University of North Dakota
UND Scholarly Commons

Theses and Dissertations

Theses, Dissertations, and Senior Projects

January 2022

An Innovative Method For Estimating Planting Dates Using Remotely Sensed Data

Jacob Edward Zanker

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://commons.und.edu/theses

Recommended Citation

Zanker, Jacob Edward, "An Innovative Method For Estimating Planting Dates Using Remotely Sensed Data" (2022). *Theses and Dissertations*. 4302.
https://commons.und.edu/theses/4302

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact und.commons@library.und.edu.

AN INNOVATIVE METHOD FOR ESTIMATING PLANTING DATES USING REMOTELY SENSED DATA

by

Jacob Edward Zanker
Bachelor of Science, University of North Dakota, 2020

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science

Grand Forks, North Dakota

May
2022

Copyright 2022 Jacob Zanker

This thesis, submitted by Jacob Zanker in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

DocuSigned by:
Jianglong Zhang
2787AED1C19A41B...

Jianglong Zhang

DocuSigned by:
Jon Starr
70F2D0889402482...

Jon Starr

DocuSigned by:
Michael Poellot
1A5B4FC10EDD472...

Michael Poellot

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

DocuSigned by:
Chris Nelson
2E0AF088C733403...

Chris Nelson
Dean of the School of Graduate Studies

4/20/2022

Date

PERMISSION

Title	An Innovative Method for Estimating Planting Dates using Remotely Sensed Data
Department	Atmospheric Sciences
Degree	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Jacob Zanker
April 20, 2022

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor, Dr. Jianglong Zhang, for his patience and support through this process. I also would like to recognize my committee members, Jon Starr and Michael Poellot, for their guidance and flexibility while completing this thesis. Additionally, I am appreciative of Dr. Jeffrey S. Reid for providing some valuable suggestions and comments.

Thank you to my parents and grandparents as well, who have provided tremendous and continual support for my studies over the years.

Lastly, I offer sincere gratitude to my wonderful professors and classmates throughout my time at UND, for their support and friendship throughout my undergraduate and graduate career.

For the funding of research presented, I acknowledge the support of the National Science Foundation under NSF EPSCoR Track-1 Cooperative Agreement OIA #13355466.

ABSTRACT

Accurate planting dates are a critical component of several processes in agricultural research, such as crop modeling, agricultural monitoring, and yield forecasts. Yet, field-scale estimations of yearly planting dates on a larger spatial domain are still a challenging task. Using Normalized Difference Vegetation Index (NDVI) values derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) coupled with Growing Degree Days (GDD) calculated from the North American Regional Reanalysis (NARR) dataset, methods for estimating planting dates over individual fields are attempted over corn fields across North Dakota. Through these methods, mean statewide planting dates from the National Agricultural Statistics Service (NASS) crop progress reports across North Dakota were used for calibration of the developed algorithm for 2012. The coupled satellite-based/GDD method for calculating planting date was then further evaluated against weekly NASS data for the years 2003-2020 with a maximum r^2 of 0.96 between the two variables. A calculated median planting date mean difference of just 1.3 days and root-mean-square-error (RMSE) of 3.6 days was achieved. This contrasts with using a set number of calendar days as a calibration, which resulted in a mean difference of 4.7 days and RMSE of 7.6 days. The results of this study suggest that satellite-based remotely sensed data coupled with high resolution meteorological data has the potential of being applied in estimating field-scale planting dates for agricultural evaluations and modeling efforts.

Chapter 1

Introduction

Monitoring and forecasting of crop growth cycles using crop models is essential for the understanding and prediction of crop yields and soil health, as well as for food security and long-term sustainability. The advent of crop modeling occurred in the 1960s, when modern computing capabilities were first achieved. A desire for hypothesis testing to predict crop performance and ongoing advancements in crop science were also contributing factors. The ability to model how minute changes in farming practices and climate can fundamentally produce consequential results is a key strength (Sinclair and Seligman, 1996). Recently, connections were also drawn between crop models and economic model simulations which incorporate market and policy changes that factor into agricultural production. Thus, crop modeling can significantly aid individual farmers in decision making that leads to higher yields and productivity (e.g. Starr et al., 2019).

Accurate modeling of crop growing cycles requires accurate estimation of key growth parameters. One such required input is the crop planting date, which defines the starting point of the plant growth cycle for each selected crop type for a given region. Although field-specific planting date datasets with a high temporal accuracy and spatial resolution are needed in area (non-point) based crop modeling applications (e.g. Starr et al., 2019), only weekly, state-aggregated ground-level estimation of percentage planted is currently available for a given crop type through farmer surveys. Thus, improvements beyond a simple statewide historical benchmark would potentially allow for adaptations based on localized differences and ongoing climate change trends.

As an alternative, remotely sensed data can be utilized to detect green vegetation covered regions by incorporating the Normalized Difference Vegetation Index (NDVI), from which planting dates can be inferred when combined with local meteorological conditions. The NDVI is computed from measured surface reflectivity at near infrared (NIR) and red spectral channels based on the theory that for green vegetation the photosynthesis process can be detected within the red but not the NIR portions of the spectrum (Tucker, 1979). NDVI values are designed, and often used, for detecting vegetation vigor and drought monitoring applications (e.g. Peters et al., 2002). Among various applications, Moulin et al. (1997) used NDVI for studying phenology and climatological variations in the growing season. Various studies have also assessed storm damage to crops for insurance purposes, such as hail swaths indicated by significantly lower NDVI values (e.g. Gallo et al., 2019).

Vegetated areas correspond to NDVI values greater than zero, with values close to one being found in an area of dense vegetation, such as a rainforest. Near-zero or negative values refer to non-vegetated regions such as bare ground, water or snow-covered areas (e.g. Fontana et al., 2012). As the crop growth cycle experiences planting, growing, maturation and harvest states, NDVI values also go through a cycle from near-zero values at the crop planting state to near one during the peak growing season. Clearly, the temporal changes in NDVI values can be used to detect the plant growth cycle and potentially used in calculating the planting date.

In this study, a method is proposed to estimate the planting date of individual corn fields by combining remotely sensed NDVI data from the Moderate Resolution Imaging

Spectroradiometer (MODIS) instrument on board the Terra and Aqua satellites with the high resolution North American Regional Reanalysis (NARR) temperature data. Available state-averaged planting date data were also used for evaluating the proposed methods over North Dakota. The crop type analyzed is corn, which has the second highest acreage of North Dakota crops and has expanded significantly in coverage over the past couple decades. Longer term variations in planting dates over North Dakota are also studied. This work serves as a first step towards improving crop modeling data integration through strategic use of remote sensing coupled with atmospheric modeling to reduce uncertainty in initial parameters.

Chapter 2

Datasets and Models

2.1 Moderate Resolution Imaging Spectroradiometer (MODIS) Level 1B calibrated reflectance

The Moderate Resolution Imaging Spectroradiometer (MODIS) instrument, which is on board both Terra (equatorial crossing time ~10:30 am) and Aqua (equatorial crossing time ~1:30 pm) satellites, provides daytime coverage every one to two days over the equator. Daily coverage is available north of 30º latitude, which includes this study's scope of North Dakota. A total of 18 years (2003-2020) of Terra and Aqua MODIS level 1B calibrated reflectance data (MOD02QKM and MYD02QKM) were used in this study (https://mcst.gsfc.nasa.gov/sites/default/files/file_attachments/M1054E_PUG_2017_0901_V6.2.2_Terra_V6.2.1_Aqua.pdf, last accessed on Sep. 29, 2021). The MODIS level 1B dataset includes calibrated radiance or reflectance values for 36 spectral channels at a spatial resolution of 250-1000 m at nadir. MODIS reflectance at red (620-670 nm, channel 1) and NIR (841-876 nm, channel 2) wavelengths, both at a spatial resolution of 250 m at nadir, are used for computing NDVI based on Eq. 1:

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad (1)$$

Note that NDVI values were derived from both Terra and Aqua MODIS in this study and different observation conditions, including different viewing geometries, are expected for Terra and Aqua satellites. Thus, the difference in NDVI values from Terra and Aqua MODIS were examined as shown in Fig. 1. To construct Fig. 1, spatially collocated Terra and Aqua NDVI values were used over corn field locations from April to June of 2012, the year chosen as the calibration year.

To spatially collocate Terra and Aqua MODIS data, a filter was applied to use coordinates centered in corn fields with 125m of continuous corn field located on each side. The feasibility of using available data from both Aqua and Terra to derive NDVI values through a statistical comparison was also explored. NDVI values from Terra and Aqua (Fig. 1) show a very high correlation with r^2 of 0.80. The mean difference (Terra-Aqua) and the RMSE in NDVI are 0.006 and 0.096, respectively. As suggested from Fig. 1, NDVI values over corn fields are comparable for observations from Terra and Aqua MODIS.

The close relationship between MODIS Terra and Aqua is expected as it was also seen in an Australian seasonal grassland study (Nguyen Tran et al., 2020). An NDVI overestimation of up to 5% or a NDVI value of about 0.03 has been noted due to viewing and illumination angle variations, particularly in the winter season when the solar zenith angle is higher. Differences in the late spring and early summer timeframe were negligible (Nguyen Tran et al., 2020). Note that this study used data collected primarily in June, when the solar zenith angle is at a minimum. Merging similar satellite datasets with different overpass times and distances from nadir maximizes available cloud-free data coverage. Thus, NDVI values from both Terra and Aqua are jointly used for this study. A majority of corn fields were found to have valid NDVI measurements from both Terra and Aqua. Some of the differences or errors between two satellite measurements for a site are accounted for by requiring multiple days of NDVI threshold exceedance (e.g. 0.275 as discussed later) when detecting initial corn emergence from the soil.

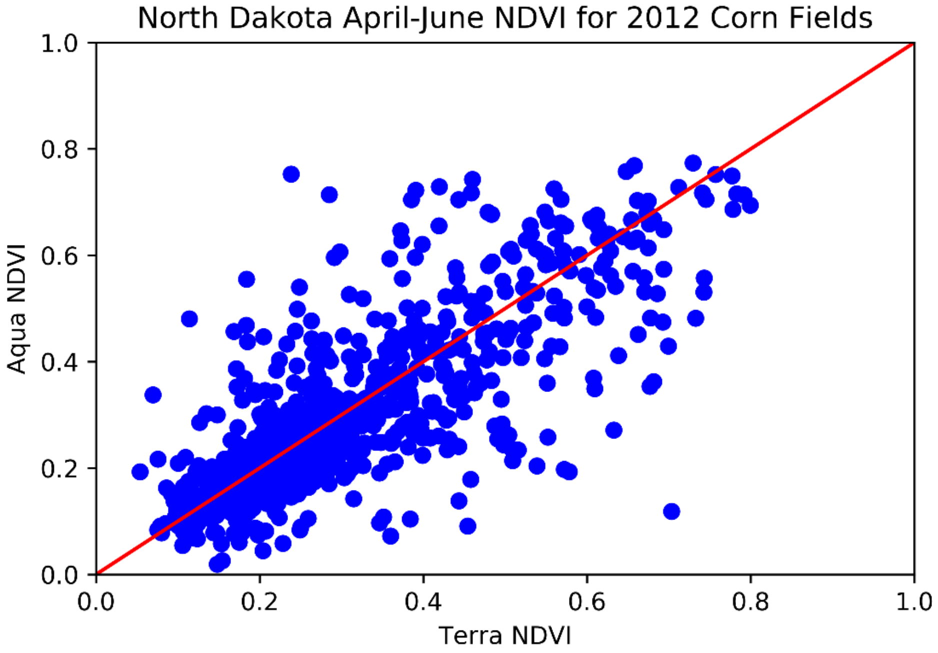


Figure 1. Comparison of NDVI values derived using MODIS Terra and Aqua data for collocated corn fields from April-June 2012

NDVI values can only be derived over cloud free regions and so readers shall be aware that there are uncertainties relating to MODIS-based NDVI values including cloud contamination as well as overlap between the boundaries of crop fields and satellite pixel resolution. For the purposes of this study only cloud-free satellite data were used.

2.2 The North American Regional Reanalysis (NARR) Data

The North American Regional Reanalysis (NARR) is a high-resolution combined model dataset across the continent generated through the use of National Centers for Environmental Prediction (NCEP) Eta model that includes the assimilation of extensive amount of observational data (Mesinger et al., 2006). Daily NARR data are recorded with eight temporal periods at approximately a 32 km resolution.

NARR 3-hourly temperature data is used for computing Growing Degree Days (GDD), that account for and accumulate during the time between planting dates and the first detectable dates for corn fields by NDVI. GDD instead of calendar days are used, as plant growth is strongly affected by environmental temperature (Gilmore and Rogers, 1958), and is computed based on daily minimum (T_{min}) and maximum (T_{max}) temperatures. For corn, the base temperature is considered to be 10 °C with an upper limit of 30 °C (Cross and Huber, 1972). This is considered the ideal temperature range for corn, since a sufficient amount of heating is necessary for growth, but exceedingly high temperatures which provide no additional benefit (Cross and Huber, 1972). If the daily minimum or maximum temperature is below 10°C, that daily minimum or maximum temperature is set to 10°C. Similarly, if the daily minimum or maximum temperature is above 30°C, that daily minimum or maximum temperature is set to an upper limit of 30°C. After these restraints that prevent or limit growth are accounted for, the GDD is computed by subtracting averaged daily temperature (T_{avg}) by the temperature floor (10°C), where T_{avg} is computed by averaging of daily maximum and minimum temperatures as mentioned above. Details of the steps are illustrated in Eqs. 2-4:

$$If\ \text{Tmax}\ or\ Tmin > 30°\text{C}, T = 30°\text{C}\ and\ if\ Tmax\ or\ Tmin < 10°\text{C}, T = 10°\text{C} \quad (2)$$

$$Daily\ Tavg\ (°\text{C}) = \frac{Tmax + Tmin}{2} \quad (3)$$

$$Daily\ Corn\ GDD\ (°\text{C}) = Tavg - 10°\text{C} \quad (4)$$

2.3 The Cropland Data Layer (CDL) data

For crop location, the Cropland Data Layer (CDL) is a product of the USDA National Agricultural Statistics Service (NASS) with a crop-specific land cover classification of more than 100 crop categories grown in the United States. These CDLs are derived from satellite observations through the use of a supervised land cover classification (Boryan et al., 2011). The classification relies on first manually identifying pixels within certain images that represent the same crop or land cover type. Using these pixel sites, a spectral signature is then developed for each crop type that is then used by analysis software to identify all other pixels in the satellite image representing the same crop (Johnson, D. and Mueller, R., 2010). In this study, CDL data were obtained from https://nassgeodata.gmu.edu/CropScape/ (last accessed on Sep. 29, 2021) with a 30 m spatial resolution for the study period of 2003-2020 over North Dakota. Homogeneous corn fields of at least 23 ha. are utilized in this study, based on contiguous areas that are 16 CDL pixels wide by 16 CDL pixels tall. 16 by 16 CDL pixels in size.

2.4 The NASS Crop Progress Reports

The NASS Crop Progress Reports from the U.S. Department of Agriculture (USDA) also give weekly values of the cumulative progress for each crop at key stages at a state-wide scale. Crop progress metrics include statewide areal percentages of corn that has been planted or is emerging from the soil. These values are obtained from a survey of a sample of farmers across a given state on a weekly basis. Thus, any particular percentiles of crops planted, such as the 50th (or median), generally must be interpolated between two weeks with the error margin being 1-2 days. The dataset for North Dakota corn was obtained as reference points to allow for calibration and a comparison for an estimated seasonal planting curve. The NASS crop progress reports can be manually retrieved for each week from the webpage at https://usda.library.cornell.edu/concern/publications/8336h188j?locale=en#release-items (last accessed Oct 26, 2021). A total of 110 weekly surveys were available for analysis over the 18-year period, an average of just over 6 weeks per year.

Chapter 3

Methodology

For this study, the CDL-determined corn locations are collocated with MODIS NDVI values to detect regions with healthy growing corn at the minimal level of NDVI that is distinctly distinguishable from the underlying soil for the April-June timeframe. This covers the period for potential planting and time for corn to emerge from the soil, which owing to cooler regional temperatures and soil drainage timing, occurs later in North Dakota than most other states. Combined with GDD calculated from the NARR, these datasets were merged in order to determine the amount of GDD accrued before a NDVI threshold date, which then estimates the date of planting for a particular location. Finally, the combined dataset is calibrated through use of the median planting dates from NASS crop progress reports for North Dakota for the selected calibration year. The expected GDD was then compared with other years to determine the accuracy of this method with selected NDVI thresholds. This process is illustrated visually in Fig. 2 with a chart showing overall data flow.

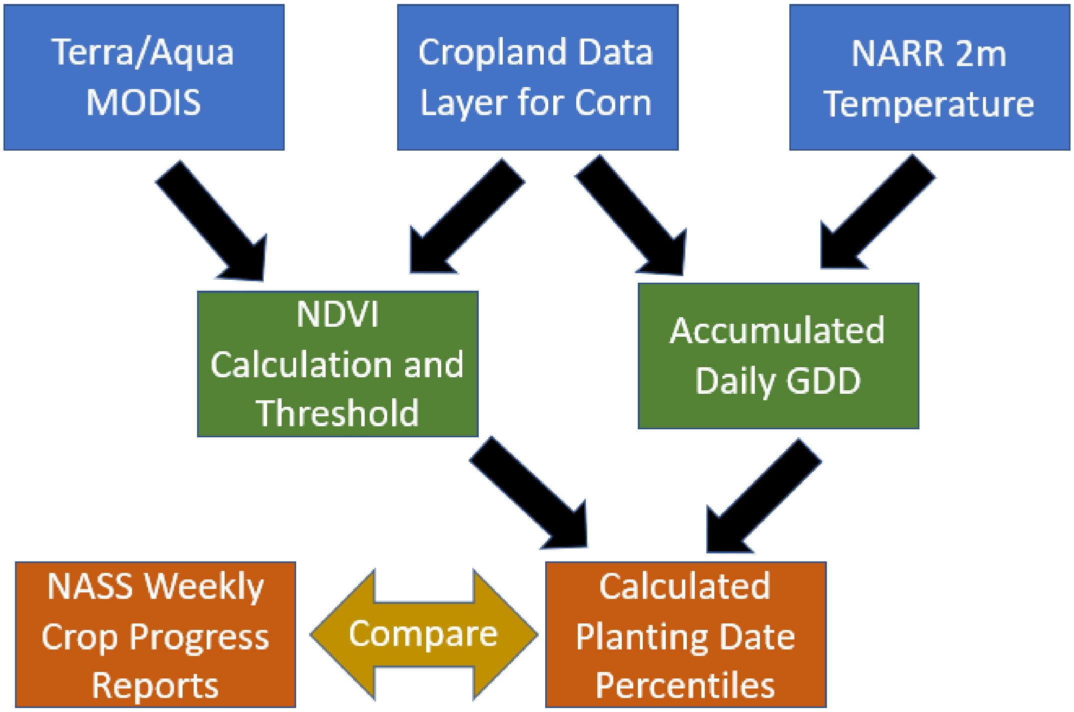


Figure 2. Flow chart for data sources and methodology to reach final results/analysis

3.1 Theoretical basis for detecting planting dates using NDVI

The theoretical basis for detecting planting dates using NDVI is illustrated in Fig. 3. Figure 3 shows the averaged NDVI computed as a function of calendar date, covering the entire growing season of April through October using 2015 Terra and Aqua MODIS data from North Dakota. Note that with the satellite pixel width of 250 m from MODIS at the red and NIR channels, a square with sides of 125 m from the center is needed to guarantee a corn field's placement within this pixel. In Fig. 3, the full sample of North Dakota NDVI measurements located inside a corn field with a buffer of 125 m shows a curve with a peak in July/August. The horizontal pink line represents the NDVI threshold of 0.275 used in this study (a sensitivity study of this threshold is discussed later) along with the vertical red lines indicating the start of each month, beginning with April. As indicated in Fig. 3 with 2015 as an example, NDVI values in April are steady and concentrated around 0.20 due to bare ground, before starting to rise by early-June due to plant germination. A peak NDVI value near 0.75 is seen in July, followed by a slight reduction in August attributed to the corn tasseling/pollination stage.

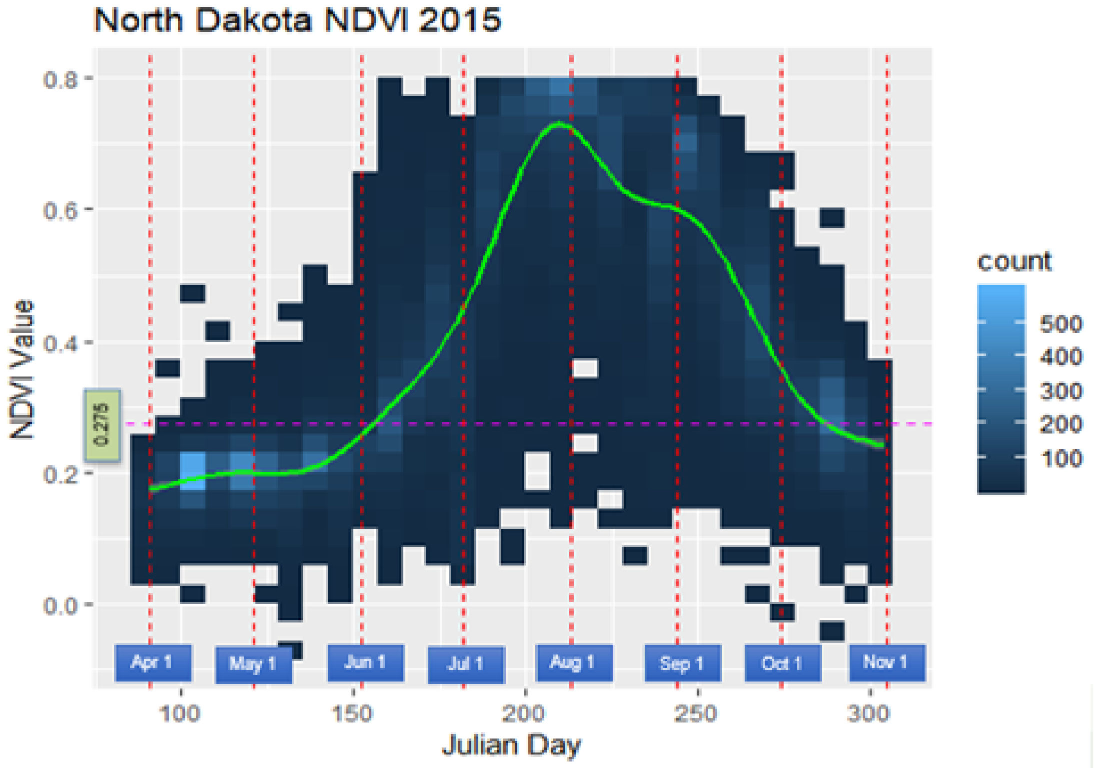


Figure 3. Example of North Dakota NDVI data distribution from 2015 using a 0.275 threshold

A statewide spatial distribution in Fig. 4 shows NDVI values collocated with corn fields for April in the calibration year of 2012. For plotting purposes, the latest available NDVI value for each point is shown. This would be near or just before the time of planting (median May 5th) after the snow cover diminished and so likely representative of bare soil NDVI. The mean NDVI shown for April 2012 was 0.18 with a 95th percentile of 0.25. Thus, after experimentation with various thresholds at or above 0.25, high confidence was given to using a 0.275 threshold that would be just above that of bare ground for almost all fields. The selected threshold (0.275) is a marker of an early stage of growth where the crop is first detectable ($DATE_{FD}$). As the life cycle of corn is often considered a function of accumulated GDD, the total GDD accumulation up to date of detectability (GDD_{FD_NDVI}) was used to indicate the time between planting date and first detectable date based on NDVI values, or $DATE_{FD}$. The planting dates can thus be derived by subtracting the number of calendar days that have corresponding GDD_{FD_NDVI} values (based on calibration) from $DATE_{FD}$.

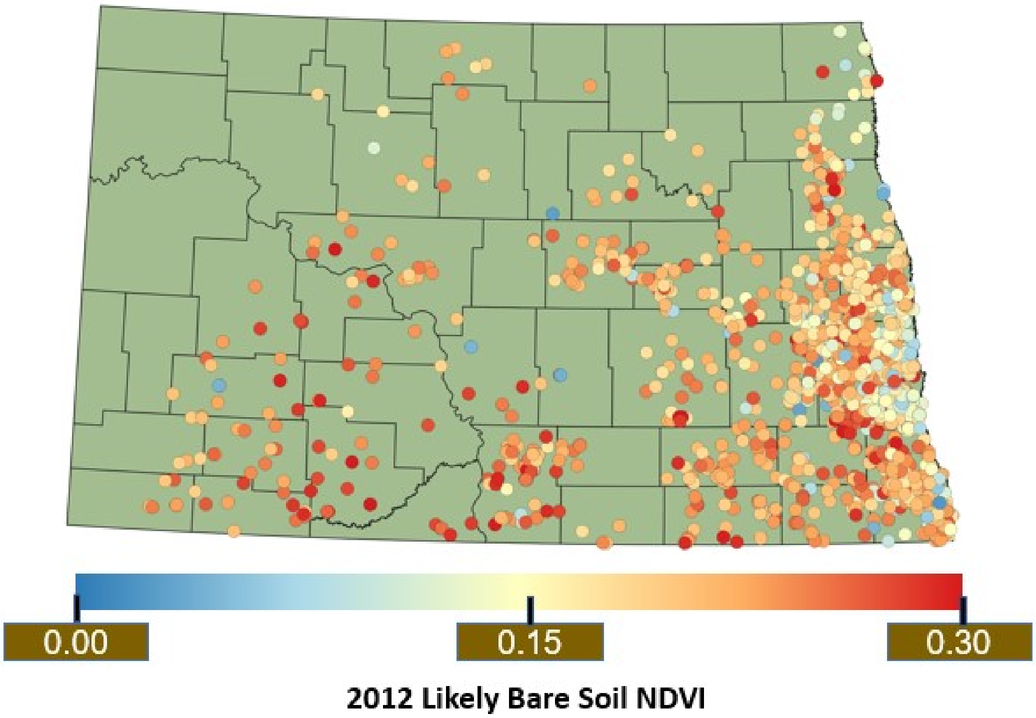


Figure 4. Latest available North Dakota NDVI values collocated with corn fields for April 2012, near or just before time of planting when soil is likely bare

3.2 Usage of NARR-based GDD in calculation of time span between planting and detectability

Corn will be indistinguishable from the underlying soil for a large period of time post-planting due to a number of factors. These include the length of time (approximately 2 weeks) between planting and first emergence, the small leaf area index of early-stage seedlings, and the large row spacing that is often up to 90 cm. To account for this period of undetectability an offset method must be selected, such as a fixed number of days, to accurately estimate the original planting date. For this study GDD was selected as the method of offsetting due to the previously documented close relationship between GDD and plant growth stages. Indeed, as documented in detail in section 4.1, much larger biases are found in estimated planting dates using calendar days instead of GDD. As surface weather stations are generally sparse and irregularly spaced across North Dakota, the NARR dataset was utilized to calculate GDD independently for each field location. Figure 5 shows 2012 as an example these daily values averaged over each month of the growing season of the calibration year of 2012 across North Dakota.

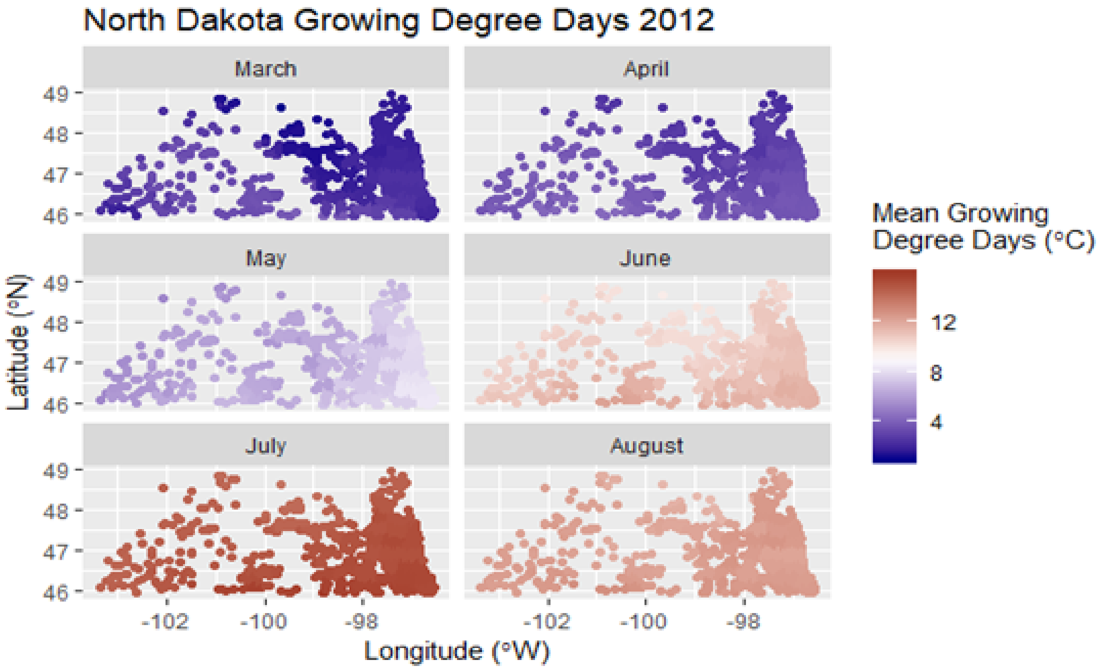


Figure 5. Growing Degree Days dataset for each corn field point with the average daily GDD given for each month. This is calculated from NARR data, with the highest monthly sum of GDD for 2012 being July for all locations.

3.3 Practical method for detecting planting dates through a combination of MODIS and NARR data

In practice, the Cropland Data Layer (CDL) data is used to determine which locations across North Dakota had entirely corn within a 250 m wide area for each year. The results of this CDL analysis are collocated with MODIS NDVI data. Once the NDVI and GDD datasets are created for each field, an earliest detection date which meets the set threshold for NDVI (0.275) is determined. A sensitivity study was performed to aid in selecting this NDVI threshold for the most accurate calibration, discussed further in section 4.2.

Next, these output calculations are merged in order to decipher locations with both a GDD and NDVI calculation available. The planting date for a given location is first estimated by detecting the date when the MODIS-based NDVI values met the predetermined NDVI threshold (or $\mathrm{DATE}_{\mathrm{FD}}$). This date indicates the earliest date that growing crop can be detected from satellite observations for the current growing season at a given location. Then, the planting date is estimated by offsetting the detected date by summing the prior days' GDD needed to reach a set value of accumulated GDD.

The requisite accumulated GDD threshold is estimated using NASS and MODIS data from 2012, which is selected as the calibration year for the study. For 2012, the median date of corn planted in North Dakota was May 5 according to NASS and the average date for $\mathrm{DATE}_{\mathrm{FD}}$ is June 12. That is, 38 days from planting to first detection. For this control year, the mean accumulated GDD between May 5 and June 12 is calculated to be 340 GDD (°C). Then, the combination of the threshold value and GDD is applied to

MODIS data from 2003-2020 to evaluate the proposed method, as well as study the longer-term variations in satellite detected planting dates.

Lastly, the spatial distribution of estimated planting dates for the control year are shown in Fig. 6. Calculated 2012 planting dates for North Dakota are overlayed, with corn field points used for calibration indicated. While the spring of 2012 was anomalously warm and dry, a couple of outliers with unrealistic dates several weeks earlier in March were set aside in order to display a range of closer to 30 days from mid-April to mid-May. Higher elevations away from the Red River in southeastern North Dakota had slightly earlier planting dates, potentially due to a less saturated water table.

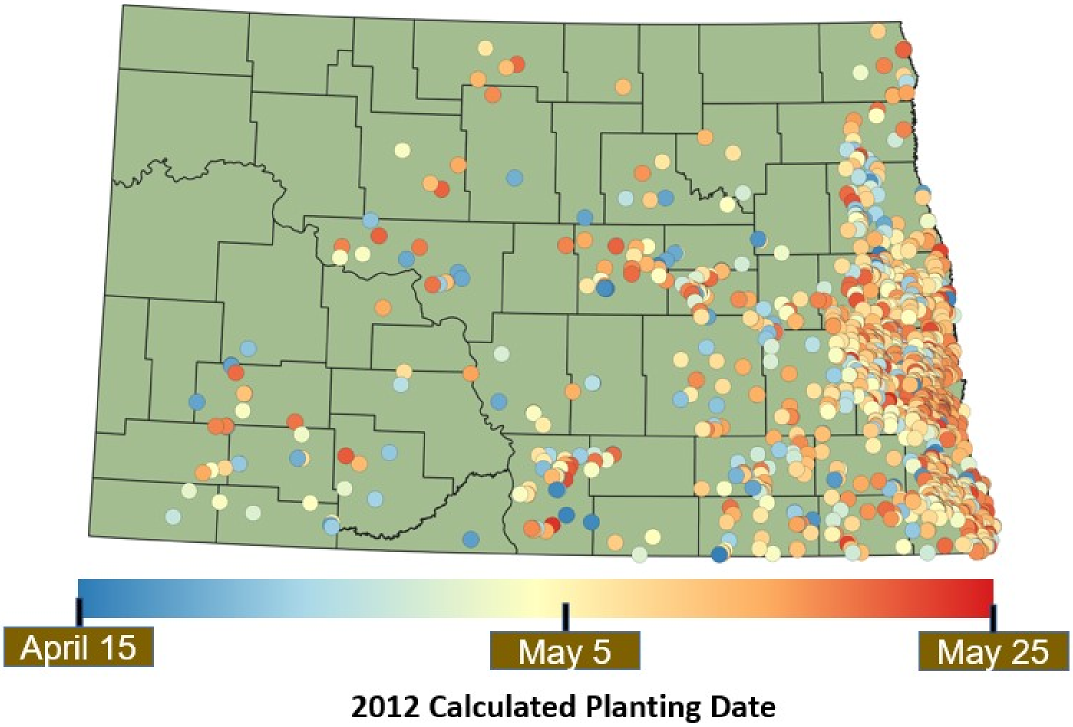


Figure 6. Planting dates for 2012 control year (from April 15-May 25), later dates shown in red

Chapter 4

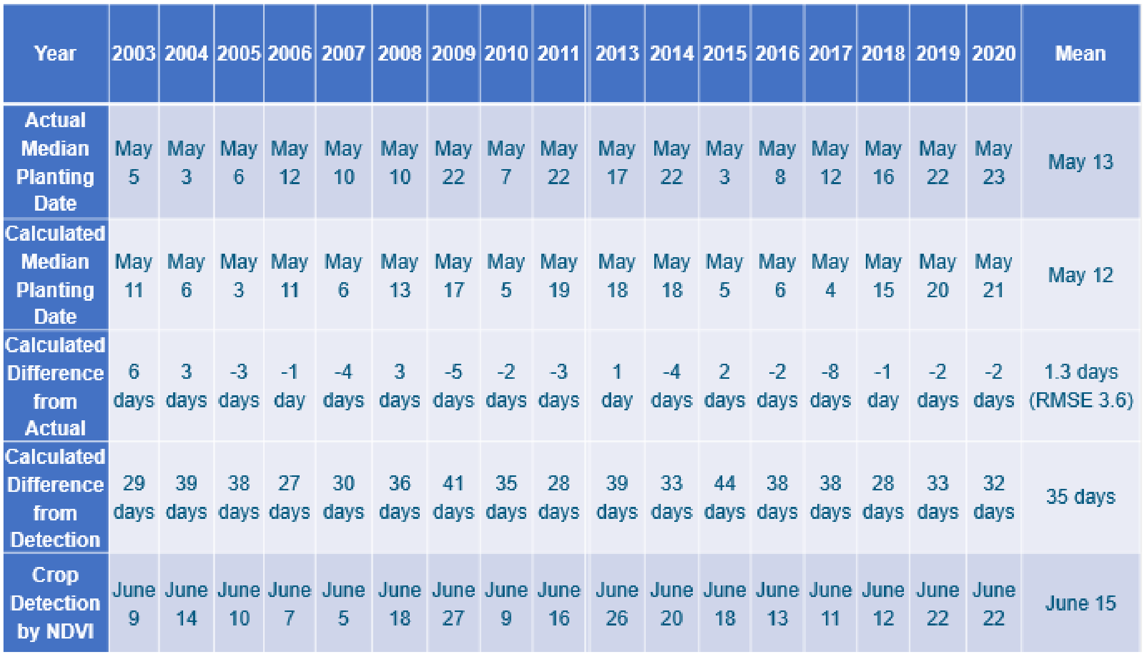
Results and Discussions

In this section, the method described in the previous section is implemented for Terra and Aqua MODIS data from 2003-2020. As a validation effort, the satellite-derived planting dates are inter-compared with statewide NASS crop progress reports. Longer term changes in planting dates for corn over North Dakota are also studied.

4.1 Evaluation of the derived planting dates from MODIS using NASS crop progress reports data

For this study, the percent of corn estimated to be planted by a particular date is compared against the NASS crop progress reports for the weekly percent planted. A critical threshold in the NASS crop progress reports is the median planting date, referring to having 50 percent of areal corn planted. Table 1 shows the comparison between the actual (NASS crop progress reports) and expected (MODIS derived) median planting dates across all of North Dakota. Note that for 2012 the actual median planting date is May 5th, which is used to derive GDD_{FD_NDVI} values for the proposed method, and thus, the MODIS derived planting date is also the same day as expected. For years other than 2012 (2003-2011 and 2013-2020), the mean difference between actual and calculated median dates is 1.3 days and the RMSE is 3.6 days. Note that if the initial calibration is based solely on calendar days (38), instead of GDD between 2012 median detection and planting dates, a mean difference of 4.7 days earlier median planting dates and a RMSE of 7.6 days are found. These values are significantly higher than the results using GDD.

Table 1. Target median planting dates and calculated dates for 2003-2011 and 2013-2020, along with median date of NDVI threshold (0.275) detection



Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2013	2014	2015	2016	2017	2018	2019	2020	Mean
Actual Median Planting Date	May 5	May 3	May 6	May 12	May 10	May 10	May 22	May 7	May 22	May 17	May 22	May 3	May 8	May 12	May 16	May 22	May 23	May 13
Calculated Median Planting Date	May 11	May 6	May 3	May 11	May 6	May 13	May 17	May 5	May 19	May 18	May 18	May 5	May 6	May 4	May 15	May 20	May 21	May 12
Calculated Difference from Actual	6 days	3 days	-3 days	-1 day	-4 days	3 days	-5 days	-2 days	-3 days	1 day	-4 days	2 days	-2 days	-8 days	-1 day	-2 days	-2 days	1.3 days (RMSE 3.6)
Calculated Difference from Detection	29 days	39 days	38 days	27 days	30 days	36 days	41 days	35 days	28 days	39 days	33 days	44 days	38 days	38 days	28 days	33 days	32 days	35 days
Crop Detection by NDVI	June 9	June 14	June 10	June 7	June 5	June 18	June 27	June 9	June 16	June 26	June 20	June 18	June 13	June 11	June 12	June 22	June 22	June 15

For most years, calculated median dates were within 1-4 days of the reported date. The calculations for 2003 and 2017 were the farthest from the survey dates, both having a discrepancy of about a week. Also listed in row 5 of Table 1 are the median dates for corn fields to be detected by MODIS based on the given NDVI threshold of 0.275, or $DATE_{FD}$. In general, $DATE_{FD}$ occurs just over a month after the calculated median planting dates, as indicated by row 4 with a mean difference of 35 days.

In addition, note that weekly, state-averaged data are available from the NASS crop progress reports, representing the percentage of corn fields planted for a given state. Thus, a planting date curve, with x-axis representing Julian day and y-axis representing percentage of corn planted, can be constructed from both NASS crop progress reports (orange dots) and MODIS derived (blue curve) planting dates as shown in Fig. 7 for years

2015 (Fig. 7a) and 2018 (Fig. 7b). The curve of NDVI detection dates is also plotted in green, allowing for a visual representation of the gap between planting in which GDD is accumulated. For example, as shown in Fig. 7b, with the actual median planting date for 2018 being May 16 and the calculated date May 15, both curves closely match near the median, with the MODIS derived planting date curve matching reasonably close to points from NASS statewide crop progress reports from each week. Since the 2012 calibration for GDD was only applied to the median planting date, the most notable differences between the actual and calculated curves are found near the lowest and highest percentiles.

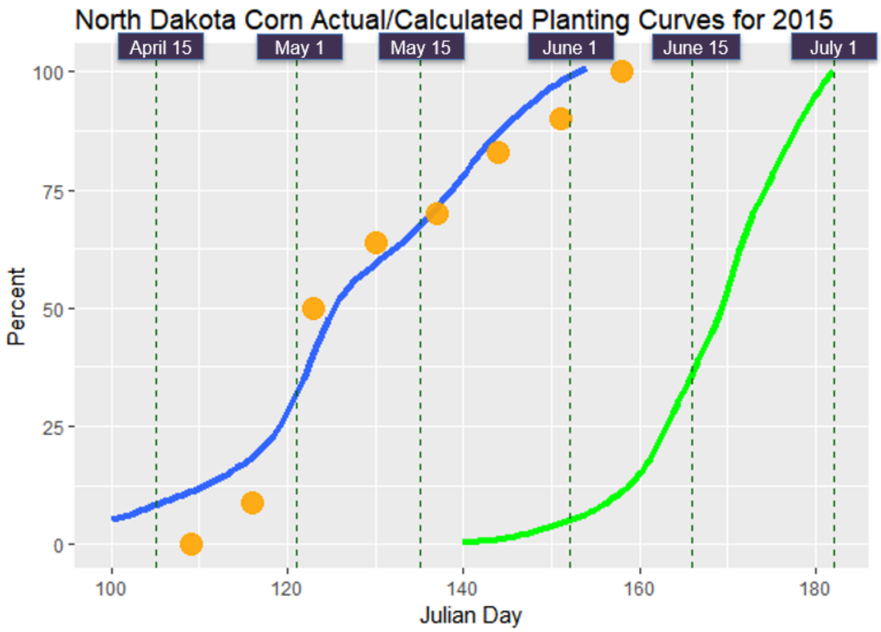


Figure 7a. Calculated planting curves (blue) for 2015 with NASS survey points (orange) and NDVI detection curves (green)

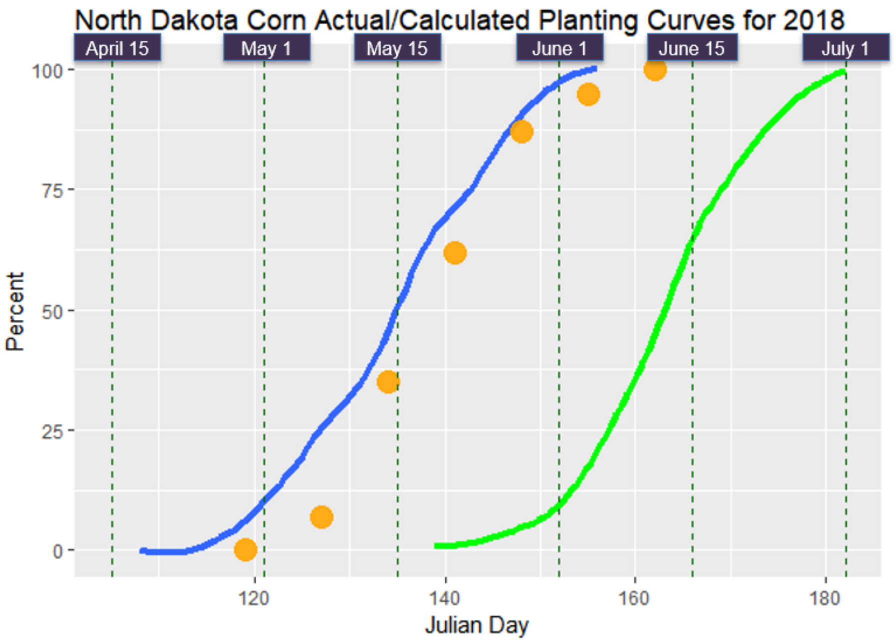


Figure 7b. Calculated planting curves (blue) for 2018 with NASS survey points (orange) and NDVI detection curves (green)

For 2018, the integrated area of the curve in Fig. 7b is almost entirely within the month of May, with a range from the last few days of April to the beginning of June. About half of corn planted was within the 2nd and 3rd quartiles of 25-75 percent, occurring in slightly less than a two-week period, of which the calculated curve closely matches. Vertical lines correspond to the indicated calendar dates of both plots, which are with respect to Julian days. Other years analyzed generally followed a similar pattern of accuracy with the middle 50 percent of planting occurring within 10-15 days.

Figure 7 was extended to compare actual (from NASS crop progress reports) and estimated (from MODIS) percentage of corn fields planted on a weekly basis as shown in Fig. 8 using temporally collocated NASS crop progress reports and MODIS data from

2003-2020. Each data point in Fig. 8 represents a pair of weekly statewide percentages of corn planted from both the NASS crop progress reports and MODIS derived method. The r^2 value is 0.96, mean difference (Calculated-Survey) is 6.66% area planted and RMSE 11.9% area planted. As expected with calibration based on the median, actual and calculated dates match most closely around 50%. They also maintain a particularly strong correlation up until nearly 100% planted. However, survey planting percentiles less than 40% were consistently exceeded by higher calculated percentages. This indicates that most calculated planting curves do not slope down towards 0 percent, or time when first planting begins, fast enough or as accurately as it does when approaching 100% planted. Much of the error can be accounted for within each of the first available NASS survey weeks, which range from mid-April to early-May.

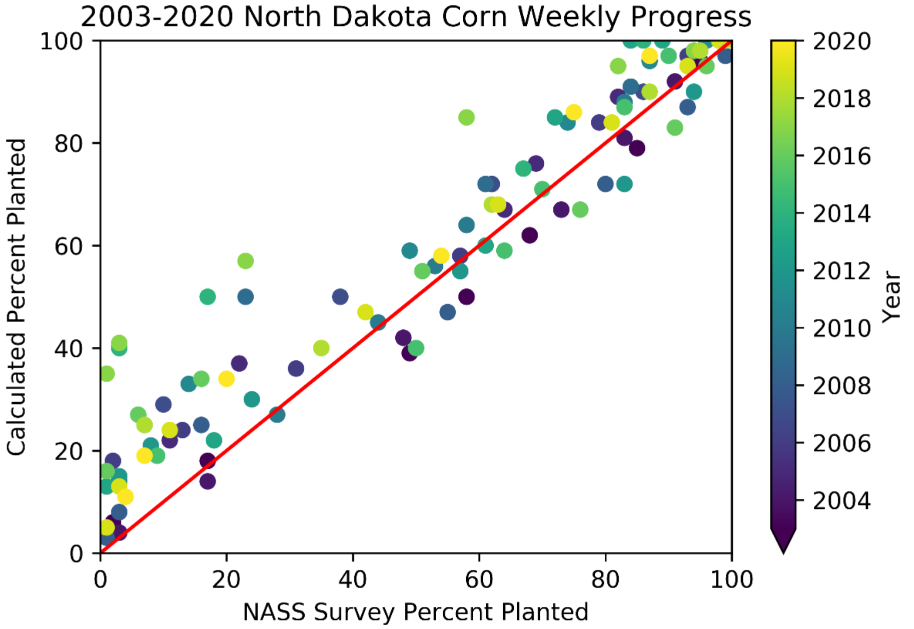


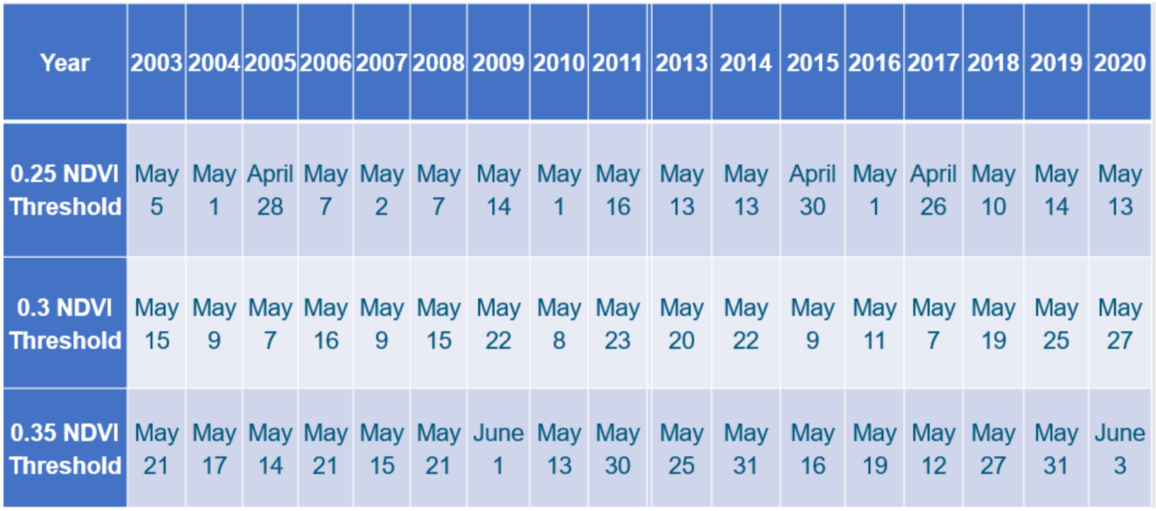
Figure 8. Comparison of weekly NASS survey percent planted versus calculated percentile, 2003-2020

A 2012 calibration based on the 25th planting percentile was performed for 2013-2018 that resulted in a mean difference between calculated and actual dates (for having 25 percent planted) of 4 days earlier, compared to 11 days when calibrating from the median. A total of 385°C in GDD before detection was necessary, which is 45°C GDD more than the median-based calibration. Thus, inconsistencies and wider variations in a relatively small amount of GDD during in the early spring can explain at least part of the earlier bias in calculations for lower percentiles. It should be noted that since North Dakota is one of the last states to finish planting corn, NASS reports for most years end when around 90-95% of corn has been planted.

4.2 Sensitivity Study

The NDVI threshold of 0.275 is selected for this study through a calibration process using MODIS and NASS crop progress reports data from 2012. Then, the sensitivity of this NDVI threshold was tested for all other years in the study. Table 2 shows examples of calculated median planting dates from MODIS as a function of the NDVI threshold. Initially, a 0.35 NDVI threshold was chosen; however, when holding the number of GDD constant for the 2012 calibration, corn detection was often 1-2 weeks too late, with a mean difference between the actual and calculated planting dates to be -9.4 days. By reducing to a value of 0.3, the mean difference reduced to -2.6 days but still a few days later on average. A 0.25 threshold proved to be just as valid, but typically several days too early (mean difference of 6.4 days). Lastly, the final selected value of 0.275 lined up well as a calibration and performed best in accuracy when comparing to survey dates from the NASS crop progress report (mean difference of 1.3 days). This exercise suggests that the NDVI threshold of 0.275 as derived using data from 2012 is also applicable for other years, which confirms the robustness of the method proposed in the study. Also note that seasonal plots of NDVI also aided in justifying the threshold choice being marginally high enough to eliminate most contamination from bare ground values, such as the example shown in Fig. 3.

Table 2. Calculated median planting dates using NDVI detection thresholds of 0.25-0.35



Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2013	2014	2015	2016	2017	2018	2019	2020
0.25 NDVI Threshold	May 5	May 1	April 28	May 7	May 2	May 7	May 14	May 1	May 16	May 13	May 13	April 30	May 1	April 26	May 10	May 14	May 13
0.3 NDVI Threshold	May 15	May 9	May 7	May 16	May 9	May 15	May 22	May 8	May 23	May 20	May 22	May 9	May 11	May 7	May 19	May 25	May 27
0.35 NDVI Threshold	May 21	May 17	May 14	May 21	May 15	May 21	June 1	May 13	May 30	May 25	May 31	May 16	May 19	May 12	May 27	May 31	June 3

4.3 Long Term Planting Date Trends

It is also interesting to study the longer-term change in planting dates for corn for the study period (2003-2020) using both the NASS survey and MODIS data. Fig. 9 shows the variation of the median corn planting date from the NASS survey for North Dakota over the past 20 years. The 2010s saw more variability and overall slightly later planting dates, but there is not a significant linear correlation ($r^2 = 0.16$). Also notable is that there is some positive relationship between actual median planting dates and our NDVI detection dates from 2003-2020 ($r^2 = 0.39$). Corn production has expanded significantly in area during this time and the distribution across the state shifted more north/west. This major expansion to northern and western parts of the state that have typically have relatively later last freeze dates may skew the results of an overall statewide trend in planting dates.

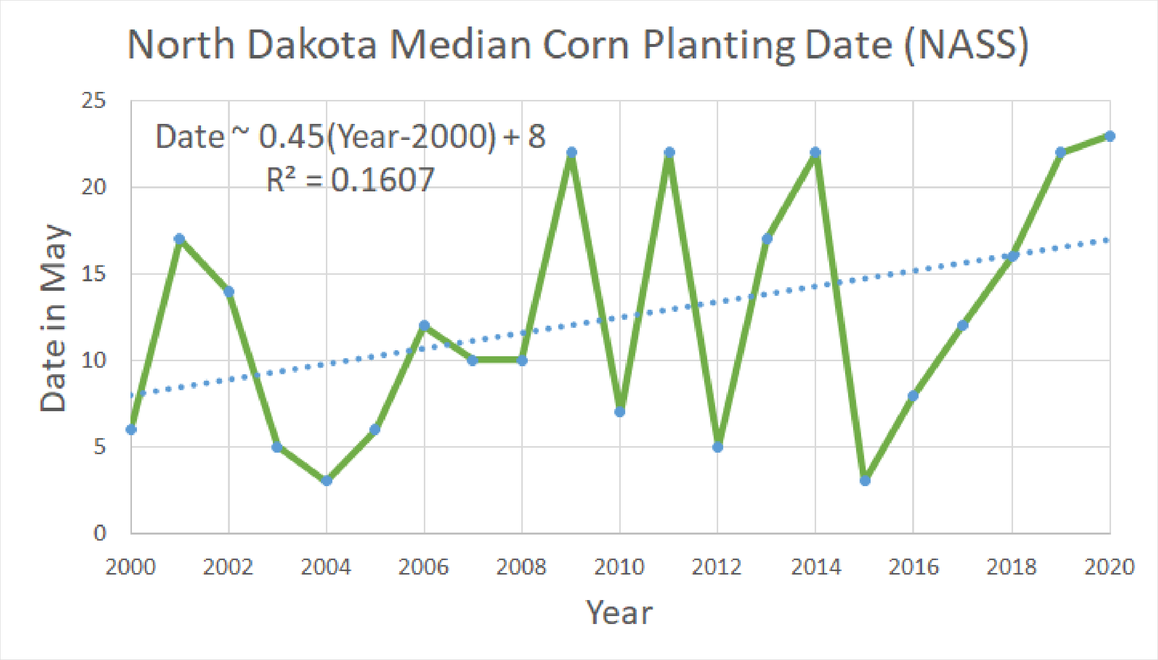


Figure 9. 2000-2020 planting date trend/correlation, based on median dates from NASS survey

To address this change, corn field locations that have consistently been in use from 2003 through 2020 were isolated to calculate trends in planting dates using the developed method. Fig. 10 shows the spatial distribution of trends in planting dates for corn fields that have had corn consistently in the crop rotation since 2003. On average those fields trended about 2 days later over the 18-year period, but with some regions seeing corn planting up to 7 days earlier or later. There has been zero trend in annual temperature at Fargo Hector Int'l Airport from 2003-2020, which is nearby many of the points shown in Fig. 10. A statistically insignificant decrease in April temperatures was observed over this period ($r^2 = 0.18$) along with an even more marginal increase in May temperature ($r^2 = 0.04$). The last spring freeze is trending slightly later ($r^2 = 0.13$) and there is zero correlation for last measurable snowfall over this timeframe. So, the minimal but overall slightly later change in planting dates calculated over these sites is plausible. Also, note that the locations shown mostly reflect the 2003 planting distribution, which is significantly less in coverage and much more concentrated in the southeast portion of the state. A majority of fields along the Red River trended several days later, such as around Fargo in eastern Cass County. A relative increase in the frequency of floods, including the Sheyenne River in this area, may explain the need to plant later that is independent of GDD and spring temperature trends.

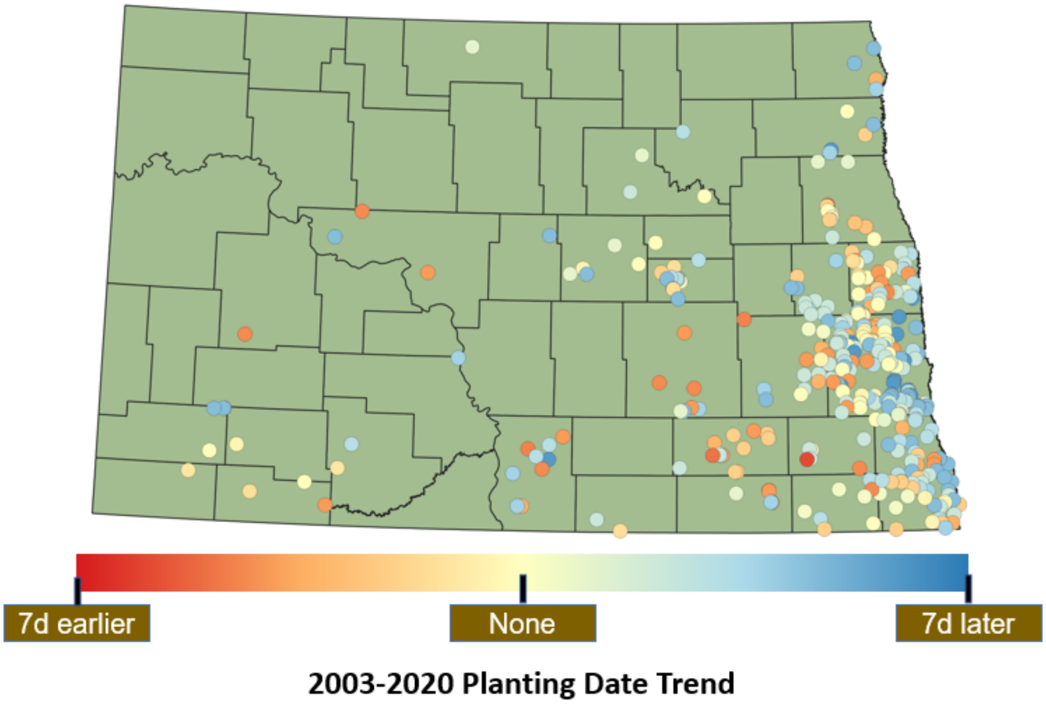


Figure 10. Median planting date trend for consistent corn field locations from 2003-2020

4.4 Limitations

While the results show a degree of consistency between derived and measured planting date, there are several uncertainties and improvements that could be made to the methodology. Estimated planting curves for 2003-2020 are solely based on a calibration of the actual median planting date for 2012. Thus, when some of these weekly points of percentage planted in Fig. 8 are plotted together, a clear bias is shown for most years with the curve skewed toward earlier dates. One potential reason for this is a contamination of GDD from early-season warm periods, particularly in 2012, that fall well before the threat of frost diminishes in the spring. Reported NASS survey data and the climate of North

Dakota generally do not suggest that any significant corn can be planted before April 15 and crop insurance plans do not include freeze protections until the end of April.

However, errors can also be attributed to reported NASS survey data, as it has a limited sample size and is only reported weekly. Thus, imprecise values for the percentage of corn planted used for calibration result from having to interpolate between these weekly data. Since the most significant weekly increases in statewide percent planted are often observed near the median planting dates, the most uncertainty in interpolation lies then. The estimated planting curve can then be considered within a standard margin of error from actual dates, or potentially forming a curve more accurately than that made by the survey. Finally, with the scope of this study limited to only North Dakota and a 2012 calibration, there is limited confidence that this method functions universally. Extending this method to before 2003 would be limited by the lack of availability of enhanced satellite data resolution, compromising the accuracy; this is further compounded by the complication of a less complete cropland data layer. Overall, the concept of a comprehensive dataset of planting dates assists all other considerations as a key starting point in model simulations.

4.5 Alternate Interpretations

A few major components of crop modeling are soil temperature and moisture/precipitation. These certainly have an influence on overall crop growth and planting dates. Therefore, using temperatures from NARR for GDD may be considered as only one perspective of remote sensing. While tools such as radar and satellite can be used to determine these other parameters, models typically account for factors such as precipitation and solar radiation well. The GDD calibration could be expanded to include other meteorological factors, though the most influential change might result from using additional percentiles for calibration points besides solely the median. Finally, the NDVI threshold used is critical and can be arbitrary, but overall bare ground values for corn fields only vary slightly across all years between snow cover and planting, assuming consistent tillage practices. The threshold must remain just high enough to detect plant growth, but not to the extent that a high number of GDD needs to be counted back, increasing the margin of error.

Chapter 5

Conclusions

A method is developed to derive crop planting dates by combining satellite-based NDVI values with atmospheric model derived GDD. In this proposed method, the crop planting date is estimated through finding the first date that a crop field can be identified from the satellite based NDVI value after planting. Then, the time (estimated based on GDD) is accounted for from the planting date backward to the first day when crops can be detected using an NDVI threshold. This concept is tested over corn fields in North Dakota using 18 years of Terra and Aqua MODIS data. It was found that:

1. Compared against NASS crop progress reports data, which are weekly farmer surveys available at the state level and are used as ground-truth in this study, the difference in observed (from NASS data) and calculated planting dates is 1.3 days, with a RMSE of 3.6 days. This is in contrast to using a set number of calendar days to calibrate, which resulted in a mean difference of 4.7 days and RMSE of 7.6 days.

2. In addition, week-to-week changes in percent planted that are detected by the estimated curves for this method match closely to reported points near the median calibration point from the NASS data. This suggests that developed method can be used for estimating crop planting dates with a high spatial resolution and a large spatial coverage through the use of satellite based NDVI data. Clearly, the ability to determine planting dates of individual fields over time as the method demonstrated in this study, is more complete than currently available statewide weekly averages.

3. Using 18 years of Terra and Aqua MODIS data, longer term changes in planting dates were examined over North Dakota. While the mean change is 2 days later in planting date, some regions over the southern part of North Dakota are found to have planting dates up to 7 days earlier after 18 years. This may or may not link to climate variations that the North Dakota region is currently experiencing.

Knowledge of planting times at a high spatial resolution with a large spatial coverage is needed, even with respect to weather forecasts days or weeks ahead of time. The combination of using past crop detection dates and temperature data to determine planting dates enhances the accuracy and capability of crop models to represent when particular vegetation is present for each field. This remote sensing approach has the potential to reduce the uncertainty in the initial parameters needed for crop modeling. Simulations over large regions with regional considerations may also be possible; however, improvements and clarity would result from a more complete dataset and calibration. For instance, the skew toward earlier dates in calculated planting curves should be adjusted to within a historical planting window based on other data sources, such as insurance dates.

The focus of any future work on this approach would be on refining the calibration to include multiple percentiles of reported planting dates, along with using other or multiple years as an alternate basis of calibration. Additionally, future years beyond 2020 can be tested to expand the sample size. Any findings related to corn in North Dakota could be expanded to other major crops in terms of acreage for the state.

Lastly, various regions and climate zones may also be of interest to explore, in order to test the validity of the remotely sensed temperature and GDD methodology.

References

Boryan, C., Z. Yang, R. Mueller, and M. Craig, 2011: Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.*, **26**, 341-358, https://doi.org/10.1080/10106049.2011.562309.

Cross, H.Z., and M.S. Zuber, 1972: Prediction of Flowering Dates in Maize Based on Different Methods of Estimating Thermal Units. *Agron. J.,* **64**, 351-355, https://doi.org/10.2134/agronj1972.00021962006400030029x.

Fontana, F., N. Coops, K. Khlopenkov, A. Trishchenko, M. Riffler, and M. Wulder, 2012: Generation of a novel 1km NDVI data set over Canada the northern United States and Greenland based on historical AVHRR data. *Remote Sens. Environ.,* **121**, 171-185, https://doi.org/10.1016/j.rse.2012.01.007.

Funk, C., and Coauthors, 2019: Recognizing the Famine Early Warning Systems Network: Over 30 Years of Drought Early Warning Science Advances and Partnerships Promoting Global Food Security. *Bull. Amer. Meteor. Soc.,* **100**, 1011–1027, https://doi.org/10.1175/BAMS-D-17-0233.1.

Gallo, K., P. Schumacher, J. Boustead, and A. Ferguson, 2019: Validation of Satellite Observations of Storm Damage to Cropland with Digital Photographs. *Wea. Forecasting,* **34**, 435-446, https://doi.org/10.1175/WAF-D-18-0059.1.

Gilmore, E.C., and J.S. Rogers, 1958: Heat Units as a Method of Measuring Maturity in Corn. *Agron. J.*, **50**, 611–615, https://doi.org/10.2134/agronj1958.00021962005000100014x.

Johnson, D.M., and R. Mueller, 2010: The 2009 Cropland Data Layer. *Photogramm. Eng. Remote Sens.*, **76**, 1201-1205.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Moulin, S., L. Kergoat, N. Viovy, and G. Dedieu, 1997: Global-Scale Assessment of Vegetation Phenology Using NOAA/AVHRR Satellite Measurements. *J. Clim.*, **10**, 1154-1170, https://doi.org/10.1175/1520-0442(1997)010<1154:GSAOVP>2.0.CO;2.

Nguyen Tran, N., and Coauthors, 2020: Seasonal Comparisons of Himawari-8 AHI and MODIS Vegetation Indices over Latitudinal Australian Grassland Sites. *Remote Sens.*, **12**, 2494, https://doi.org/10.3390/rs12152494.

Peters, A.J., E.A. Walter-Shea, L. Ji, A.Vina, M. Hayes, and M.D. Svoboda, 2002: Drought monitoring with NDVI-based standardized vegetation index. *Photogramm. Eng. Remote Sens.,* **68**, 71-75.

Peterson, T.C., and Coauthors, 1998: Homogeneity adjustments of *in situ* atmospheric climate data: a review. *Int. J. Climatol.,* **18**, 1493–1517, https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.

Sinclair, T.R., and N.G. Seligman, 1996: Crop Modeling: From Infancy to Maturity. *Agron. J.,* **88**, 698-704, https://doi.org/10.2134/agronj1996.00021962008800050004x.

Starr, J., H. Zheng, J. Zhang, J.S. Reid, 2019: Evaluating Sensitivities of Economic Factors through Coupled Economics-Almanac Model System. *Agron. J.,* **111**, 1865–1878, https://doi.org/10.2134/agronj2018.08.0537.

Tucker, C.J., 1979: Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.*, **8**, 127-150, https://doi.org/10.1016/0034-4257(79)90013-0.