< 0.001
All Fields	0.80	15.5	6.72×10^{-14}	1.77×10^{-14}	< 0.001	0.77	16.8	6.78×10^{-14}	-3.54×10^{-14}	< 0.001

C. Comparing the Polynomial Regression Equations and R^2 Values when using GDD Cumulative and Days After Planting for 2019

	LAI v. GDD				**LAI v. DAP**			
Field Name	R^2	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*	R^2	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*
R1	0.84	0.92	-2.57	-2.54	0.81	0.97	2.88×10^{-15}	-1.15×10^{-14}
R2	0.88	0.89	6.0×10^{-14}	5.32×10^{-14}	0.82	1.09	-3.54×10^{-15}	0
R3	0.86	0.98	-6.0×10^{-14}	-7.00×10^{-14}	0.83	1.08	8.07×10^{-15}	1.75×10^{-14}
R4	0.92	0.78	-1.52×10^{-14}	-2.03×10^{-14}	0.92	0.80	0	0
R5	0.83	0.75	2.00×10^{-15}	1.30×10^{-14}	0.79	0.83	-1.10×10^{-14}	1.30×10^{-14}
R6	0.93	0.70	2.57×10^{-14}	3.17×10^{-14}	0.92	0.72	1.14×10^{-14}	0
R8	0.88	0.71	-2.22×10^{-14}	-2.01×10^{-14}	0.87	0.74	2.43×10^{-14}	2.01×10^{-14}
R9	0.79	1.00	-3.11×10^{-14}	-2.19×10^{-14}	0.76	1.06	-1.26×10^{-14}	0
R10	0.86	0.89	4.93×10^{-15}	0	0.86	0.90	-1.56×10^{-14}	0
W3	0.90	0.76	1.23×10^{-14}	0	0.89	0.79	-5.94×10^{-15}	0
W4	0.82	0.91	-2.12×10^{-14}	-1.26×10^{-14}	0.81	0.95	0	0
CN	0.81	1.16	-3.51×10^{-14}	-1.86×10^{-14}	0.80	1.19	-1.07×10^{-14}	0
CS	0.85	0.88	1.56×10^{-14}	1.87×10^{-14}	0.84	0.89	3.90×10^{-15}	-1.87×10^{-14}
BK30	0.78	0.94	2.73×10^{-14}	4.86×10^{-14}	0.77	0.95	0	-2.43×10^{-14}
BK50	0.92	0.66	1.28×10^{-14}	1.13×10^{-14}	0.92	0.67	5.64×10^{-15}	-2.26×10^{-14}
All Fields	0.78	1.15	-5.02×10^{-14}	-4.10×10^{-14}	0.71	1.31	-2.75×10^{-14}	0

	H_{can} v. GDD				**H_{can} v. DAP**			
Field Name	R^2	*RMSE* (cm)	*pbias*	*% Diff.*	R^2	*RMSE* (cm)	*pbias*	*% Diff.*
R1	0.90	10.3	-2.57	-2.54	0.87	11.3	1.17×10^{-14}	0
R2	0.93	10.3	-1.31×10^{-14}	-2.03×10^{-14}	0.88	13.1	7.76×10^{-15}	2.03×10^{-14}
R3	0.90	12.4	-1.74×10^{-14}	-1.74×10^{-14}	0.86	14.4	2.12×10^{-14}	3.49×10^{-14}
R4	0.97	6.43	-2.00×10^{-14}	0	0.96	6.86	-2.51×10^{-14}	-2.04×10^{-14}
R5	0.90	10.3	7.99×10^{-15}	2.18×10^{-14}	0.83	13.1	-3.41×10^{-14}	-4.36×10^{-14}
R6	0.91	10.3	1.54×10^{-14}	0	0.89	11.2	1.98×10^{-14}	2.18×10^{-14}
R8	0.93	9.21	1.05×10^{-14}	-2.18×10^{-14}	0.92	9.81	4.36×10^{-14}	2.18×10^{-14}
R9	0.92	10.1	0	-2.10×10^{-14}	0.89	11.8	3.08×10^{-14}	2.10×10^{-14}
R10	0.92	8.89	1.03×10^{-14}	0	0.91	9.37	5.13×10^{-15}	0
W3	0.90	11.2	1.09×10^{-14}	0	0.89	11.8	1.82×10^{-15}	0
W4	0.93	9.03	-4.69×10^{-15}	0	0.92	9.63	-1.88×10^{-14}	-2.03×10^{-14}
CN	0.96	6.94	-2.66×10^{-14}	-2.10×10^{-14}	0.95	7.26	-1.21×10^{-14}	0
CS	0.96	7.06	-1.01×10^{-14}	0	0.95	7.30	-5.62×10^{-15}	0
BK30	0.95	7.17	3.71×10^{-14}	2.05×10^{-14}	0.95	7.28	5.97×10^{-15}	0
BK50	0.95	7.84	5.33×10^{-15}	2.13×10^{-14}	0.95	8.01	-1.33×10^{-14}	0
All Fields	0.89	11.6	-4.17×10^{-14}	0	0.81	15.2	-4.15×10^{-14}	4.11×10^{-14}

*Note: All p-values for these regressions were < 0.001 and were therefore not included in the table

D. Comparing the Sigmoid Regression Equations and R^2 Values when using GDD Cumulative and Days After Planting for 2018:

	LAI v. GDD				LAI v. DAP			
Field Name	*R^2*	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*	*R^2*	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*
R1	0.90	0.56	-0.03	1.97	0.90	0.56	0.01	2.17
R2	0.95	0.50	-0.33	7.32	0.95	0.50	-0.29	7.20
R3	0.96	0.43	-0.09	7.92	0.96	1.49	-0.03	7.98
R4	0.97	0.40	-0.23	13.9	0.97	0.40	-0.22	14.10
R5	0.85	0.74	-0.49	4.91	0.77	0.90	1.01	0.56
R7	0.92	0.62	-0.77	0.64	0.93	0.61	-0.76	0.55
R8	0.42	2.21	35.7	39.7	0.34	2.39	30.1	29.5
R9	0.97	0.36	0.08	-1.83	0.98	0.36	0.03	-1.91
R10	0.51	0.46	-0.44	0.20	0.51	0.45	-0.45	0.16
W3	0.65	0.71	-0.48	12.31	0.61	0.70	-0.39	12.7
W4	0.51	0.73	-0.40	12.50	0.53	0.73	-0.35	12.0
CN	0.71	1.05	-1.05	-2.84	0.79	1.24	2.29	-0.21
CS	0.12	0.72	-0.36	5.84	0.12	0.71	-0.31	5.88
All Fields	0.52	0.97	-0.45	2.61	0.59	1.02	-0.40	1.04

	H_{can} v. GDD				H_{can} v. DAP			
Field Name	*R^2*	*RMSE* (cm)	*pbias*	*% Diff.*	*R^2*	*RMSE* (cm)	*pbias*	*% Diff.*
R1	0.95	5.58	-0.41	1.40	0.95	5.44	-0.32	1.24
R2	0.94	7.34	-0.50	0.29	0.94	7.19	-0.53	0.17
R3	0.98	5.27	-0.31	-2.58	0.98	5.15	-0.34	-2.59
R4	0.95	7.18	-0.54	-2.92	0.96	6.98	-0.49	-2.89
R5	0.98	3.93	-0.07	1.44	0.98	3.93	0.03	1.38
R7	0.97	7.20	-0.59	-3.62	0.97	7.02	-0.53	-3.54
R8	0.92	9.02	-0.79	-2.41	0.92	8.92	-0.71	-2.35
R9	0.98	4.56	-0.19	-1.11	0.98	4.56	0.01	-0.82
R10	0.88	7.68	-0.58	-2.56	0.88	7.50	-0.52	-2.52
W3	0.86	7.16	-0.52	1.61	0.85	7.14	-0.42	1.73
W4	0.89	6.62	-0.25	1.97	0.88	6.58	-0.07	2.16
CN	0.98	7.61	-0.72	-5.69	0.97	7.33	-0.63	-5.60
CS	0.95	7.06	-0.69	-4.68	0.94	6.95	-0.61	-4.62
All Fields	0.88	15.25	-0.35	-1.52	0.86	16.6	-0.31	-1.63

*Note: All p-values for these regressions were < 0.001 and were therefore not included in the table

E. Comparing the Sigmoid Regression Equations and R² Values when using GDD Cumulative and Days After Planting for 2019:

	LAI v. GDD				*LAI v. DAP*			
Field Name	*R^2*	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*	*R^2*	*RMSE* (m^2m^{-2})	*pbias*	*% Diff.*
R1	0.98	0.29	-0.46	0.14	0.98	0.33	-0.53	-4.21
R2	0.98	0.41	-0.50	12.29	0.98	0.39	-0.55	12.3
R3	0.99	0.27	0.40	-2.50	0.99	0.25	0.40	-2.34
R4	0.97	0.47	0.19	2.12	0.97	0.46	0.29	2.20
R5	0.97	0.34	0.04	4.66	0.96	0.35	0.03	4.76
R6	0.99	0.20	-0.07	15.58	0.99	0.19	-0.07	15.73
R8	0.96	0.42	-0.11	3.47	0.96	0.42	-0.08	3.50
R9	0.81	0.78	-0.77	4.69	0.82	0.76	-0.64	4.81
R10	0.85	0.55	-0.40	2.87	0.84	0.54	-0.42	2.84
W3	0.83	0.28	-0.49	-0.63	0.84	0.27	-0.48	-0.68
W4	0.71	0.32	-0.71	1.10	0.71	0.32	-0.68	1.06
CN	0.64	0.31	0.42	-1.75	0.64	0.31	0.42	-1.75
CS	0.79	0.36	0.28	7.61	0.79	0.36	0.26	4.36
BK30	0.23	0.34	0.44	8.45	0.21	0.33	0.38	8.32
BK50	0.84	0.41	-0.40	2.09	0.90	0.42	0.11	-0.52
All Fields	0.82	0.89	-0.10	2.51	0.82	1.15	-0.19	1.80

	H_{can} v. GDD				**H_{can} v. DAP**			
Field Name	*R^2*	*RMSE* (cm)	*pbias*	*% Diff.*	*R^2*	*RMSE* (cm)	*pbias*	*% Diff.*
R1	0.98	3.77	-0.46	3.62	0.99	3.39	-0.33	3.25
R2	0.99	3.46	-0.52	7.26	0.99	2.89	-0.37	8.19
R3	0.99	4.62	-0.29	-1.91	0.98	5.00	0.14	-2.28
R4	0.99	4.43	-0.32	0.37	0.98	4.56	-0.17	-0.13
R5	0.98	4.40	-0.45	-1.82	0.98	4.32	-0.27	-1.51
R6	0.99	2.24	-0.24	-2.32	0.99	3.22	0.16	-3.74
R8	0.96	6.81	-0.33	-2.98	0.96	6.87	0.01	-2.94
R9	0.85	5.17	-0.75	-2.83	0.84	4.52	-0.64	-2.41
R10	0.55	3.35	-0.36	-0.11	0.77	6.47	0.23	-2.21
W3	0.79	5.82	-0.39	-0.63	0.85	6.30	0.05	-3.05
W4	0.86	4.26	-0.44	-0.28	0.89	4.37	-0.13	-0.54
CN	0.92	6.35	-0.66	-4.43	0.92	6.32	-0.58	-4.36
CS	0.91	4.98	-0.49	-3.79	0.91	4.95	-0.47	-3.77
BK30	0.90	5.77	-0.64	-3.39	0.95	5.84	-0.48	-4.61
BK50	0.81	3.22	-0.42	-2.21	0.92	4.65	-0.07	-5.68
All Fields	0.85	9.09	-0.45	-1.54	0.88	13.85	-0.14	-1.80

*Note: All p-values for these regressions were < 0.001 and were therefore not included in the table

F. Sigmoid Regression Final Values from Excel Solver Based on Calibration Datasets

Model	Base	Inflection Point	Spread	Optimum GDD	Optimum Shape
GDD v. LAI	9.78	757.51	152.57	3145.13	2378.60
GDD v. H_{can}	369.37	730.22	208.68	-9.72×10^5	1.05×10^6

G. Sigmoid Regression Final Values from Excel Solver for Cultivar-Specific Models

Model	Base	Inflection Point	Spread	Optimum GDD	Optimum Shape
CL XL745 GDD v. LAI	10.6	749.2	142.5	2999.5	3499.3
CL XL745 GDD v. H_{can}	893.1	799.0	221.1	-2.94×10^4	6962.6
XP753 GDD v. LAI	10.6	740.7	130.0	2794.5	2367.8
XP753 GDD v. H_{can}	219.9	721.5	160.3	-7.04×10^5	7.48×10^6
Gemini214 CL GDD v. LAI	9.70	699.95	158.0	1.11×10^7	1.56×10^7
Gemini214 CL GDD v. H_{can}	314.4	763.3	251.9	-6.00×10^5	1.06×10^6