Satellite Based Remote Sensing for Estimating Crop Yield: Examining the Use of Various Functional Forms and Vegetation Indices

by

Shania Miralda

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department of Agribusiness and Agricultural Economics

University of Manitoba

Winnipeg, Manitoba

Abstract

This research includes two studies on estimating crop yield using satellite-based vegetation indices. For the first study, the objective is to compare eight different functional forms for estimating U.S crop yield using satellite based NDVI. For the second study, the objective is to examine 10 satellite-based vegetation indices for estimating U.S crop yield.

For both studies, corn, soybeans, spring wheat, and winter wheat data for crop yield (bushel/acre) were obtained from the USDA NASS from 2008 to 2019 for a total of 12 years covering all 48 states in the United States except Hawaii and Alaska (though different states are included, based on where the crops are grown). Data for the vegetation indices were obtained from the MODIS satellite using 250m resolution level and selecting for maximum Vegetation Index values. The methodology used regression with crop yield as the dependent variable. The main independent variable is the selected vegetation index (e.g NDVI, GOSAVI, etc). A time trend variable is also included, and dummy variables for U.S States.

Results for the first study indicated that relationship between NDVI and crop yield was mostly nonlinear, and piecewise regression was generally found to be the most suitable functional form. Results for the second study showed that for all 10 indices analyzed, that RDVI, GOSAVI, and GSAVI provided better estimates of crop yield than the commonly used NDVI. These results should be useful in providing a better understanding of various functional forms and various satellite based vegetation indices for improving crop yield estimation.

Keywords: Satellite Data, Remote Sensing, Crop Yield Estimation, NDVI (Normalized Difference Vegetation Index), RDVI (Renormalized Difference Vegetation Index), GOSAVI (Green Optimized Soil Adjusted Vegetation Index), GSAVI, Polynomial Regression, Piecewise (Segmented) Regression, GAM

Acknowledgments

I would like to express my gratitude to my advisor Dr. Milton Boyd. Thank you for your dedication, mentorship, and patience, and I am forever grateful for your guidance not only throughout my Master's studies but also about life skills. I would also like to thank my committee members, Dr. Barry Coyle and Dr. Lysa Porth, for their excellent and kind instruction and guidance. Additionally, I extend my gratitude to Dr. Mitchell Roznik who helped me in learning about remote sensing. Finally, I would like to thank my family for their encouragement and for believing in me throughout my studies at the University of Manitoba.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

THESIS INTRODUCTION

Estimating crop yield with satellite based remote sensing has a number of advantages over traditional on ground crop survey methods. Using remote sensing helps to lower costs by avoiding intense labor and time needed for crop surveys. As technology has improved, the cost of satellite data and computing has decreased, making satellite-based crop estimation more appealing.

Satellite-based vegetation indices make use of spectral bands, and usually the visible spectrum bands and the near-infrared are of interest for satellite-based crop yield estimation. The most used vegetation index is the Normalized Difference Vegetation Index (NDVI) due to its simplicity and ability to measure vegetation cover (greenness) quite accurately. However, it may be possible to use improved vegetation indices rather than NDVI, and it may also be possible to find improved functional forms over the linear regression form, for estimating crop yield.

Therefore, the first study of this thesis (Chapter 2) compares eight regression functional forms for estimating crop yield using the commonly used satellite-based vegetation index, NDVI. The eight regression functional forms include four groups: linear, polynomial (quadratic and cubic), piecewise/segmented (linear piecewise, quadratic piecewise, cubic spline, and natural spline), and Generalized Additive Model (GAM). NDVI is the main independent variable, other independent variables including the time trend and dummy variables for U.S. States. Crop yield is the dependent variable. Data is from 2008 to 2019 for the U.S., covering crop yield for corn, soybeans, spring wheat, and winter wheat. MODIS satellite data is used at the 250m resolution level. The relationship between NDVI and crop yield can be non-linear due to the limitations of NDVI, such as saturation. This study should be useful for those who seek to understand how using improved regression functional forms may help to more accurately estimate crop yield satellite-based vegetation indices.

The second study (Chapter 3) compares 10 vegetation indices for estimating crop yield, using two functional forms. The selected vegetation index is be the main independent variable, other independent variables included time trend and dummy variables for U.S states, and crop yield is the dependent variable. This study should be useful to understand how improved satellite-based vegetation indices, rather than traditional NDVI, may help to more accurately estimate crop yield.

CHAPTER 2

COMPARING VARIOUS REGRESSION FUNCTIONAL FORMS FOR ESTIMATING CROP YIELD USING SATELLITE-BASED NDVI

2.1 INTRODUCTION

The objective of this study is to compare various regression functional forms for estimating crop yield using satellite based NDVI. For this study, crop yield serves as the dependent variable and NDVI as a main independent variable.

Traditionally, crop yield estimation has been done mainly through surveys. This approach is not only labor intensive but also time consuming. This can lead to slower generation of yield estimation because the data is not quickly available. However, as technology advanced, government and the agricultural industry has taken advantage of satellites to monitor large areas for mapping or observing crop and vegetation growth without having to physically be at the location. There are a number of advantages of using remote sensing such as lower cost, faster in obtaining data and updating data fairly frequently. Also, using data from remote sensing for vegetation and crop estimation has been relatively accurate.

One type of data generated by remote sensing satellites is the different spectral bands that can be used to identify and classify certain landscape and terrain. Vegetation indices make use of these spectral band properties and often the visible spectrum and the near-infrared are of interest for such purposes. The most commonly used vegetation index is the Normalized Difference Vegetation Index (NDVI) due to its relatively simplicity that is able to measure vegetation cover quite accurately. NDVI, also called the measure of 'greenness' due to its correlation with photosynthetic capacity, has values ranging from -1.00 to +1.00 with higher values implying denser vegetation in the area (Pettorelli, 2013). Numerous literature has noted the usefulness of NDVI for agriculture and ecology sector (Vescovo et al, 2012; Purevdorj et al, 1998; Baugh and Groeneveld, 2006; Carreiras et al, 2006, Roznik, 2021).

But one limitation of NDVI is saturation at higher biomass density. This mean that as vegetation biomass gets higher, NDVI index value will not increase as much as expected. In other words, saturation is a situation where NDVI is not as sensitive to higher biomass (crop yield) and does not increase as much as it should. Knowing this situation, relationship between NDVI and crop yield may not be as linear and therefore using only linear regression functional form will lead to inaccurate estimation. Linear regression functional form is often used perhaps due to its simplicity in both form and interpretation (Carreiras et al, 2006), but in real applications, such form may oversimplify the relationship between crop yield and NDVI.

A non-linear functional form may be needed instead to address this saturation issue and provide more accuracy. A polynomial regression functional form is one approach to estimate crop yield and NDVI relationships, rather than linear regression form (Abdolalizadeh et al, 2020). Another nonlinear functional form, Generalized Additive Model (GAM) may be quite useful, as it is relatively flexible (Melin et al, 2017). However, one disadvantage is that a GAM approach may overfit the data given that it is a relatively flexible functional form.

2.1.1 Objective and Importance

Therefore, the focus of this study is to compare various regression functional forms for estimating crop yield using satellite based NDVI. The regression functional forms include linear regression form, polynomial regression form, piecewise (segmented) regression form, and Generalized Additive Model (GAM). NDVI will be one of the main independent variables and crop yield will be the dependent variable. This study can be useful for those who seek to understand how different nonlinear regression functions may help to better estimate yield of crops and identify which forms may be better for each crop. Using more suitable regression functional forms to estimate crop yield should improve accuracy of crop yield estimations when using NDVI.

The next part of this paper will discuss the literature about the various regression functional forms and how they may provide better estimation than linear regression form. The next section covers data, followed by methodology. Then results are presented, followed by a summary of the paper

2.2 LITERATURE REVIEW

2.2.1 Linear Regression Forms

The linear regression form is the most used form due to its simplicity, though sometimes it may not be the most accurate. Many studies have used the linear functional form (Shi and Xingguo, 2011), (Mkhabela et al, 2005), (Rasmusen, 1992), (Benedetti and Rossini, 1993), (Quarmby et al., 1993), (Groten, 1993), and (Manjunath et al, 2002) often for reasons of simplicity, and a number of researchers felt that the linear form was sufficient (Zaigham Abbas Naqvi et al, 2021), (Bolton and Friedl, 2013), and (Stepanov et al, 2020).

Despite its popularity, the linear regression form is not always the best form for estimation. This is due to saturation problem mentioned earlier involving nonlinearity when crop yield gets higher. More advanced functional forms may be needed to estimate the relationship of variables. (Kayad et al, 2019) used machine learning to estimate corn yield in Italy and found that machine learning, Random Forest technique, performed better and more robust than the simple regression forms.

2.2.2 Polynomial Regression Forms

Polynomial regression forms allow for more flexibility in estimation while still being easily interpretable. (Abrantes et al, 2021) used polynomial regression forms to help explain relationship of vegetation indices with soybean yield. Similarly, (Dadhwal and Sridhar, 1997) compared linear and non-linear forms to try and estimate wheat grain yield in India and found higher R^2 value (better fit) for the polynomial form compared to linear regression. Although (Jamali et al, 2014) did not focus specifically on agricultural crops, they indicated that vegetation amount/growth may progress non-linearly over time due to different external and natural factors such as insect infestation and change in moisture content, respectively. They fitted linear and polynomial regressions, including quadratic and cubic forms.

In general, the polynomial regression is straight forward to use but may still underestimate how data is behaving because such a form only captures the general trend of data. (Cummings et al, 2021) compared different regression forms including polynomial with machine learning for determining the nitrogen status in corn as it influences the growth and yield. They found that linear and polynomial regression forms are easy to use but machine learning that may be a better functional form approach.

2.2.3 Piecewise/Segmented Regression Forms

Another interesting functional form with flexibility but does not lose its simplicity is piecewise/segmented regression. Piecewise regression is similar to linear and polynomial regression, but the independent variable divides data into segments (e.g two different regression line slopes) through the use of different intervals/breakpoints. The breakpoints/knots can be set arbitrarily based on visual inspection, or else prior understanding of the data. Some statistical software or machine learning may also be used to determine the best location of knots based for the data.

Choudhury et al (2015) used piecewise regression to detrend crop yield and stated that this form may be more useful at identifying patterns and threshold points over time than linear or polynomial approaches. (Prasad et al, 2007) used the linear piecewise form to estimate wheat and rice crop yield in India and found this form to be quite promising for estimating yield.

Bateman et al (2020) used piecewise regression to determine the best planting date to maximize yield of soybeans by observing change of slope when the crops are planted at different times.

2.2.4 Generalized Additive Model (GAM) Regression Form

GAM is a very flexible form that combined the generalize linear model and additive models in which the response (dependent) variable effect is captured by smoothing function of the explanatory (independent) variable(s). There are many different smoothing functions in GAM and user can based on need (Wood, 2006). A main reason for choosing GAM over other functional forms is that it allows nonlinearity without the user having to identify which order or forms to fit.

Chen et al (2019) compared linear regression with GAM for predicting wheat yield in Australia and found that GAM model had better goodness of fit measures than linear regression. Crane-Droesch et al (2013) used GAM to estimate relationship of biochar on crop yields and some soil properties and found GAM performed quite well. Similarly, (Marcillo et al, 2021) used GAM to predict soybean time to maturity. Bera et al (2021) compared linear, polynomial, and GAM functional forms to measure canopy cover of forestry in India and found that non-linear forms performed better than linear, though by a small margin.

2.3 DATA

2.3.1 Crop Yield Data

For the dependent variable, crop yield, this study focuses on analyzing four U.S crops including corn, soybeans, winter wheat, and spring wheat. Data are obtained from the USDA NASS database and are measured as bushels per acre yield. This county crop yield data spans from 2008 to 2019 for total of 12 years of crop yield data for each crop. The data initially cover all 48 states in the United States except Hawaii and Alaska, though different U.S states are included, based on where the crops are grown. Also, even though the USDA NASS database include most counties that produce corn and soybeans, only some counties that produce spring wheat and winter wheat are included. Therefore, there will be a lesser number of counties analyzed for spring wheat and winter wheat.

Data for corn covers 28 states, which are Alabama, Arkansas, Colorado, Delaware, Georgia, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Michigan, Minnesota, Mississippi, Missouri, Nebraska, New York, North Carolina, North Dakota, Ohio, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Virginia, and Wisconsin. There are a total of 651 counties for corn, giving total of 7812 data observed for 12 years. For soybeans, there are 27 states, which are Alabama, Arkansas, Delaware, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Michigan, Minnesota, Mississippi, Missouri, Nebraska, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, South Carolina, South Dakota, Tennessee, Virginia, and Wisconsin. There are 7524 data observation for soybeans, coming from 627 counties over 12 years.

There are six states for Spring Wheat, which are Minnesota, Idaho, Montana, North Dakota, South Dakota, and Washington. This gives total data of 384 coming from 32 counties over 12 years. There are 19 states included for Winter Wheat data, which are Colorado, California, Idaho, Illinois, Indiana, Kansas, Maryland, Michigan, Missouri, Montana, Nebraska, North Carolina, Ohio, Oklahoma, Oregon, South Dakota, Texas, Virginia, and Washington, giving total data of 2196 observation across 183 counties over 12 years.

2.3.2 NDVI Data

NDVI is used as the main independent variable. For this and other vegetation indices data, the USDA NASS provide a program called Cropland Data Layer (CDL). CDL uses satellite imagery to identify major crop types and to produce digital, crop-specific, categorized geo-referenced output products to the Agricultural Statistics Board. Using NASA’s Moderate

Resolution Imaging Spectroradiometer (MODIS), the vegetation indices data series (e.g NDVI) are extracted from location specified by the Cropland Data Layer. MODIS then creates 16-day vegetation index products based on different spectral reflectance bands, and for this paper, 250m spectral band resolution level is used. The reflectance bands are measured daily and sometimes there may be some obstruction, such as cloud cover, that may cause errors to the data. MODIS will apply algorithms to choose best pixel value over the 16-days period, creating less errors in the image, correlations, etc.

There are many NDVI metrics that can be used to estimate crop yield (Wilton, 2021) but for this study, the maximum NDVI will be considered because it is known to be relatively accurate in measuring productivity (biomass) and crop yield. First, the NDVI data is averaged over the county, selecting for the maximum value of the index. Then it is merged with the corresponding county crop yield per year. Through this process, the maximum vegetation index series (e.g MaxNDVI), along with corresponding yield for each county, the dataset is completed.

2.4 METHODOLOGY

The basic goal of this study is to use NDVI (main independent variable) to estimate crop yield (dependent variable), using various regression functional forms for comparison. Three steps are used:

Step 1. Processing the satellite data to select for maximum NDVI and its corresponding crop yield data for all counties across the United States from 2008 to 2019.

Step 2. Fit data into different regression forms using the software R and plot the results.

Step 3. Compare the regression forms based on visual inspection, coefficient of determination, R^2, and whether the functional form appears theoretically correct.

In general, the above are the steps to generate results for this paper. Although within *Step 1* and *Step 2*, some preliminary statistical tests are also conducted to ensure that estimation is done acceptably.

2.4.1 Comparing Eight Regression Functional Forms for Estimating Crop Yield using NDVI

This paper compares eight different functional forms: the linear regression, polynomial regression (quadratic and cubic), piecewise/segmented regression (linear piecewise, quadratic piecewise, cubic spline, and natural spline), and Generalized Additive Model (GAM). Crop yield (dependent variable) is in bushels per acre and the main explanatory (independent) variable is NDVI. Also a Time trend is used to capture improvement in technology that may influence crop production, and a U.S State dummy variable is used to capture yield variation across states, due to soil type, weather, and other factors.

2.4.1.1 Linear Regression Form

Linear regression is the most basic functional form used here to estimate the relationship between variables.

$$Yield_t = \alpha_t + \beta_{1,t} NDVI + \beta_{2,t} Time + \beta_{3,t} State + \varepsilon_t \tag{2.1}$$

Equation (2.1) represents the linear regression equations used where $Yield$ is the $n \times 1$ vector of crop yields (i.e corn, soybeans, winter wheat, and spring wheat) and $NDVI$ is the vegetation index used for the analysis. *Time* is a time trend variable and *State* represent a dummy variable for each U.S State in data except one, which will act as reference state to avoid the dummy variable trap. These dummy variables are used to control for variation between states in data.

2.4.1.2 Polynomial Regression Form

The relationship between NDVI and crop yield is not linear due to several factors including the saturation problem at higher biomass density (yield). Saturation arises when crop yield increases in value, NDVI value does not increase as much, leading to a nonlinear relationship between crop yield and NDVI. Therefore, using a linear form for this study may lead to inaccurate estimation of the data. Some polynomial regression forms are therefore used here including a polynomial up to the third degree, which are quadratic and cubic regression forms.

$$Yield_t = \alpha_t + \beta_{1,t}NDVI + \beta_{2,t}NDVI^2 + \beta_{3,t}Time + \beta_{4,t}State + \varepsilon_t \tag{2.2}$$

$$Yield_t = \alpha_t + \beta_{1,t}NDVI + \beta_{2,t}NDVI^2 + \beta_{3,t}NDVI^3 + \beta_{4,t}Time + \beta_{5,t}State + \varepsilon_t \tag{2.3}$$

Equation (2.2), and (2.3) represent the quadratic regression and cubic regression equations used respectively, where $Yield$ is the $n \times 1$ vector of crop yields (i.e corn, soybeans, winter wheat, and spring wheat) and $NDVI$ is the vegetation index used for the analysis. *Time* is a time trend variable and *State* is a dummy variable.

2.4.1.3 Piecewise (Segmented) Regressions Form

Piecewise/segmented regression allows setting of breakpoints to divide the data into “segments” and fit separate regression lines through each one. For example, if the data looks linear (e.g in low yield and low NDVI) for the first half but quadratic for the rest (e.g in high yield and high NDVI), then by making use of the dummy variables feature in piecewise, one can address such situation and fit the data better.

Equation (2.4) and (2.5) represent the linear piecewise and quadratic piecewise equations used in this paper. Basically, piecewise regressions are created by specifying a dummy variable for each side of breakpoint/knot value and to avoid discontinuity problem, the piecewise functions are constrained to be continuous for the entire domain.

$$Yield_t = \alpha + \beta_{1,t}NDVI + \beta_{2,t}\big(NDVI - x(k)\big)X_k + \beta_{3,t}Time + \beta_{4,t}State + \varepsilon_t \quad (2.4)$$

$$Yield_t = \alpha + \beta_{1,t}NDVI + \beta_{2,t}\,NDVI^2 + \beta_{3,t}\big(NDVI - x(k)\big)X_k + \beta_{4,t}\big(NDVI - x(k)\big)^2 X_k + \beta_{5,t}Time + \beta_{6,t}State + \varepsilon_t \quad (2.5)$$

where $$X_k = \begin{cases} 0 \; if \; NDVI \le x(k) \\ 1 \; if \; NDVI > x(k) \end{cases}$$

The *x(k)* term represents the location of breakpoint or commonly referred as the knot value. In this paper, the knot value for linear piecewise and quadratic piecewise will be set based on visual observation on crop yield versus NDVI plot to determine best location for specifying the breakpoint/knot. More details on knot value selection will be discussed later. The X_k term represents dummy variable depending on what side the values fall from the knot. In here, the dummy variable X_k will be 0 if NDVI value is less than equal to the specified knot value, x(k) and the dummy variable X_k will be 1 if NDVI value is more than the specified knot value.

Another extension of piecewise regressions that will be considered in this study are the cubic spline and natural spline. Cubic spline is a third-order polynomial piecewise that enforce continuous first and second derivatives at the knot, meaning the function has continuous slope and continuous slope of the slope. But one attribute of the cubic spline that may be undesirable is that it has higher variability in the boundary region, leading to higher uncertainty (standard errors). Such a problem can be addressed by using a natural cubic spline or commonly referred to as natural spline that imposed another constraint on top of constraints for cubic spline, that the function will be linear in the boundary region. This constraint smooths the boundary region, making it more robust between data points that leads to lower variability.

The cubic spline and natural spline are fitted using *bs()* function and *ns()* function respectively in R software included in package *splines*. The dependent variable for these two functional forms is crop yield and independent variables are NDVI as well as Time and State dummy variables. In linear piecewise and quadratic piecewise only one knot value is used to make the model simple but for the cubic spline and natural spline, a few knot values will be considered to see whether applying more knots will improve the fitness of the functional forms. Looking at Figure 2.1, corn and soybeans knot location seems apparent, for example.

2.4.1.4 Generalized Additive Model (GAM) Regression

If one does not want to assume any form, but knows that the data has nonlinearity, the Generalized Additive Model (GAM) may be useful because it is very flexible. GAM combines the generalized linear model and additive models in which the response variable effect is captured by smoothing function of the explanatory variable(s). GAM is fitted using the *mgcv* package available in R and the thin-plate spline smoother function will be applied as the default choice (Wood, 2006). Similar to cubic spline and natural spline, Time and State dummy variables will be added as independent (explanatory) variables in this functional form.

2.5 RESULTS

This study compares eight different regression functional forms: the linear regression, polynomial regression (quadratic and cubic), piecewise/segmented regression (linear piecewise, quadratic piecewise, cubic spline and natural spline), and Generalized Additive Model (GAM). Crop yield (dependent variable) is estimated for corn, soybeans, spring wheat, and winter wheat. The independent variables used are NDVI (main variable), and Time trend, and U.S State dummy variables.

There are three approaches here for assessing functional form suitability: goodness of fit (adjusted R^2 value), visual inspection, and theoretical considerations. The intercept will be a focus as theoretically, in a single variable model with crop yield as the dependent variable and NDVI is the independent variable, if NDVI is zero, then crop yield should likely be somewhere around zero as well. Therefore, the intercept should likely be within range of zero. In other words, if NDVI value is zero, then estimated crop yield value should likely be not too far from zero as well.

The Augmented Dick-Fuller with lag one test was done to check for stationarity of the data. Based on the results, the p-value for all variables is small (< 0.05), therefore the null hypothesis can be rejected, concluding that the data does not have unit root problem. To correct for heteroscedasticity, robust standard error for all functional forms were used (White, 1980). In general, robust standard errors generated were about the same or slightly higher than the normal standard errors for all crops and functional forms.

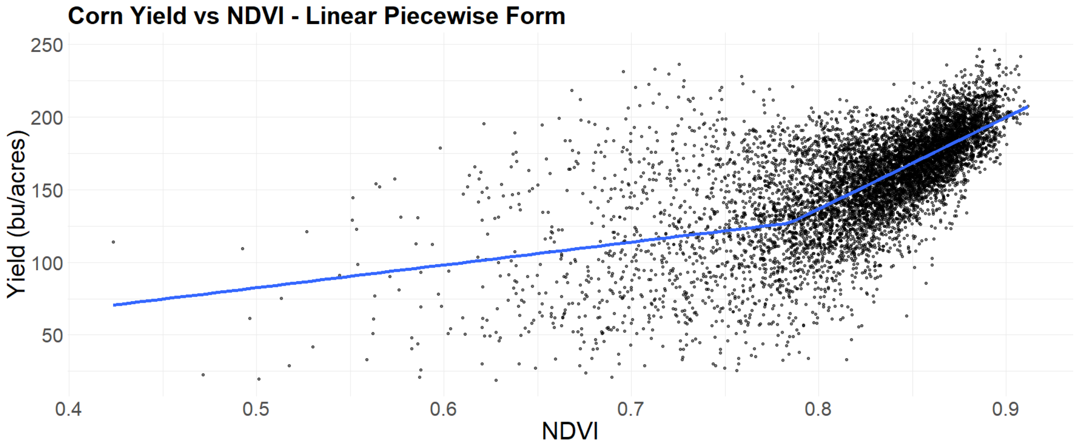
2.5.1 Corn: Regression Functional Form Results

Figure 2.1 shows plots of crop yield with NDVI for all four crops using linear piecewise regression. Results for corn shows that the linear regression functional form provides poorer goodness of fit (adjusted R^2) in estimating crop yield using NDVI. The linear piecewise regression form on the other hand, provided a better adjusted R^2 value and intercept value, that is suitable.

In general Table 2.2 shows that, the Time Trend variable is significant and has positive coefficients for all functional forms, as expected. This means that corn yield increased over time. This was due to several reasons such as improved technology, improved crop management, and improved seeds and inputs over time. For corn, Alabama is the reference state for the U.S State dummy variable and in general, there were at least 15 states out of 26 included that are significantly different (<0.05) from corn yield in Alabama. About half of those have a positive coefficient, indicating higher yields than Alabama, which may be due to several factors such as weather, soil conditions, and geographic location related factors.

Figure 2.1 *All Four U.S. Crops: Linear Piecewise Regression Function for NDVI Plotted for Crop Yield vs NDVI data, 2008 to 2019*

(a)



(b)

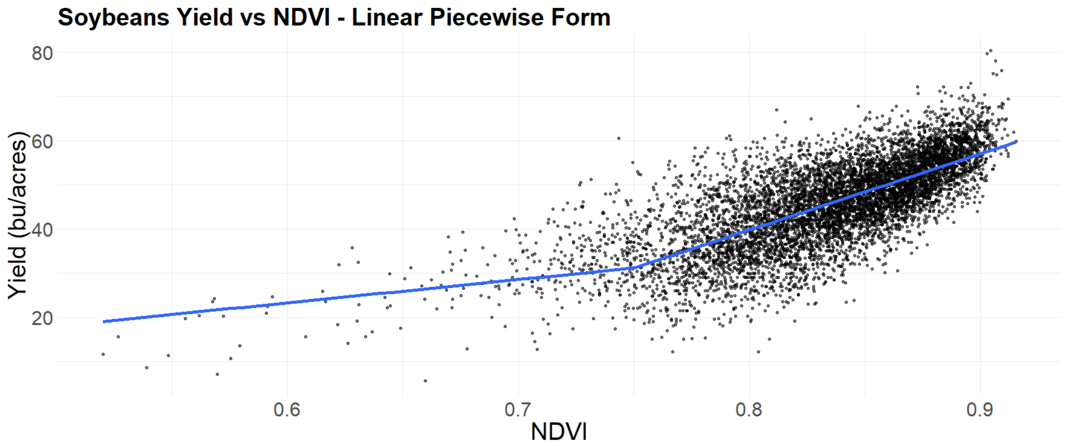
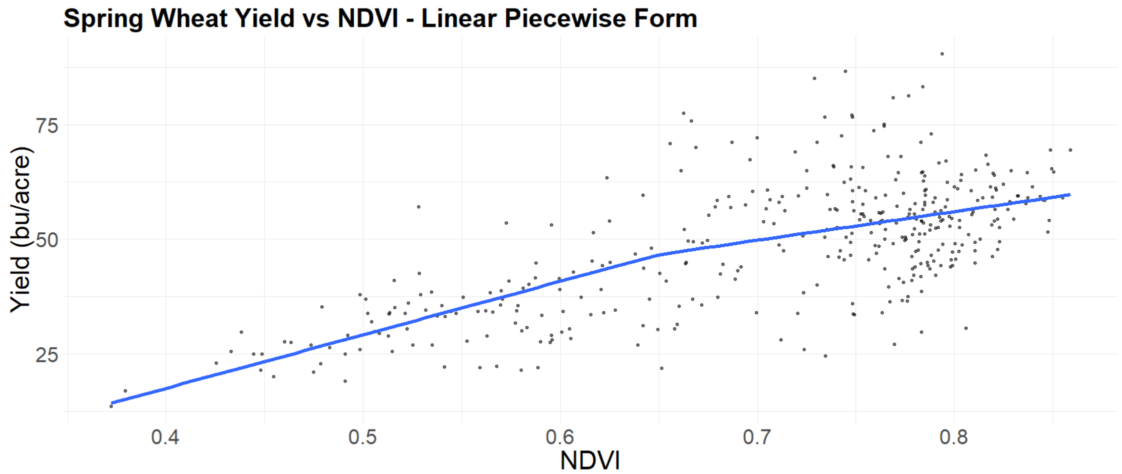
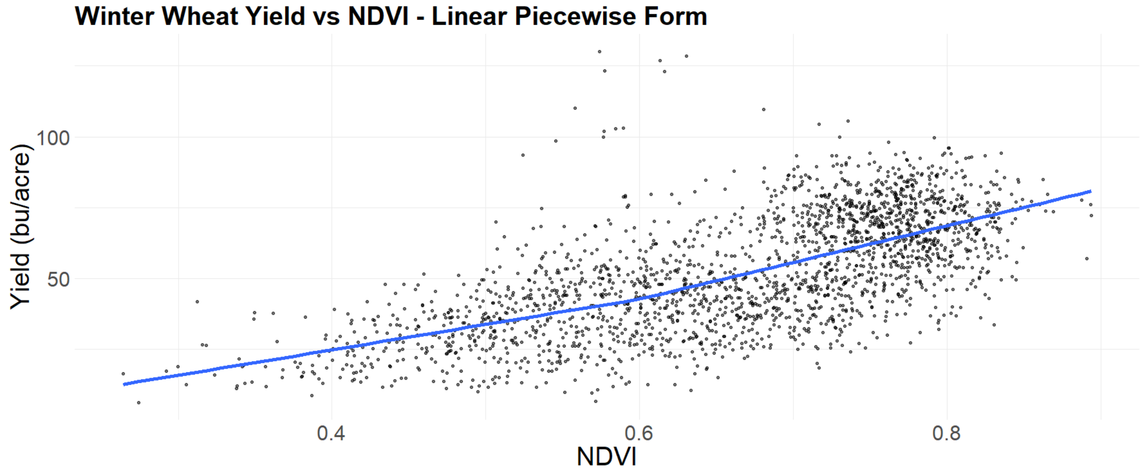


Figure 2.1 (continued)

(c)



(d)



Notes: Figures (a), (b), (c), (d) Linear Piecewise Regression for corn, soybeans, spring wheat, and winter wheat respectively. Knot values (x(k)) used for corn: NDVI = 0.785, soybeans: NDVI = 0.75, spring wheat: NDVI = 0.65, for winter wheat: NDVI = 0.6.

Note: Corn: N=7812 (651 counties x 12 years), Soybeans: N=7524 (627 countries x 12 years), Spring Wheat: N=384 (32 counties x 12 years), and Winter Wheat: N=2196 (183 counties x 12 years).

Note: Compared to all eight functional forms used in this study, linear piecewise is able to provide suitable balance between the visual nonlinear relationship between NDVI and yield at higher NDVI levels, suitable intercept and sufficiently high adjusted R^2 value.

The relationship between corn yield and NDVI became more curved (greater upward slope) with NDVI above approximately 0.7, but more linear below that. The selected knot value found was x(k):NDVI=0.785 for linear piecewise functional form (Figure 2.1) and x(k):NDVI=0.775 for quadratic piecewise functional form (not shown Figure 2.1). Knot values for the piecewise cubic spline and natural spline were selected to be NDVI=0.5,0.7,0.85 (values not shown in Figure 2.1 for both functions).

Table 2.1 *All Four U.S. Crops: Adjusted R^2 Values for All Functional Forms [Dependent Variable: Crop Yield, Independent Variables: NDVI, U.S State Dummy Variables, Time Trend], 2008 to 2019*

Corn

Forms	R^2	Adj. R^2	F-statistic	p-value	Residual Std. Error
Linear	0.6308	0.6294	458.5 (df=29; 7782)	$<2.2\text{x}10^{-16}$	21.93 (df=7782)
Quadratic	0.6524	0.6511	486.9 (df=30; 7781)	$<2.2\text{x}10^{-16}$	21.28 (df=7781)
Cubic	0.6524	0.6511	486.9 (df=30; 7781)	$<2.2\text{x}10^{-16}$	21.28 (df=7781)
Linear Piecewise	0.6497	0.6483	481 (df=30; 7781)	$<2.2\text{x}10^{-16}$	21.36 (df=7781)
Quadratic Piecewise	0.6533	0.6519	458 (df=32; 7779)	$<2.2\text{x}10^{-16}$	21.26 (df=7779)
Cubic Spline	0.654	0.6525	432.4 (df=34; 7777)	$<2.2\text{x}10^{-16}$	21.24 (df=7777)
Natural Spline	0.6535	0.652	458.4 (df=32; 7779)	$<2.2\text{x}10^{-16}$	21.25 (df=7779)
GAM	NA	0.653	498.9	$<2\text{x}10^{-16}$	NA

Soybeans

Forms	R^2	Adj. R^2	F-statistic	p-value	Residual Std. Error
Linear	0.6803	0.6791	569.5 (df= 28; 7495)	$<2.2\text{x}10^{-16}$	5.615 (df= 7495)
Quadratic	0.6858	0.6846	564 (df= 29; 7494)	$<2.2\text{x}10^{-16}$	5.566 (df= 7494)
Cubic	0.6885	0.6872	552 (df= 30; 7493)	$<2.2\text{x}10^{-16}$	5.543 df= 7493
Linear Piecewise	0.6825	0.6813	555.6 (df=29; 7494)	$<2.2\text{x}10^{-16}$	5.595 (df= 7494)
Quadratic Piecewise	0.6898	0.6886	537.5 (df= 31; 7492)	$<2.2\text{x}10^{-16}$	5.531 (df= 7492)
Cubic Spline	0.6898	0.6885	504.8 (df= 333; 7490)	$<2.2\text{x}10^{-16}$	5.532 (df= 7490)
Natural Spline	0.6894	0.6881	536.4 (df=31; 7492)	$<2.2\text{x}10^{-16}$	5.535 (df=7492)
GAM	NA	0.689	586.3	$<2.2\text{x}10^{-16}$	NA

Note: Knot value selected (x(k)): Corn for Linear Piecewise: NDVI = 0.785, for Quadratic Piecewise: NDVI = 0.775, for Cubic Spline and Natural Spline: NDVI = 0.5, 0.7, 0.85. Soybeans for Linear Piecewise: NDVI = 0.75, for Quadratic Piecewise: NDVI = 0.8, for Cubic Spline and Natural Spline: NDVI = 0.65, 0.75, 0.85.

Table 2.1 (continued)

Spring Wheat

Forms	R^2	Adj. R^2	F-statistic	p-value	Residual Std. Error
Linear	0.5724	0.5644	71.89 (df=7; 376)	$<2.2x10^{-16}$	9.299 (df=376)
Quadratic	0.5724	0.5632	62.74 (df=8; 375)	$<2.2x10^{-16}$	9.311 (df= 375)
Cubic	0.5828	0.5727	58.04 (df= 9; 374)	$<2.2x10^{-16}$	9.21 (df= 374)
Linear Piecewise	0.5724	0.5633	62.74 (df= 8; 375)	$<2.2x10^{-16}$	9.311 (df= 375)
Quadratic Piecewise	0.5948	0.5839	54.75 (df= 10; 373)	$<2.2x10^{-16}$	9.088 (df= 373)
Cubic Spline	0.6013	0.5896	51.01 (df= 11; 372)	$<2.2x10^{-16}$	9.027 (df= 372)
Natural Spline	0.592	0.5822	60.29 (df= 9; 374)	$<2.2x10^{-16}$	9.107 (df= 374)
GAM	NA	0.591	40.97	$<2.2x10^{-16}$	NA

Winter Wheat

Forms	R^2	Adj. R^2	F-statistic	p-value	Residual Std. Error
Linear	0.7399	0.7375	309.3 (df=20; 2175)	$<2.2x10^{-16}$	10.44 (df=2175)
Quadratic	0.7451	0.7427	302.7 (df= 21; 2174)	$<2.2x10^{-16}$	10.34 (df= 2174)
Cubic	0.7455	0.7429	289.3 (df=22; 2173)	$<2.2x10^{-16}$	10.33 (df=2173)
Linear Piecewise	0.7469	0.7444	305.4 (df=21; 2174)	$<2.2x10^{-16}$	10.3 (df= 2174)
Quadratic Piecewise	0.7471	0.7444	279 (df=23; 2172)	$<2.2x10^{-16}$	10.3 (df=2172)
Cubic Spline	0.7481	0.7453	268.6 (df= 24; 2171)	$<2.2x10^{-16}$	10.28 (df=2171)
Natural Spline	0.7461	0.7435	290.2 (df= 22; 2173)	$<2.2x10^{-16}$	10.32 (df= 2173)
GAM	NA	0.746	58.95	$<2.2x10^{-16}$	NA

Note: Knot value selected (x(k)): Spring Wheat for Linear Piecewise: NDVI = 0.65, for Quadratic Piecewise : NDVI = 0.625, for Cubic Spline and Natural Spline : NDVI = 0.6, 0.7. Winter Wheat for Linear Piecewise: NDVI = 0.6, for Quadratic Piecewise: NDVI = 0.561, for Cubic Spline and Natural Spline: NDVI = 0.5, 0.7.

Note: For estimating crop yield using NDVI (and U.S States dummy variables and a time trend), based on goodness of fit, adjusted R^2, all functional forms fit fairly similar. Though linear functional form for corn showed a slightly worse fit than the other functional forms.

Based on Table 2.1, the Generalized Additive Model had the best adjusted R^2 (0.653), followed by Cubic Spline (0.6525), Natural Spline (0.652) and Quadratic Piecewise (0.6519), so all had about the same values. Table 2.2 shows regression coefficients and its p-values for all variables in linear piecewise regression for all crops.

Table 2.2 *All Four U.S. Crops: Regression Coefficients and p-values for Linear Piecewise Regression [Dependent Variable: Crop Yield; Independent Variables: NDVI, U.S State dummy variables, Time Trend], 2008 to 2019*

CORN

Linear Piecewise Form

Variable	Coefficient	p-value
Intercept	-40.8312	$1.56x10^{-5}$
NDVI	200.19986	$<2x10^{-16}$
(NDVI-x(k)) Xk	376.03986	$<2x10^{-16}$
Time trend	2.05974	$<2x10^{-16}$
Arkansas	36.88095	$<2x10^{-16}$
Colorado	72.09034	$<2x10^{-16}$
Delaware	8.82498	0.043097
Georgia	60.85253	$<2x10^{-16}$
Illinois	1.17465	0.667332
Indiana	-7.31766	0.007431
Iowa	5.06077	0.059834
Kansas	-3.60773	0.276288
Kentucky	-4.40288	0.115428
Louisiana	32.33108	$<2x10^{-16}$
Maryland	-0.69483	0.839847
Michigan	1.06671	0.705263
Minnesota	2.89764	0.28862
Mississippi	25.21971	$<2x10^{-16}$
Missouri	-7.9432	0.009214
Nebraska	17.07632	$1.76x10^{-10}$
New York	-18.44222	$2.63x10^{-7}$
North Carolina	-28.17074	$<2x10^{-16}$
North Dakota	-6.34055	0.111799
Ohio	-2.38445	0.377905
Pennsylvania	-11.2351	0.000131
South Carolina	-23.8663	$2.87x10^{-15}$
South Dakota	-1.44754	0.629107
Tennessee	-2.59251	0.370488
Texas	-11.57004	0.000142
Virginia	-14.03108	$2.12x10^{-6}$
Wisconsin	-4.21628	0.120166

SOYBEANS

Linear Piecewise Form

Variable	Coefficient	p-value
Intercept	-35.8063	$3.34x10^{-15}$
NDVI	83.7218	$<2x10^{-16}$
(NDVI-x(k)) Xk	50.5228	$2.07x10^{-13}$
Time Trend	0.6607	$<2x10^{-16}$
Arkansas	7.5486	$<2x10^{-16}$
Delaware	0.0173	0.9924
Illinois	5.3728	$2.15x10^{-10}$
Indiana	5.0235	$3.28x10^{-9}$
Iowa	5.0327	$2.13x10^{-9}$
Kansas	4.5577	$4.31x10^{-7}$
Kentucky	2.8687	0.0010
Louisiana	8.3723	$<2x10^{-16}$
Maryland	3.2445	0.0028
Michigan	2.2119	0.0103
Minnesota	0.3612	0.6714
Mississippi	2.8841	0.0008
Missouri	1.6207	0.0672
Nebraska	8.6694	$<2x10^{-16}$
New Jersey	0.1233	0.9140
New York	3.4424	0.0003
North Carolina	-5.5861	$8.68x10^{-11}$
North Dakota	-3.6268	$5.42x10^{-5}$
Ohio	5.4712	$1.03x10^{-10}$
Oklahoma	-1.4481	0.1890
Pennsylvania	6.5792	$6.68x10^{-12}$
South Carolina	-6.0261	$3.2x10^{-10}$
South Dakota	2.3597	0.0125
Tennessee	0.9646	0.2866
Virginia	-0.7481	0.4143
Wisconsin	2.2974	0.0084

Table 2.2 (continued)

SPRING WHEAT

Linear Piecewise Form

Variable	Coefficient	p-value
Intercept	-47.5097	$5.75x10^{-9}$
NDVI	127.962	$<2x10^{-16}$
(NDVI-x(k)) Xk	2.9172	0.895531
Time Trend	0.3478	0.012708
Idaho	9.7389	$6.02x10^{-5}$
Montana	10.5498	0.000533
North Dakota	-5.168	0.00537
South Dakota	2.4556	0.295467
Washington	7.513	0.021703

WINTER WHEAT

Linear Piecewise Form

Variable	Coefficient	p-value
Intercept	-7.28892	0.01125
NDVI	88.85636	$<2x10^{-16}$
(NDVI-x(k)) Xk	-68.16538	$1.36x10^{-14}$
Time Trend	0.55612	$<2x10^{-16}$
California	17.43799	$4.01x10^{-7}$
Idaho	28.59773	$<2x10^{-16}$
Illinois	15.72373	$5.94x10^{-13}$
Indiana	17.54621	$<2x10^{-16}$
Kansas	-6.84442	0.00012
Maryland	10.39843	0.00283
Michigan	20.36414	$<2x10^{-16}$
Missouri	3.53415	0.12501
Montana	-3.02028	0.09522
Nebraska	0.63104	0.73190
North Carolina	-0.52886	0.81313
Ohio	16.06124	$<2x10^{-16}$
Oklahoma	-18.29855	<2x10-16
Oregon	3.12246	0.13462
South Dakota	6.76286	0.01246
Texas	-12.94177	$3.56x10^{-15}$
Virginia	7.91957	0.00145
Washington	20.08080	$<2x10^{-16}$

Note: Knot values (x(k)) used for corn: NDVI = 0.775, soybeans: NDVI = 0.8, spring wheat: NDVI = 0.625, for winter wheat: NDVI = 0.561.

Note: As expected, Time Trend variable had positive coefficients for all functional forms, meaning yield increased over time. U.S State dummy variables are mostly significant and positive for all crops. This means most states had higher crop yield than their reference states. Alabama is the reference state for corn and soybeans, Minnesota is the reference state for spring wheat, and Colorado is the reference state for winter wheat.

Linear piecewise and GAM seem to provide the most suitable functional forms to estimate crop yield. However, the linear piecewise regression was able to estimate the relationship of corn yield and NDVI slightly better than other functional forms. A possible challenge of the GAM function is the tendency to overfit the data, given that it has a relatively large number of parameters. The linear piecewise had the best balance between relatively simple functional form, higher adjusted R^2 value and suitable intercept value that is within a suitable range.

2.5.2 Soybeans: Regression Functional Form Results

From Table 2.1, adjusted R^2 values showed that the linear function fit was very similar to fit of other functions. However, the linear piecewise regression functional worked quite well, with a suitable R^2 value and an intercept value is suitable.

The selected knot value was $x(k): NDVI = 0.75$ for the linear piecewise form (Figure 2.1), for Cubic Spline and Natural Spline: $x(k): NDVI = 0.65, 0.75, 0.85$ (not shown in Figure 2.1) and $x(k): NDVI = 0.8$ for quadratic piecewise form (not shown in Figure 2.1), similar to corn values. These knot values provided sufficient balance between higher adjusted R^2 value and intercept value that is suitable. Based on Table 2.1, all eight functions had similar adjusted R^2 values.

The linear piecewise and GAM form seem to provide a more suitable fit for soybeans data and capture increase in slope as NDVI value gets higher. Overall, the piecewise regressions were able to estimate the relationship of soybeans yield and NDVI better than other functional forms. While the GAM function appeared reasonable, it may overfit the data, as it uses more parameters.

The linear piecewise offered the best balance between relatively simple functional forms, higher adjusted R^2 value and intercept value relatively suitable for estimating soybeans yield using NDVI.

2.5.3 Spring Wheat: Regression Functional Form Results

Results for spring wheat (Table 2.1 and Figure 2.1) are different than for corn and soybeans. The linear regression form may be appropriate for estimating spring wheat yield due to more linear relationship between spring wheat yield and NDVI, based on visual inspection of plots. But in general, the piecewise regression forms are able to provide better results of adjusted R^2 value and intercept value that is suitable.

The knot value selected was $x(k){:}\,NDVI = 0.65$ for linear piecewise form (Figure 2.1), for Cubic Spline and Natural Spline $x(k){:}\,NDVI = 0.6, 0.7$ (not shown in Figure 2.1) and $x(k){:}\,NDVI = 0.625$ for quadratic piecewise form (not shown in Figure 2.1). These knot values provided balance between a higher adjusted R^2 value and reasonable intercept value.

Based on Table 2.1, the GAM has adjusted R^2 (0.591) value, Cubic Spline (0.5896), Quadratic Piecewise (0.5839), and Natural Spline (0.5822), were all fairly similar.

2.5.4 Winter Wheat: Regression Functional Form Results

The linear regression form looks suitable in estimating winter wheat yield using NDVI. However, the piecewise regression forms able to provide a relatively suitable balance between adjusted R^2 value and intercept value closer to that is suitable.

The knot value selected was $x(k){:}\,NDVI = 0.6$ for linear piecewise form (Figure 2.1), for Cubic Spline and Natural Spline: $x(k){:}\,NDVI = 0.5, 0.7$ (not shown in Figure 2.1) and $x(k){:}\,NDVI = 0.561$ for quadratic piecewise form (not shown in Figure 2.1).

Based on Table 2.1, the Generalized Additive Model with adjusted R^2 (0.746) value, followed by Cubic Spline (0.7453), Quadratic Piecewise (0.7444), and Linear Piecewise (0.7444). GAM had the highest R^2 though not much higher than other functional forms.

Similar to spring wheat, nonlinearity is not as observable for winter wheat compared to corn and soybeans. This may be due wheat having less biomass density. Results also show that intercepts of linear piecewise for winter wheat have a negative value, similar to other crops. Although linear regression may be a suitable functional form to estimate winter wheat yield, the linear piecewise functional form may provide a suitable balance between relatively simple functional forms and an intercept with a reasonable value.

2.5.5 Results Summary

The goal of this study was to compare various regression functional forms for estimating crop yield (dependent variable) using satellite based NDVI (main independent variable), and a Time Trend and U.S State dummy variables were also included. The linear regression functional form was not suitable for all crops except spring wheat while GAM may require a larger number of parameters and may overfit the data. The linear piecewise regression provides the best balance between adjusted R^2 values and intercept value for all crops. Therefore, given linear piecewise provides a higher adjusted R^2 and is theoretically suitable, the linear piecewise functional form may be the appropriate functional form to estimate four crop yields here using NDVI.

2.6 SUMMARY

The objective of this study is to compare various regression functional forms for estimating crop yield (dependent variable) using satellite based NDVI (main independent variable). There are many advantages of using satellite based remote sensing to estimate crop yield compared to the traditional approach of using surveys. These advantages include lower cost, faster in obtaining data and updating data fairly frequently. Also using data from remote sensing is relatively accurate.

The linear regression form is the most commonly used functional form due to its simplicity in form and interpretation. But in real applications, the relationship between NDVI and crop yield maybe not linear due to several factors such as saturation of NDVI. In other words, at high levels of NDVI, crop yield may increase but NDVI may not increase enough in response. Therefore, this paper examined eight functional forms: linear regression, polynomial regression (quadratic and cubic), piecewise/segmented regression (linear piecewise, quadratic piecewise, cubic spline and natural spline), and Generalized Additive Model (GAM). The purpose was to estimate crop yield using NDVI and examine how those functional forms compare to the linear regression functional form.

Data for the crop yield was from USDA NASS database and was measured in bushel per acre. These county crop yields were from 2008 to 2019 for total of 12 years for corn, soybeans, spring wheat, and winter wheat. For each crop, the data cover all 48 states in the United States except Hawaii and Alaska, although fewer states are used as some states grow very small amounts of some of the crops. Data for NDVI was obtained using the MODIS at the 250m resolution level. The NDVI data was averaged over the county. Then NDVI data is merged with the corresponding county crop yield each year and maximum NDVI (i.e MaxNDVI) values is used for the analysis.

The methodology of this study can be summarized into three steps. First, the data is processed to select for maximum NDVI and its corresponding crop yield data for all counties across the United States from 2008 to 2019. The second step is to fit the data with various regression functional forms using the software R and plot the results. Third, the functional forms are compared based on visual inspection, coefficient of determination (R^2), and theoretical considerations. A dummy variable for U.S states and a time trend are also included as independent variables.

Results show that the crop yield vs NDVI data plot exhibits nonlinear relationship. This can be more obviously seen for corn, soybeans, and winter wheat crop. However, for spring wheat, the yield and NDVI relationship is somewhat more linear. For piecewise/spline regressions, knot value placements were selected based on visual inspection. Generally, the Generalized Additive Model (GAM) resulted in a slightly higher adjusted R^2 value compared to all eight functional forms tested for all crops, but GAM may have the tendency to overfit the data as this function uses a larger number of parameters. Instead, the linear piecewise functional form provides a suitable balance between producing a sufficiently high adjusted R^2 value and an intercept value that is reasonable.

Overall, this study may be useful for those who seek to understand how different regression functional forms may help to estimate crop yield and identify which functional forms may be better for each crop.

CHAPTER 3

COMPARING VARIOUS VEGETATION INDICES FOR ESTIMATING CROP YIELD USING TWO REGRESSION FUNCTIONAL FORMS

3.1 INTRODUCTION

The objective of this study is to compare 10 satellite-based vegetation indices for estimating crop yield using two functional forms. The study uses crop yield as the dependent variable and vegetation index (e.g SAVI, etc) as the main independent variables (and also time trend and dummy variables for U.S states).

One of the most well-known satellite based indices is NDVI, though it has a few known limitations that can cause inaccuracy in estimating crop yield. Therefore, there is a need to explore other vegetation indices that may possibly overcome NDVI limitations.

The Normalized Difference Vegetation Index (NDVI) is the most commonly used index in agriculture and ecology due to its simpler calculations, relying solely on vegetation reflectance and absorption of ultraviolet region on red and near-infrared (NIR). NDVI, also called measure of 'greenness' due to its correlation with photosynthetic capacity, has values ranging from -1.00 to +1.00 with higher values implying denser vegetation (e.g higher crop yield) in the area (Pettorelli, 2013). Despite the advantages of NDVI, some limitations of this index include sensitivity towards atmospheric contaminants, sensitivity towards soil-reflectance for sparsely vegetated areas, and saturation in densely vegetated areas (Pettorelli, 2013). The saturation problem in NDVI is due to absorption saturation in red and reflectance saturation of NIR in high biomass (yield), leading to sensitivity degradation of the vegetation index from moderate to high biomass (Dong et al, 2015). For the soil background sensitivity, it was noted that dark soil often resulted in higher vegetation indices value compared to the bright-colored soil (Mróz and Sobieraj, 2004) for this type of vegetation index. Therefore, soil conditions such as the water content and organic matter may influence index values. Also, the atmospheric effect may create “noise” as molecules, such as gasses or aerosols, influence the radiation and absorption activity and cause variation in the vegetation index from proper values (Myneni and Asrar, 1994).

Currently, there are numerous types of vegetation indices for monitoring and analyzing vegetation and some indices may have a slight advantage over others. This study analyzes nine other indices in comparison to NDVI as the benchmark, to determine which index best estimate the crop yield. Corn, soybeans, winter wheat, and spring wheat are analyzed. Given that NDVI is the most common and studied index, this study will be using NDVI as the performance benchmark when comparing with other indices being analyzed. The nine indices include: Renormalized Difference Vegetation Index (RDVI), Transformed Difference Vegetation Index (TDVI), Wide Dynamic Range Vegetation Index (WDRVI), Two-Band Enhanced Vegetation Index (EVI2), Soil Adjusted Vegetation Index (SAVI), Green Optimized Soil Adjusted Vegetation Index (GOSAVI), Green Soil Adjusted Vegetation Index (GSAVI), Modified Soil Adjusted Vegetation Index 2 (MSAVI2), and Optimized Soil Adjusted Vegetation Index (OSAVI).

3.1.1 Objective

The objective of this study is to compare various vegetation indices for estimating crop yield using two functional forms. In the previous study (Chapter 2), several functional forms were considered to estimate crop yield. This study aims to compare performance of NDVI with the nine other vegetation indices mentioned above using two functional forms which are: linear regression form and linear piecewise form. The two forms are being used as they are considered to have a relatively suitable fit for corn, soybeans, spring wheat, and winter wheat.

The next part of this paper includes a literature review of the various vegetation indices used in this study. Next data and methodology are discussed. Finally, results are presented, and a summary of study is presented.

3.1.2 Background

Remote sensing often includes satellites that allow mapping of the earth's land surface, oceans, or atmosphere by recording the unique properties of electromagnetic energy emitted by areas or objects. This allows for distinctions and identification of the areas or objects (Khorram et al, 2012). Moderate-resolution Imaging Spectroradiometer (MODIS) is a program installed in NASA Terra and Aqua satellites that produce a wide spectral range used to generate images of earth every one to two days. There are many applications of remote sensing using MODIS for agriculture such as for field mapping, nitrogen management, and monitoring plant conditions for stress or water content. Crop yield estimation can also use vegetation indices, which exploit the vegetation property of absorbing the red and blue visible spectrum and reflecting the near-infrared, enabling them to be recognized (Pettorelli, 2013). When using vegetation indices to monitor crop development, one must be aware that vegetation reflectance is influenced by many factors including soil background, atmospheric conditions, leaf structure, and many others. Therefore different vegetation species in a particular area may respond differently towards a specific index (Gitelson, 2013).

3.1.2.1 Soil-Adjusted Vegetation Indices: [SAVI, MSAVI2, OSAVI, GOSAVI, GSAVI, TDVI, RDVI]

Different vegetation indices have been developed in order to make improvements over NDVI. The soil-adjusted vegetation index group aims to take into account the effect of soil background in arid and semi-arid environments. Sensitivity of indices towards soil brightness is important especially for areas that are sparsely vegetated. There are several indices that fall into this category and some of them are analyzed in this study, including Soil-Adjusted

Vegetation Index (SAVI), Modified Secondary Soil-Adjusted Vegetation Index (MSAVI2), Optimized Soil-Adjusted Vegetation Index (OSAVI), Green Optimized Soil Adjusted Vegetation Index (GOSAVI), and Green Soil Adjusted Vegetation Index (GSAVI).

SAVI includes a soil conditioning index (L) that ranges from 0 to 1, with a higher value implying that the soil background has no effect on the vegetation information, while the lower it is, the closer the index to NDVI (Xue and Su, 2017). Most common value for L being used in SAVI is 0.5 as it is considered the median value that can be a suitable fit in various environment. Many have identified SAVI as the appropriate index for agricultural areas (González-Dugo and Mateos, 2008; Liaqata et al, 2017; da Silva et al, 2020), but depending on the type of vegetation, soil conditions such as the amount of organic matter, and atmospheric conditions, this index may still produce some errors.

OSAVI is one extension of soil-adjusted index that does not require any preliminary knowledge of the soil line parameters (Rondeaux et al, 1996). OSAVI is not only more sensitive to leaf chlorophyll area, it removes the soil effect more effectively than others and it may also work better for agricultural crops than SAVI (Rondeaux et al, 1996; Piegari et al, 2021).

MSAVI2 is proposed to reduce the effect of bare soil on SAVI and incorporate other properties than the soil line principle (Xue and Su, 2017). For MSAVI2, it uses a dynamic soil adjusting factor, compared to SAVI and OSAVI, that use a fixed adjusting factor. However, all three indices are suitable to measure above ground vegetation (Celleri et al, 2019).

GOSAVI and GSAVI are also indices proposed to help eliminate the soil background effect. However, rather than using the near-infrared and red band, GOSAVI is similar to OSAVI but using the near-infrared and green band, while GSAVI is similar to SAVI but using the near-infrared and green band (Sripada et al, 2006). In terms of mapping, the GSAVI may be able to differentiate between crops and weeds better than NDVI. It also is less sensitive to changes between crop growth phases (e.g changes from bare soil to crop presence) (Stroppiana et al, 2018).

Another index that was proposed to be better on sparsely vegetated area is Transformed Difference Vegetation Index (TDVI). It is claimed to not only be minimally affected by soil background compared to SAVI, but will not saturate as easy as NDVI in a densely vegetated area (Bannari et al, 2002).

The Renormalized Difference Vegetation Index (RDVI) is a ratio-based index that was developed to lessen the influence of soil reflectance but was derived from a different method than the SAVI group. RDVI combined the property of Band Difference (DVI) to be less affected by soil background and perform better in both low and high-density areas than NDVI (Roujean and Breon, 1995, Payero et al, 2004; Dong et al, 2015).

3.1.2.2 High Biomass and Atmospheric Resistant Index: [WDRVI, EVI2]

It is noted above that one limitation of NDVI is that it saturates as the vegetation density gets higher. Some indices were proposed to address such issues including the Wide Dynamic Range Vegetation Index (WDRVI) and Two-Band Enhanced Vegetation Index (EVI2). EVI2 is an extension of EVI (Enhanced Vegetation Index) which aimed to improve vegetation monitoring in high density area (full canopy cover) by taking advantage of blue band rather than just red and near-infrared (Vescovo et al, 2012). By using the blue band, the index become more aerosol resistant and therefore able to control for atmospheric influence better (Wang et al, 2016). EVI2 was developed to be a comparable index to EVI when the blue band is low or unavailable.

On the other hand, WDRVI was developed in trying to resolve the saturation problem, meaning this index is supposed to be better than NDVI in a high biomass area, by applying a weighted coefficient to the near-infrared reflectance, therefore increasing the dynamic range of

NDVI while still using the same bands (Xue and Su, 2017). (Towers et al, 2019) also found that in dense green biomass, e.g when NDVI is more than 0.4, WDRVI shows better accuracy and sensitivity than NDVI for several types of vegetation, including cropland.

3.2 LITERATURE REVIEW

3.2.1 Normalized Difference Vegetation Index (NDVI)

NDVI is the benchmark to be used in this study and is the most extensively used vegetation index. There are several studies throughout the years that have discussed the suitability of NDVI for remote sensing on croplands. On average, NDVI performs satisfactory in monitoring crop development, many have found the index to have a suitable correlation with crop yield in United States (Labus et al, 2002 and Becker-Reshef et al, 2010) and other parts of the world (Ren et al, 2008; Mkhabela et al, 2011; Dempewolf et al, 2013).

NDVI works reasonably well for estimating yield of various type of crops but there is a need to explore other vegetation indices that may potentially be able to address NDVI limitations, including saturation, especially for yield estimation. There are a few studies that have compared NDVI with other indices such as by (GopalaPillai and Tian, 1999; Panda et al, 2010, Kayad et al, 2019) for corn and (Bolton and Friedl, 2013) for soybeans. Stepanov et al (2020) assessed relationship of soybean yield and NDVI in the Khabarovsk District, Russia. They also incorporated different variables in their model including soil humidity and temperature, and a seasonality factor.

3.2.2 Soil-Adjusted Vegetation Indices: [SAVI, MSAVI2, OSAVI, GOSAVI, GSAVI, TDVI, RDVI]

As mentioned, NDVI has some limitations, typically in very high or very low-density areas. The SAVI was one of the pioneer indices proposed to correct soil influence on NDVI using its L factor that can be adjusted based on vegetation coverage. As the L value approaches to 0, that is when the area is densely vegetated, and SAVI approximates NDVI. Although for balance, L=0.5 is most commonly used in typical research. Some studies found SAVI to be a better index than NDVI such as (Baugh and Groeneveld, 2006), who compared the two indices in shrubland area with low vegetation cover in Colorado. While for agriculture, (da Silva et al, 2020) found that SAVI is more suitable than NDVI. They found that NDVI that is better suited for forest lands, while SAVI is better suited for agricultural and native fields.

MSAVI2 is one extension of SAVI that has self-adjusting L value so prior knowledge of vegetation density in the area is not required (Qi et al, 1994). MSAVI2 is able to raise the vegetation signal while keeping the soil noise low in both sparsely vegetated and higher density biomass areas even better than SAVI. (Baugh and Groeneveld, 2006) have found MSAVI2 to have higher a R^2 value than SAVI. (Barillé et al, 2011) conducted studies to compare background reflectance among indices and found that MSAVI2 performed better than SAVI. However, they mentioned that the two are affected by soil background color that can cause variation in both indices. The literature also noted that MSAVI2 and SAVI still saturate at higher biomass values.

Another extension of SAVI is OSAVI, which used 0.16 as the fixed adjustment L factor. According to (Celleri et al, 2019), OSAVI does not exhibit superiority compared to SAVI and MSAVI2 in sparse vegetated areas, and one reason may be because L=0.5 is the optimal value in a semiarid moderately vegetated area. OSAVI had a reasonable correlation with the vegetation in their study, but an important aspect in their findings is that OSAVI had a significant correlation with soil salinity, leading to the ability to indicate water stress in vegetation. Shang et al (2015) showed that OSAVI performed better than NDVI in their spring wheat and canola regression on estimating the green effective plant area index. They also indicated that combination of OSAVI and Transformed Chlorophyll Absorption Reflectance Index (TCARI) could potentially perform well in dense canopy leaf chlorophyll content estimation.

Other indices in the SAVI family that are potentially more robust, are GOSAVI and GSAVI. These two indices are not as extensively used as the ones mentioned previously, but there are some studies that noted the benefit of using green bands for monitoring crop growth. (Cao et al, 2015) used many indices, including GOSAVI and GSAVI, in their study to help estimate winter wheat nitrogen status in China and they mentioned that GOSAVI seemed to perform quite well to measure above ground biomass, while both GOSAVI and GSAVI performed quite well to measure plant nitrogen concentration. They also mentioned the importance of using green band indices to measure nitrogen status in vegetation as that has a high correlation with the nitrogen concentration.

(Barzin et al, 2020) showed that GOSAVI has a reasonable correlation with corn yield in their study conducted in Mississippi, but many other studies actually showed the usefulness of GOSAVI and GSAVI in evaluating nitrogen content in plants such as by (Sripada et al, 2006) for corn crops and (Sripada et al, 2007) for winter wheat crops as well. Nitrogen is an essential nutrient for crops to grow and mature so it is quite important to be able to assess plant health status for nitrogen deficiency, which impact yield. (Diego et al, 2021) emphasized this in their study regarding coffee in Brazil. Their literature indicated that among indices they used, GOSAVI was one of the best indices to differentiate between nitrogen content class in the coffee plantation and they attributed this to the use of the green band rather than red band, such as in NDVI.

There are some indices derived to address the soil influence problem as well, but more NDVI-like as they are not based on SAVI. One is RDVI and (Haboudane et al, 2004) tried to quantify the green leaf area index for corn, soybeans, and wheat crops in Canada. RDVI performed quite well as it does not saturate as easily as NDVI but tended to overestimate soybeans and corn canopies while underestimating the wheat canopy as density gets higher. (Dong et al, 2015) used RDVI as one of the vegetation indices for estimating crop fraction of absorbed photosynthetically active radiation in China, and according to their findings, RDVI does not saturate as easily in higher leaf area index compared to NDVI and performed at least as good as NDVI for winter wheat, and slightly below for corn. They also mentioned that RDVI is less influenced by background effects and is able to perform well in denser canopies.

TDVI does not have many applications in agriculture but (Chandel et al, 2021) showed that TDVI appeared to estimate alfalfa yield relatively well. (Suarez et al, 2020) used TDVI as one of vegetation indices to estimate carrot yield in Australia, and found that RDVI was the best index, while TDVI had potential to perform better than NDVI and SAVI in certain cases. (Morier et al, 2015) included TDVI to estimate potato yield in Canada and again, TDVI was not the best index used as OSAVI, MSAVI, and NDVI outperformed TDVI in this case. (Al Shidi et al, 2019) found similar results when comparing the performance of NDVI and TDVI.

3.2.3 High Biomass and Atmospheric Resistant Index: [WDRVI, EVI2]

A limitation of NDVI is the saturation effect, as the vegetation gains higher leaf area index/biomass. As NDVI is influenced by the soil-background effect, NDVI is also influenced by the atmospheric particles that can hinder accuracy of the index value.

EVI2 was developed specifically to address the issue of both saturation and atmospheric influence effect of NDVI, while still using the red and NIR bands, unlike the EVI that also included the blue band. (Sun et al, 2012) successfully mapped the major winter wheat-producing regions in China using MODIS EVI2 data to classify land-cover information that can be used for yield assessment or agricultural management. They chose EVI2 specifically due its advantages to be less influenced by soil background and atmospheric effect with less saturation compared to NDVI, while also simpler than EVI that requires the blue band which is only available on 500m MODIS resolution.

In the United States, (Bolton and Friedl, 2013) used EVI2 as one of the indices to estimate corn and soybeans yield. They indicated that EVI2 has a reasonable correlation with both corn and soybeans. Based on R^2 values, they found that EVI2 models estimate yield better than NDVI models for both crops. In their study, EVI2 is better used in non-semi arid counties for predicting corn and soybeans yield, therefore they suggested to have a different model for semi-arid and non-semi-arid areas to estimate yield, especially for corn. (Zhang and Zhang, 2016) measured global cereal production and yield using EVI2 as measure of greenness because compared to NDVI, EVI2 was able to differentiate between vegetation diversity better, measure vegetation conditions better, and perform better at estimating yield of crops including corn, soybeans, and coffee.

WDRVI is another index proposed to perform better than NDVI especially in higher biomass areas. WDRVI is quite widely used in ecology and agriculture due to lower saturation at higher leaf area index levels compared to NDVI. (Viña et al, 2004) noted this characteristic and suggested that WDRVI would be suited for use in croplands and more humid areas. Several studies proved strong relationships of WDRVI with crop yield such as (Sakamoto et al, 2013), who estimated corn yield across U.S using WDRVI, as the index is believed to have close to linear relationship with corn (Guindin-Garcia et al, 2012; Sibley et al, 2014). (Dempewolf et al, 2013) which compared indices including WDRVI to best estimate wheat yield in the Punjab Province of Pakistan and found that WDRVI exhibited less deviation with tighter distribution around mean. (de Souza et al, 2015) compared NDVI and WDRVI in their study of mapping corn and soybeans field in Brazil and found that WDRVI was able to identify crops better than NDVI.

3.3 DATA

3.3.1 Crop Yield Data

Data for crop yield used in this study is the same as Chapter 2.

3.3.2 Vegetation Index Data

Data for vegetation indices used in this study is similar to Chapter 2. Equations for all vegetation indices can be found below.

Equations for all Vegetation Indices

Index	Equation	Reference
1. NDVI	$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$	(Rouse et al, 1973)
2. RDVI	$RDVI = \frac{(NIR - Red)}{\sqrt{(NIR + Red)}}$	(Roujean and Breon, 1995)
3. TDVI	$TDVI = 1.5\left[\frac{(NIR - Red)}{\sqrt{NIR^2 + Red + 0.5}}\right]$	(Bannari et al, 2002)
4. WDRVI	$WDRVI = \frac{(a * NIR - Red)}{(a * NIR + Red)}$	(Gitelson, 2004; Henebry et al, 2004)
5. EVI2	$EVI2 = 2.5 * \frac{(NIR - RED)}{NIR + (2.4 * RED) + 1}$	(Jiang et al, 2008)
6. SAVI	$SAVI = \frac{1.5 * (NIR - Red)}{(NIR + Red + 0.5)}$	(Huete, 1988)
7. GOSAVI	$GOSAVI = \frac{NIR - Green}{NIR + Green + 0.16}$	(Sripada et al, 2005)
8. GSAVI	$GSAVI = 1.5 * \frac{(NIR - Green)}{(NIR + Green + 0.5)}$	(Sripada et al, 2005)
9. MSAVI2	$MSAVI2 = \frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8(NIR - Red)}}{2}$	(Qi et al, 1994)
10. OSAVI	$OSAVI = \frac{(NIR - Red)}{(NIR + Red + 0.16)}$	(Rondeaux et al, 1996)

3.4 METHODOLOGY

The basic goal of this study is to use various vegetation indices (main independent variable) to estimate crop yield (dependent variable) using regression with various vegetation indices for comparison, and NDVI is a benchmark.

Methodology of this study can be summarized into three steps, which are similar to Chapter 2.

3.4.1 Regression Functional Forms: [Linear Regression and Linear Piecewise Regression]

NDVI and nine other vegetation indices used: RDVI, TDVI, WDRVI, EVI2, SAVI, GOSAVI, GSAVI, MSAVI2, and OSAVI. Two functional forms are used including linear regression and linear piecewise regression. Linear piecewise regressions appear to estimate crop yield for corn, soybeans, spring wheat, and winter wheat quite well using NDVI. Piecewise regressions offer the advantage of being more flexible yet still simple at the same time. While for the linear regression form, there are two reasons for the inclusion of this functional form. First, linear regression form is the most commonly used and simplest functional form, therefore one may be curious as to how it fits using different vegetation indices.

Second, other vegetation indices such as the SAVI group or WDRVI were developed to mitigate the saturation problem appearing in NDVI. Therefore, it is expected when using vegetation indices other than NDVI that the data will behave more linear like and the linear regression form may be adequate to explain the crop yield. In all functional forms, the output (dependent variable) is crop yield (bushels per acre) and the explanatory (independent) variables are the vegetation index (VI), U.S. State dummy variables, and Time trend.

According to Chapter 2, linear piecewise regression seems to be an appropriate form to estimate crop yield using NDVI as the explanatory variable as this functional form give not only a higher R^2 value but also has intercept that is reasonable and is suitable visually. Piecewise regression forms can be hypothesized to produce a suitable estimation for corn and soybeans yield using different vegetation indices, as when using NDVI. For both spring wheat and winter wheat on the other hand, relationship between yield and NDVI appear more linear compared to corn and soybeans, so one may be able to estimate wheat yields using linear regression. The piecewise regressions may capture the slight nonlinearity in spring wheat and winter wheat.

Equation (3.1) and (3.2) represent the linear regression form and linear piecewise equations used in this study. Basically, piecewise regressions are created by specifying a dummy variable for each side of breakpoint/knot value and to avoid discontinuity problem, the piecewise functions will be constrained to be continuous for the entire domain.

$$Yield_t = \alpha + \beta_{1,t}VI + \beta_{2,t}State + \beta_{3,t}Time + \varepsilon_t \tag{3.1}$$

$$Yield_t = \alpha + \beta_{1,t}VI + \beta_{2,t}\big(VI - x(k)\big)X_k + \beta_{3,t}State + \beta_{4,t}Time + \varepsilon_t \tag{3.2}$$

where $$X_k = \begin{cases} 0 \; if \; VI \leq x(k) \\ 1 \; if \; VI > x(k) \end{cases}$$

where $Yield$ is the $n \times 1$ vector of crop yields (i.e corn, soybeans, winter wheat, and spring wheat) and *VI* represent the specific vegetation index being used. *Time* is a time trend variable and *State* represent a dummy variable for each U.S state in data except one, which will act as reference State to avoid the dummy variable trap. These dummy variables are used to control for variation between states in data.

The *x(k)* term represents the location of breakpoint or commonly referred as the knot value and as in Chapter 2. The Xk term represents the dummy variable depending on what side the values fall from the knot. Here, the dummy variable Xk will be 0 if the vegetation index value is less than equal to the specified knot value, x(k) and the dummy variable Xk will be 1 if the vegetation index value is more than the specified knot value.

3.5 RESULTS

The focus of this study is to use various satellite-based vegetation indices (independent variable) to estimate crop yield (dependent variable) using regression. Each vegetation index is fitted using two different regression forms, to determine whether the nine other vegetation indices are an improvement over the benchmark, NDVI. The performance of an index is analyzed over the two regression forms and the other vegetation indices.

There are three approaches used here for assessing functional form suitability: goodness of fit (adjusted R^2 value), visual inspection, and theoretical considerations. To correct for heteroscedasticity, a robust standard error for all functional forms is used (White, 1980).

All ten vegetation indices behaved very similarly often showing very high correlations between the various indices of around 0.90 or higher. In fact, some of the indices have very similar formulas, such as OSAVI and GOSAVI, as well as SAVI with GSAVI.

Overall, results show that the relationship between a vegetation index and crop yield is mostly nonlinear, especially for larger vegetation index values. However, there is less nonlinearity for some vegetation indices such as GOSAVI, GSAVI, and WDRVI, because those indices were developed to solve the saturation problem of NDVI. Functional forms using these indices can be expected to perform better than functional forms using NDVI (which is subject to saturation). The breakpoint/knot value for piecewise regressions is based on goodness of fit and theoretical considerations.

3.5.1 Corn: Various Vegetation Indices Results

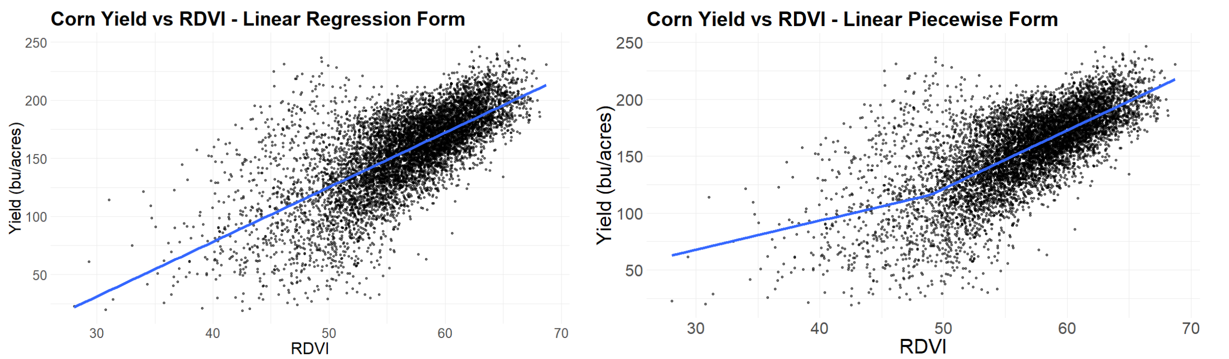
Results for corn shows that the top three vegetation indices that have relatively higher adjusted R^2 values and suitable intercept values relatively reasonable, and they are RDVI, GOSAVI, and GSAVI. In terms of functional forms, the piecewise regressions perform better compared to linear regression for corn. Similar functional form results were obtained in Chapter 2 using NDVI.

The Time Trend variable was statistically significant and had positive coefficients for all indices, indicating increasing corn yield over time. This is due mostly to improvement in crop management, technology, and inputs.

Based on Table 3.1, the linear piecewise functional form performs relatively better than the linear functional form. RDVI showed a similar adjusted R^2 value for the linear piecewise form. The top three indices with relatively higher adjusted R^2 value were GOSAVI (0.66), GSAVI (0.66), and RDVI (0.67) when averaging the adjusted R^2 values over the two forms. Most indices performed better than NDVI based on R^2 goodness of fit. However, TDVI, MSAVI2 and OSAVI had overall R^2 values slightly lower than NDVI.

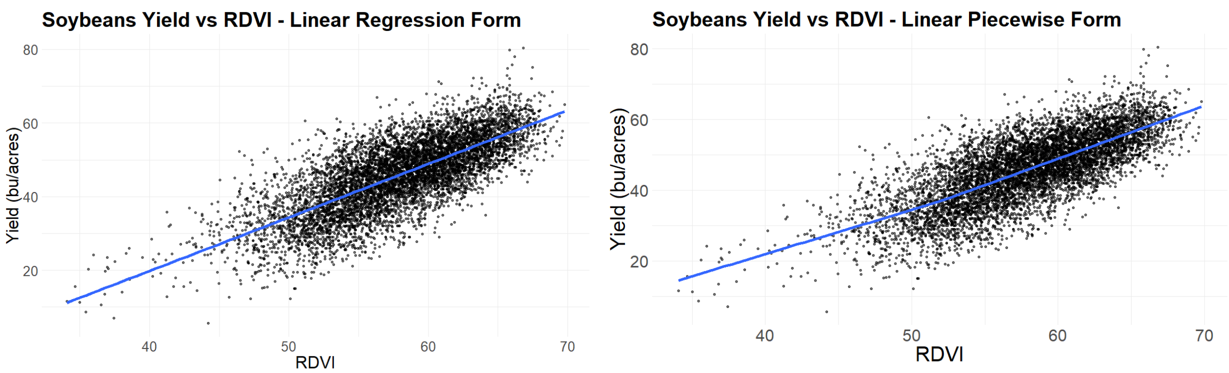
Figure 3.1 *All Four U.S. Crops: Plot of Crop Yield vs RDVI using Linear and Linear Piecewise Functions, 2008 to 2019*

Corn



(a) (b)

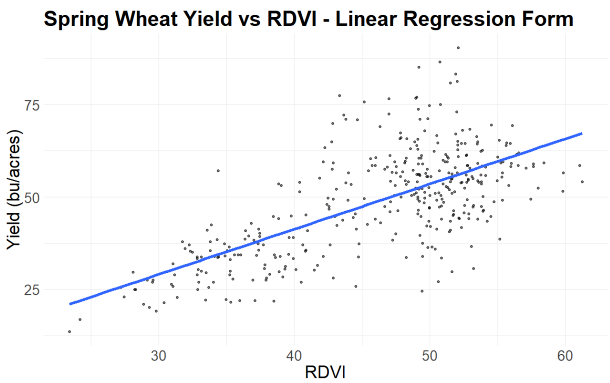
Soybeans



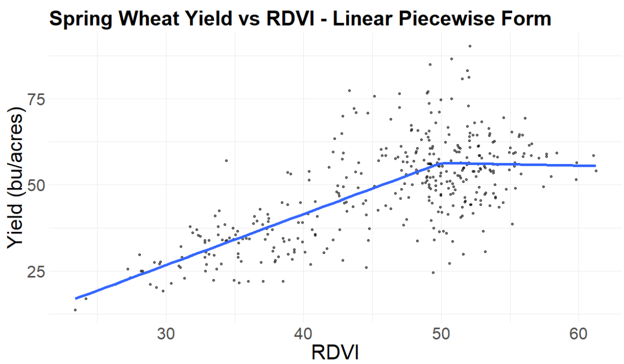
(c) (d)

Figure 3.1 (continued)

Spring Wheat

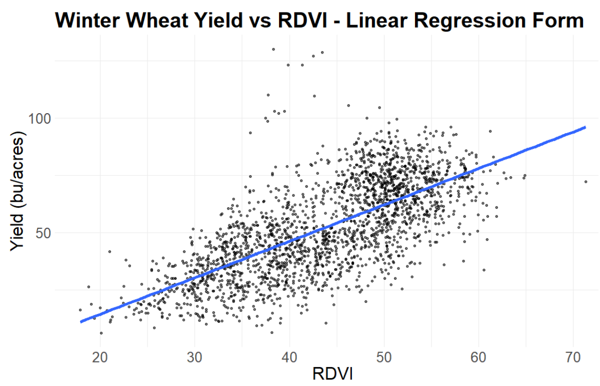


(e)

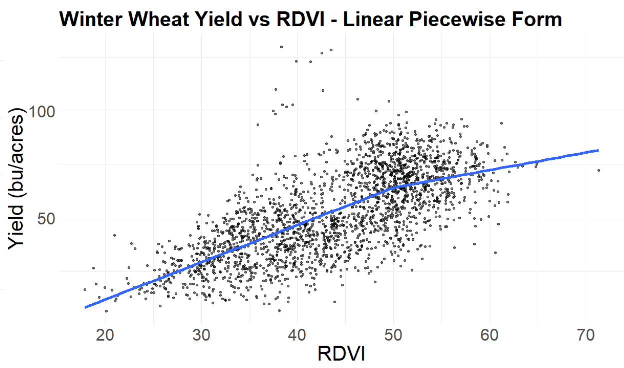


(f)

Winter Wheat



(g)



(h)

Notes: Plot (a) and (b) represents Linear Regression and Linear Piecewise Regression, respectively, using RDVI to estimate corn yield. Knot values (x(k)) used for Linear Piecewise: RDVI = 49. Plot (c) and (d) represents Linear Regression and Linear Piecewise Regression, respectively, using RDVI to estimate soybeans yield. Knot values (x(k)) used for Linear Piecewise: RDVI = 52. Plot (e) and (f) represents Linear Regression and Linear Piecewise Regression, respectively, using RDVI to estimate spring wheat yield. Knot values (x(k)) used for Linear Piecewise RDVI = 50. Plot (g) and (h) represents Linear Regression and Linear Piecewise Regression, respectively, using RDVI to estimate winter wheat yield. Knot values (x(k)) used for Linear Piecewise RDVI = 50.

Note: Corn: N=7812 (651 counties x 12 years), Soybeans: N=7524 (627 countries x 12 years), Spring Wheat: N=384 (32 counties x 12 years), and Winter Wheat: N=2196 (183 counties x 12 years).

Note: For corn, the linear regression may not be suitable to estimate corn yield due to nonlinearity at higher index value. But there is less saturation (nonlinearity) by using RDVI instead of NDVI. Soybeans have similar pattern as corn, linear piecewise able to address nonlinearity issue although RDVI able to lessen the saturation problem compared to other indices. Linear regression fit spring wheat and winter wheat quite well so linear piecewise may not be needed. Nonlinearity due to saturation does not seem to be a problem for spring wheat due to lesser biomass compared to corn and soybeans. Therefore, linear functional form may be suitable to estimate wheat.

Overall, there seems to be less saturation at higher index values for RDVI, GOSAVI, and GSAVI. Figure 3.1 shows visual description of RDVI using linear regression and linear piecewise functional form for all crops. Apart from having high adjusted R^2 value, the RDVI, GOSAVI, and GSAVI also have intercepts relatively suitable (consistent with theoretical considerations).

Table 3.1 *All Four U.S. Crops: Adjusted R^2 Values for All Vegetation Indices [Dependent Variable: Crop Yield; Independent Variables: Vegetation Index, U.S State Dummy Variable, Time Trend], 2008 to 2019*

Corn

Index	Form	Adj. R^2
NDVI	Linear	0.6294
	Linear Piecewise	0.6483
RDVI	Linear	0.6759
	Linear Piecewise	0.679
TDVI	Linear	0.6129
	Linear Piecewise	0.6431
WDRVI	Linear	0.6539
	Linear Piecewise	0.6582
EVI2	Linear	0.6421
	Linear Piecewise	0.6521
SAVI	Linear	0.6294
	Linear Piecewise	0.6482
GOSAVI	Linear	0.665
	Linear Piecewise	0.6687
GSAVI	Linear	0.665
	Linear Piecewise	0.6687
MSAVI2	Linear	0.6129
	Linear Piecewise	0.6431
OSAVI	Linear	0.6293
	Linear Piecewise	0.648

Soybeans

Index	Form	Adj. R^2
NDVI	Linear	0.6791
	Linear Piecewise	0.6813
RDVI	Linear	0.6942
	Linear Piecewise	0.6941
TDVI	Linear	0.6699
	Linear Piecewise	0.673
WDRVI	Linear	0.6923
	Linear Piecewise	0.6952
EVI2	Linear	0.686
	Linear Piecewise	0.6864
SAVI	Linear	0.6791
	Linear Piecewise	0.6808
GOSAVI	Linear	0.7001
	Linear Piecewise	0.7002
GSAVI	Linear	0.7001
	Linear Piecewise	0.6813
MSAVI2	Linear	0.6699
	Linear Piecewise	0.6727
OSAVI	Linear	0.6791
	Linear Piecewise	0.6813

Notes: Corn: Knot value (x(k)) used for Linear Piecewise: NDVI = 0.785, RDVI = 49, TDVI = 1.3, WDRVI = 0.3, EVI2 = 1.79, SAVI = 1.18, GOSAVI = 0.64, GSAVI = 0.96, MSAVI2 = 0.865, OSAVI = 0.785. Soybeans: Knot value (x(k)) used for Linear Piecewise: NDVI = 0.75, RDVI = 52, TDVI = 1.22, WDRVI = 0.5, EVI2 = 1.5, SAVI = 1.1, GOSAVI = 0.65, GSAVI = 1.1, MSAVI2 = 0.8, OSAVI = 0.75.

Table 3.1 (continued)

Spring Wheat

Index	Form	Adj. R^2
NDVI	Linear	0.5644
	Linear Piecewise	0.5633
RDVI	Linear	0.5646
	Linear Piecewise	0.5985
TDVI	Linear	0.5609
	Linear Piecewise	0.571
WDRVI	Linear	0.5537
	Linear Piecewise	0.5793
EVI2	Linear	0.5621
	Linear Piecewise	0.5682
SAVI	Linear	0.5644
	Linear Piecewise	0.5671
GOSAVI	Linear	0.5945
	Linear Piecewise	0.5983
GSAVI	Linear	0.5945
	Linear Piecewise	0.5983
MSAVI2	Linear	0.5609
	Linear Piecewise	0.571
OSAVI	Linear	0.5644
	Linear Piecewise	0.5708

Winter Wheat

Index	Form	Adj. R^2
NDVI	Linear	0.7375
	Linear Piecewise	0.7444
RDVI	Linear	0.738
	Linear Piecewise	0.7424
TDVI	Linear	0.738
	Linear Piecewise	0.7386
WDRVI	Linear	0.7351
	Linear Piecewise	0.7368
EVI2	Linear	0.7365
	Linear Piecewise	0.7375
SAVI	Linear	0.7375
	Linear Piecewise	0.7444
GOSAVI	Linear	0.7379
	Linear Piecewise	0.7451
GSAVI	Linear	0.7379
	Linear Piecewise	0.7451
MSAVI2	Linear	0.738
	Linear Piecewise	0.7383
OSAVI	Linear	0.7375
	Linear Piecewise	0.7444

Notes: Spring Wheat: Notes: Knot value (x(k)) used for Linear Piecewise: NDVI = 0.55, RDVI = 50, TDVI = 1.05, WDRVI = 0.2, EVI2 = 1.7, SAVI = 0.9, GOSAVI = 0.6, GSAVI = 0.9, MSAVI2 = 0.7, OSAVI = 0.75. Winter Wheat: Notes: Knot value (x(k)) used for Linear Piecewise: NDVI = 0.6, RDVI = 50, TDVI = 1, WDRVI = 0.25, EVI2 = 1.7, SAVI = 0.9, GOSAVI = 0.6, GSAVI = 0.9, MSAVI2 = 0.65, OSAVI = 0.6.

Overall, RDVI, GOSAVI, and GSAVI are the top three indices with relatively higher adjusted R^2 for all crops. Linear piecewise may be the better option instead with suitable balance between adjusted R^2 value and intercept relatively suitable, although visually linear regression is sufficient for spring wheat and winter wheat.

3.5.2 Soybeans: Various Vegetation Indices Results

Results for soybeans analysis follow very similarly to corn. The top three vegetation indices that produce relatively higher adjusted R^2 value and suitable intercept are RDVI, GOSAVI, and GSAVI. In general, the piecewise regressions functional forms appear more suitable than linear regression for soybeans, although for RDVI, linear regression may provide a suitable fit.

The presence of Time Trend is significant and has positive coefficients for all indices, indicating increase in soybean yield over time due to advancement in technology, management, and inputs. Many U.S state dummy variable are statistically significant, likely due to difference across states such as weather, soil type, and geographic location.

Soybeans follow similar pattern as corn in general. However, based on Table 3.1, by using RDVI to estimate soybeans yield, linear regression may be a suitable form. The top three indices with relatively higher adjusted R^2 value are GOSAVI (0.7), GSAVI (0.69), and RDVI (0.69), averaged over the two functional forms. Most indices being considered perform better than NDVI, except for TDVI and MSAVI2 which overall R^2 values are slightly lower than NDVI.

Overall, there seem to be less saturation at higher index values for RDVI, GOSAVI, and GSAVI. These indices have higher adjusted R^2 value and suitable intercept.

3.5.3 Spring Wheat: Various Vegetation Indices Results

Results for spring wheat show that the RDVI index provide the best balance of having higher adjusted R^2 as well as suitable intercept. The linear regression form provides a reasonably goodness of fit.

According to the adjusted R^2 value on Table 3.1, the top three indices with relatively higher adjusted R^2 value are GOSAVI (0.59), GSAVI (0.59), and RDVI (0.58), averaged across the two forms. Most index being considered perform quite similarly as NDVI with about 57% of variability in spring wheat yield explained by the respective index, except for WDRVI with overall R^2 values are slightly lower than NDVI. The linear piecewise may be less beneficial for some indices given that it has an insignificant coefficient (p-value>0.05), for indices such as SAVI.

Visually, there is very slight nonlinearity observed in spring wheat, which may be due to wheat having lower biomass compared to corn and soybeans. Based on visual inspection, piecewise regressions may not be necessary for spring wheat, using most of indices being considered. Instead, linear regression provides a suitable fit for all indices compared to linear piecewise. Figure 3.1 shows visual description of RDVI using linear regression and linear piecewise functional form for all crops. Even though linear regression provides a good fit, using linear piecewise may help to attain suitable intercept values, which is favorable.

Overall, even though GOSAVI and GSAVI have adjusted R^2 values similar to RDVI, their intercepts are quite far off. Therefore, based on these results, RDVI may provide a suitable balance between adjusted R^2 value, intercept, and visual inspection for spring wheat.

3.5.4 Winter Wheat: Various Vegetation Indices Results

Similar to spring wheat results, the winter wheat crop analysis shows that RDVI provides a reasonable balance of having higher adjusted R^2 as well as suitable intercept. In general, linear regression provides a suitable fit for winter wheat, although for some indices, using linear piecewise allows intercepts to be relatively suitable.

According to the adjusted R^2 value on Table 3.1, the top three indices with relatively higher adjusted R^2 value are GOSAVI (0.74), GSAVI (0.74), and RDVI (0.74), averaged over the two forms. Most indices being considered perform quite similarly to NDVI with about 73% of variability in winter wheat yield explained by the respective index, except for MSAVI2 which overall adjusted R^2 values are slightly lower than NDVI. For some indices, linear piecewise functional form had insignificant coefficient ($p\text{-value}>0.05$), such as for MSAVI2.

There appears to be some nonlinearity present for all vegetation indices being considered here, although not as much for winter wheat compared to corn and soybeans. Based on visual inspection, piecewise regressions may not be needed for winter wheat using most of indices being considered. Linear regression may be sufficient compared to linear piecewise regression. Figure 3.1 shows visual description of RDVI using linear regression and linear piecewise functional form for all crops. Similar to spring wheat, even though linear regression provides a good fit, using linear piecewise may help to attain suitable intercept values, which is favorable. GOSAVI and GSAVI have adjusted R^2 values similar to RDVI, their intercepts are less acceptable. Therefore, based on these results, RDVI is able to provide a suitable balance between adjusted R^2 value, intercept, and visual for spring wheat.

3.5.4 Results Discussion

3.5.4.1 Regression Functional Forms

Based on the results, the linear piecewise fit well for corn and soybeans (for most indices) among the two forms being considered which are linear regression and linear piecewise. Using piecewise functional forms gives a better fit visually, but also intercepts relatively suitable, and higher adjusted R^2 value.

For spring wheat and winter wheat, the use of piecewise regressions is not as beneficial because linear regression form seems to fit quite well visually for these two crops. Index saturation does not seem to be much of a problem for both wheats compared to corn and soybeans which could be due to the characteristic of wheat that does not grow as dense as corn and soybeans at maturity. The only downside of using just linear regression is the inability to have the intercept value relatively reasonable. For some indices, piecewise regressions may not be needed unless specific knot value is assigned, although using linear piecewise can provide relatively higher adjusted R^2 value and intercept that is relatively suitable compared to linear regression.

3.5.4.2 Potential of Vegetation Indices that Use Green Reflectance Band

Although overall the performance of vegetation indices are quite similar to NDVI based on adjusted R^2 value, the top three that stand out are RDVI, GOSAVI, and GSAVI. GOSAVI and GSAVI use the green band in place of red for a vegetation index and overall, the green band helps to increase the sensitivity of the index as vegetation gets denser (compared to NDVI or other indices using the red band), resulting in less saturation. The OSAVI and SAVI have the exact same equations as GOSAVI and GSAVI but they use red spectral band instead of the green band. The results showed that both OSAVI and SAVI were not any better than NDVI (i.e similar adjusted R^2 values), yet GOSAVI and GSAVI are. Therefore, this higher fitness of GOSAVI and GSAVI compared to NDVI is entirely due to the use of green band.

3.5.4.3 Potential of Using RDVI in Estimating Crop Yield

RDVI performed as well as GOSAVI and GSAVI while still using the red band. RDVI provides balance between low and high vegetation, and may be suitable for agricultural land. Based on this study, RDVI may have less saturation than NDVI, therefore, RDVI can provide more accurate estimation than NDVI when using linear regression functional form, or when the green band is not available to use.

3.6 SUMMARY

The objective of this study was to compare various satellite-based vegetation indices for estimating crop yield, and two functional forms were used. Crop yield was the dependent variable and vegetation index (e.g SAVI) was one of the independent variables, along with time trend and dummy variables for U.S states. There are many advantages of using satellite based remote sensing to estimate crop yield compared to the traditional crop surveys. Such advantages include lower cost, faster in obtaining data and updating fairly frequently. Data from remote sensing is also relatively accurate.

NDVI is the most commonly used vegetation index for agricultural or ecology purposes. However, there are a few limitations of NDVI such as saturation with high density biomass, such as high crop yield or very green crops nearing maturity. Therefore, this paper discusses nine other vegetation indices to understand their performance compared to NDVI for estimating crop yield. The data and methodology of this chapter is the same as in Chapter 2.

There is benefit of using linear piecewise functional form for corn but less for soybeans. For spring wheat and winter wheat, although linear piecewise form produced a higher adjusted R^2 value, both wheat data appear more linear (less saturation) compared to corn and soybeans, leading to using piecewise regression being unnecessary. Linear regression fit well in estimating spring wheat and winter wheat yield.

For corn, soybeans, spring wheat, and winter wheat, the top three indices that provide highest adjusted R^2 value are GOSAVI, GSAVI, and RDVI. Interestingly, OSAVI and SAVI indices, have same equation as GOSAVI and GSAVI except they used red band instead of green band. However, they are not performing any better than NDVI for estimating crop yield, whereas GOSAVI and GSAVI are. It can be seen quite clearly that what made GOSAVI and GSAVI perform better than NDVI is the presence of green band, which seems overall more robust than red band. RDVI was derived to provide a balance between environments with dense biomass and environments that have sparse vegetation and may be suitable for agricultural land.

For future research of this type, researchers may wish to provide more focus on the green reflectance band rather than only red reflectance band. The green band is as widely available through MODIS as red band, unlike the blue band. This paper only discussed two green band based indices, GOSAVI and GSAVI, but further research could compare them with other green band indices such as Green Difference Vegetation Index (GDVI), Green Normalized Difference Vegetation Index (GNDVI), Green Chlorophyll Index (GCI), or many others.

CHAPTER 4

THESIS SUMMARY

Traditionally, crop yield estimation has been done mainly through crop surveys, and this process can be labor intensive and time consuming. This can lead to slower yield estimation because the data takes considerable time to collect. As technology has advanced, satellite-based remote sensing has become more popular for monitoring and observing vegetation growth and estimation of crop yield. There are a number of advantages regarding remote sensing for crop yield estimation such as lower cost, faster in obtaining and updating data, and data from remote sensing is also relatively accurate. Two studies are undertaken in this research.

The first study investigates eight regression functional forms for estimating crop yield using satellite based NDVI. Crop yield is used as the dependent variable and NDVI as the main independent variable (along with a time trend and dummy variables for U.S. States). The analysis includes eight regression functional forms from four groups: linear regression, polynomial regression (quadratic and cubic), piecewise and spline regression (linear piecewise, quadratic piecewise, cubic spline, and natural spline), and Generalized Additive Model (GAM). Crop yield data is from USDA, for corn, soybeans, spring wheat, and winter wheat and MODIS satellite data is used. Results for the first study show that by using piecewise regression, the knot/breakpoint values can be set to create a reasonable intercept, sufficient R^2 value, and reasonable visual fit at a high vegetation index level, where fit can be more challenging.

The second study investigates 10 vegetation indices for estimating crop yield using two functional forms. Crop yield is the dependent variable and vegetation index (e.g SAVI) is the main independent variable (along with a time trend and dummy variables for U.S. States). Results show the benefit of using piecewise regression compared to linear regression, particularly for corn. Results also suggest that GOSAVI, GSAVI, and RDVI perform somewhat better than other indices, particularly for corn.

Future research could further explore GOSAVI and GSAVI. They use the green band instead of the red band, and so are somewhat different than the traditional NDVI, that uses the red band. RDVI was able to perform similarly to indices that use the green band, even though it uses the red band. RDVI is designed to provide balance between an environment with dense biomass and sparse vegetation, and maybe suitable for agricultural land. RDVI also has a relatively simple equation and relatively favorable performance.

REFERENCES

Abdolalizadeh, Z., Ghorbani, A., Mostafazadeh, R., Moameri, M., 2020. Rangeland canopy cover estimation using Landsat OLI data and vegetation indices in Sabalan rangelands, Iran. Arabian Journal of Geosciences 13 (6), 245. https://doi.org/ 10.1007/s12517-020-5150-1.

Abrantes, T. C., Queiroz, A. R. S., Lucio, F. R., Mendes Júnior, C. W., Kuplich, T. M., Bredemeier, C., & Merotto Júnior, A. (2021). Assessing the effects of dicamba and 2,4 Dichlorophenoxyacetic acid (2,4D) on soybean through vegetation indices derived from Unmanned Aerial Vehicle (UAV) based RGB imagery. International Journal of Remote Sensing, 42(7), 2740–2758. https://doi.org/10.1080/01431161.2020.1832283

Al Shidi, R. H., Kumar, L., & Al-Khatri, S. A. (2019). Detecting Dubas bug infestations using high resolution multispectral satellite data in Oman. Computers and Electronics in Agriculture, 157, 1–11. https://doi.org/10.1016/j.compag.2018.12.037

Bannari, A., Asalhi, H., & Teillet, P. (2002). Transformed difference vegetation index (TDVI) for vegetation cover mapping. IEEE International Geoscience and Remote Sensing Symposium, 5, 3053–3055 vol.5. https://doi.org/10.1109/IGARSS.2002.1026867

Barillé, L., Mouget, J.-L., Méléder, V., Rosa, P., & Jesus, B. (2011). Spectral response of benthic diatoms with different sediment backgrounds. Remote Sensing of Environment, 115(4), 1034–1042. https://doi.org/10.1016/j.rse.2010.12.008

Barzin, R., Pathak, R., Lotfi, H., Varco, J., & Bora, G. C. (2020). Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn. Remote Sensing (Basel, Switzerland), 12(15), 2392–. https://doi.org/10.3390/rs12152392

Bateman, N. R., Catchot, A. L., Gore, J., Cook, D. R., Musser, F. R., & Irby, J. T. (2020). Effects of Planting Date for Soybean Growth, Development, and Yield in the Southern USA. Agronomy (Basel), 10(4), 596–. https://doi.org/10.3390/agronomy10040596

Baugh, W. M., & Groeneveld, D. P. (2006). Broadband vegetation index performance evaluated for a low-cover environment. International Journal of Remote Sensing, 27(21), 4715–4730. https://doi.org/10.1080/01431160600758543

Becker-Reshef, I., Vermote, E., Lindeman, M., & Justice, C. (2010). A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data.

Remote Sensing of Environment, 114(6), 1312–1323. https://doi.org/10.1016/j.rse.2010.01.010

Benedetti, R., Rossini, P. (1993). On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Rogna. Remote Sensing of Environment. 45, 311–326.

Bera, D., Chatterjee, N. D., & Bera, S. (2021). Comparative performance of linear regression, polynomial regression and generalized additive model for canopy cover estimation in the dry deciduous forest of West Bengal. Remote Sensing Applications, 22, 100502–. https://doi.org/10.1016/j.rsase.2021.100502

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology, 173, 74–84. https://doi.org/10.1016/j.agrformet.2013.01.007

Cao, Q., Miao, Y., Feng, G., Gao, X., Li, F., Liu, B., Yue, S., Cheng, S., Ustin, S. L., & Khosla, R. (2015). Active canopy sensing of winter wheat nitrogen status: An evaluation of two sensor systems. Computers and Electronics in Agriculture, 112, 54–67. https://doi.org/10.1016/j.compag.2014.08.012

Carreiras, J. M. ., Pereira, J. M. ., & Pereira, J. S. (2006). Estimation of tree canopy cover in evergreen oak woodlands using remote sensing. Forest Ecology and Management, 223(1), 45–53. https://doi.org/10.1016/j.foreco.2005.10.056

Celleri, C., Georgina Zapperi, Gabriela González Trilla & Paula Pratolongo. (2019). Assessing the capability of broadband indices derived from Landsat 8 Operational Land Imager to monitor above ground biomass and salinity in semiarid saline environments of the Bahía Blanca Estuary, Argentina. International Journal of Remote Sensing, 40:12, 4817-4838, DOI: 10.1080/01431161.2019.1574992

Chandel, A. K., Khot, L. R., & Yu, L.-X. (2021). Alfalfa (Medicago sativa L.) crop vigor and yield characterization using high-resolution aerial multispectral and thermal infrared imaging technique. Computers and Electronics in Agriculture, 182, 105999–. https://doi.org/10.1016/j.compag.2021.105999

Chen, K., O'Leary, R. A., & Evans, F. H. (2019). A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. Agricultural Systems, 173, 140–150. https://doi.org/10.1016/j.agsy.2019.02.009

Choudhury, A. H., Jones, J. R., Choudhury, R. "Lena", & Spaulding, A. D. (2015). Association of rainfall and detrended crop yield based on piecewise regression for agricultural insurance. Journal of Economics and Economic Education Research, 16(2), 31–.

Crane-Droesch, A., Abiven, S., Jeffery, S., & Torn, M. S. (2013). Heterogeneous global crop yield response to biochar: a meta-regression analysis. Environmental Research Letters, 8(4), 44049–. https://doi.org/10.1088/1748-9326/8/4/044049

Cummings, C., Miao, Y., Paiao, G. D., Kang, S., & Fernández, F. G. (2021). Corn Nitrogen Status Diagnosis with an Innovative Multi-Parameter Crop Circle Phenom Sensing System. Remote Sensing (Basel, Switzerland), 13(401), 401–. https://doi.org/10.3390/rs13030401

Dadhwal, V. K. & Sridhar, V. N. (1997). A non-linear regression form for vegetation index-crop yield relation incorporating acquisition date normalization, International Journal of Remote Sensing, 18:6, 1403-1408, DOI: 10.1080/014311697218520

da Silva, V. S., Salami, G., da Silva, M. I. O., Silva, E. A., Monteiro Junior, J. J., & Alba, E. (2020). Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. Geology, Ecology, and Landscapes, 4(2), 159–169. https://doi.org/10.1080/24749508.2019.1608409

de Souza, C. H. W., Mercante, E., Johann, J. A., Lamparelli, R. A. C., & Uribe-Opazo, M. A. (2015). Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. International Journal of Remote Sensing, 36(7), 1809–1824. https://doi.org/10.1080/01431161.2015.1026956

Dempewolf, J.; Adusei, B.; Becker-Reshef, I.; Barker, B.; Potapov, P.; Hansen, M.; Justice, C. Wheat production forecasting for Pakistan from satellite data. In Proceedings of 2013 IEEE Internationa Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, VIC, Australia, 21–26 July 2013; pp. 3239–3242.

Diego Bedin Marin, Gabriel Araújo e Silva Ferraz, Paulo Henrique Sales Guimarães, Felipe Schwerz, Lucas Santos Santana, Brenon Dienevam Souza Barbosa, Rafael Alexandre Pena Barata, Rafael de Oliveira Faria, Jessica Ellen Lima Dias, Leonardo Conti, & Giuseppe Rossi.

(2021). Remotely Piloted Aircraft and Random Forest in the Evaluation of the Spatial Variability of Foliar Nitrogen in Coffee Crop. Remote Sensing (Basel, Switzerland), 13(1471), 1471–. https://doi.org/10.3390/rs13081471

Dong, T., Meng, J., Shang, J., Liu, J., Wu, B., & Huffman, T. (2015). Modified vegetation indices for estimating crop fraction of absorbed photosynthetically active radiation. International Journal of Remote Sensing, 36(12), 3097–3113. https://doi.org/10.1080/01431161.2015.1042122

Gitelson, A. A. (2013). Remote estimation of crop fractional vegetation cover: the use of noise equivalent as an indicator of performance of vegetation indices. International Journal of Remote Sensing, 34(17), 6054–6066. https://doi.org/10.1080/01431161.2013.793868

González-Dugo, M. P., & Mateos, L. (2008). Spectral vegetation indices for benchmarking water productivity of irrigated cotton and sugarbeet crops. Agricultural Water Management, 95(1), 48–58

GopalaPillai, S.; Tian, L. (1999). In-field variability detection and yield prediction in corn using digital aerial imaging. Transactions-American Society of Agricultural Engineers, 42, 1911-1920.

Groten, S.M.E. (1993). NDVI-Crop monitoring and early yield assessment of Burkina Faso. International Journal of Remote Sensing. 14, 1495–1515.

Guindin-Garcia, N., Gitelson, A. A., Arkebauer, T. J., Shanahan, J., & Weiss, A. (2012). An evaluation of MODIS 8- and 16-day composite products for monitoring maize green leaf area index. Agricultural and Forest Meteorology, 161, 15–25. https://doi.org/10.1016/j.agrformet.2012.03.012

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sensing of Environment, 90(3), 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Roujean, & Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. Remote Sensing of Environment, 51(3), 375–384. https://doi.org/10.1016/0034-4257(94)00114-3

Jamali, S., Seaquist, J., Eklundh, L., & Ardö, J. (2014). Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel. Remote Sensing of Environment, 141, 79–89. https://doi.org/10.1016/j.rse.2013.10.019

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. Remote Sensing (Basel, Switzerland), 11(23), 2873–. https://doi.org/10.3390/rs11232873

Khorram, S., Koch, F. H., van der Wiele, C. F., & Nelson, S. A. C. (2012). Remote Sensing (1st ed. 2012.). Springer New York. https://doi.org/10.1007/978-1-4614-3103-9

Labus, M. P., Nielsen, G. A., Lawrence, R. L., Engel, R., & Long, D. S. (2002). Wheat yield estimates using multi-temporal NDVI satellite imagery. International Journal of Remote Sensing, 23(20), 4169–4180. https://doi.org/10.1080/01431160110107653

Liaqata, M. U., Cheemaa, M. J. M., Huangc, W., Mahmoodd, T., Zamana, M., & Khan, M. M. (2017). Evaluation of MODIS and Landsat multiband vegetation indices used for wheat yield estimation in irrigated Indus Basin. Computers and Electronics in Agriculture, 138, 39–47

Manjunath, K. R., Potdar, M. B., & Purohit, N. L. (2002). Large area operational wheat yield model development and validation based on spectral and meteorological data. International Journal of Remote Sensing, 23(15), 3023–3038. https://doi.org/10.1080/01431160110104692

Marcillo, G. S., Martin, N. F., Diers, B. W., Da Fonseca Santos, M., Leles, E. P., Chigeza, G., & Francischini, J. H. (2021). Implementation of a Generalized Additive Model (GAM) for Soybean Maturity Prediction in African Environments. Agronomy (Basel), 11(6), 1043–. https://doi.org/10.3390/agronomy11061043

Mkhabela, M. ., Bullock, P., Raj, S., Wang, S., & Yang, Y. (2011). Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agricultural and Forest Meteorology, 151(3), 385–393. https://doi.org/10.1016/j.agrformet.2010.11.012

Mkhabela, M. S., Mkhabela, M. S., & Mashinini, N. N. (2005). Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. Agricultural and Forest Meteorology, 129(1), 1–9. https://doi.org/10.1016/j.agrformet.2004.12.006

Melin, M., Korhonen, L., Kukkonen, M., & Packalen, P. (2017). Assessing the performance of aerial image point cloud and spectral metrics in predicting boreal forest canopy cover. ISPRS

Journal of Photogrammetry and Remote Sensing, 129, 77–85. https://doi.org/10.1016/j.isprsjprs.2017.04.018

Morier, T., Cambouris, A. N., & Chokmani, K. (2015). In-Season Nitrogen Status Assessment and Yield Estimation Using Hyperspectral Vegetation Indices in a Potato Crop. Agronomy Journal, 107(4), 1295–1309. https://doi.org/10.2134/agronj14.0402

Mróz, M., & Sobieraj, A. (2004). Comparison of several vegetation indices calculated on the basis of a seasonal SPOT XS time series, and their suitability for land cover and agricultural crop identification. Technical Sciences, 07, 39-66.

Myneni, R., & Asrar, G. (1994). Atmospheric effects and spectral vegetation indices. Remote Sensing of Environment, 47(3), 390–402. https://doi.org/10.1016/0034-4257(94)90106-6

Panda, S. S., Ames, D. P., & Panigrahi, S. (2010). Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. Remote Sensing (Basel, Switzerland), 2(3), 673–696. https://doi.org/10.3390/rs2030673

Payero, J. O., Neale, C. M. U., Wright, J. L. (2004). COMPARISON OF ELEVEN VEGETATION INDICES FOR ESTIMATING PLANT HEIGHT OF ALFALFA AND GRASS. Applied Engineering in Agriculture, 20(3), 385–393. https://doi.org/10.13031/2013.16057

Pettorelli, N. (2013). The Normalized Difference Vegetation Index. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199693160.001.0001

Piegari, E., Gossn, J. I., Grings, F., Barraza Bernadas, V., Juárez, Á. B., Mateos-Naranjo, E., & González Trilla, G. (2021). Estimation of leaf area index and leaf chlorophyll content in Sporobolus densiflorus using hyperspectral measurements and PROSAIL model simulations. International Journal of Remote Sensing, 42(4), 1181–1200. https://doi.org/10.1080/01431161.2020.1826058

Prasad, A. K., Singh, R. P., Tare, V., & Kafatos, M. (2007). Use of vegetation index and meteorological parameters for the prediction of crop yield in India. International Journal of Remote Sensing, 28(23), 5207–5235. https://doi.org/10.1080/01431160601105843

Purevdorj, T., Tateishi, R., Ishiyama, T., & Honda, Y. (1998). Relationships between percent vegetation cover and vegetation indices. International Journal of Remote Sensing, 19(18), 3519–3535. https://doi.org/10.1080/014311698213795

Quarmby, N.A., Milnes, M., Hindle, T.L., Silleos, N. (1993). The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. International Journal of Remote Sensing. 14, 199–210

Qi, J., Chehbouni, A., Huete, A. ., Kerr, Y. ., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. Remote Sensing of Environment, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Rasmusen, M.S. (1992). Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from AVHRR. International Journal of Remote Sensing. 13, 3431–3442

Ren, J.; Chen, Z.; Zhou, Q.; Tang, H. (2008). Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. International Journal of Applied Earth Observation and Geoinformation, 10, 403–413

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. Remote Sensing of Environment, 55(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Roznik, M., (2021). Application of weather data, satellite data, and other geospatial data for improving crop insurance and agricultural risk management. PhD dissertation, University of Manitoba.

Sakamoto, T., Gitelson, A. A., & Arkebauer, T. J. (2013). MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sensing of Environment, 131, 215–231. https://doi.org/10.1016/j.rse.2012.12.017

Shang, J., Liu, J., Ma, B., Zhao, T., Jiao, X., Geng, X., Huffman, T., Kovacs, J. M., & Walters, D. (2015). Mapping spatial variability of crop growth conditions using RapidEye data in Northern Ontario, Canada. Remote Sensing of Environment, 168, 113–125. https://doi.org/10.1016/j.rse.2015.06.024

Shi, H., & Xingguo, M. (2011). Interpreting spatial heterogeneity of crop yield with a process model and remote sensing. Ecological Modelling, 222(14), 2530–2541. https://doi.org/10.1016/j.ecolmodel.2010.11.011

Sibley, A. M., Grassini, P., Thomas, N. E., Cassman, K. G., & Lobell, D. B. (2014). Testing Remote Sensing Approaches for Assessing Yield Variability among Maize Fields. Agronomy Journal, 106(1), 24–32. https://doi.org/10.2134/agronj2013.0314

Sripada, R. P., Heiniger, R. W., White, J. G., & Meijer, A. D. (2006). Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. Agronomy Journal, 98(4), 968–977. https://doi.org/10.2134/agronj2005.0200

Sripada, R. P., Farrer, D. C., Weisz, R., Heiniger, R. W., & White, J. G. (2007). Aerial Color Infrared Photography to Optimize In-Season Nitrogen Fertilizer Recommendations in Winter Wheat. Agronomy Journal, 99(6), 1424–1435. https://doi.org/10.2134/agronj2006.0258

Stepanov, A., Dubrovin, K., Sorokin, A., & Aseeva, T. (2020). Predicting Soybean Yield at the Regional Scale Using Remote Sensing and Climatic Data. Remote Sensing (Basel, Switzerland), 12(12), 1936–. https://doi.org/10.3390/rs12121936

Stroppiana, D., Villa, P., Sona, G., Ronchetti, G., Candiani, G., Pepe, M., Busetto, L., Migliazzi, M., & Boschetti, M. (2018). Early season weed mapping in rice crops using multi-spectral UAV data. International Journal of Remote Sensing, 39(15-16), 5432–5452. https://doi.org/10.1080/01431161.2018.1441569

Suarez, L. A., Robson, A., McPhee, J., O'Halloran, J., & van Sprang, C. (2020). Accuracy of carrot yield forecasting using proximal hyperspectral and satellite multispectral data. Precision Agriculture. https://doi.org/10.1007/s11119-020-09722-6

Sun, H., Xu, A., Lin, H., Zhang, L., & Mei, Y. (2012). Winter wheat mapping using temporal signatures of MODIS vegetation index data. International Journal of Remote Sensing, 33(16), 5026–5042. https://doi.org/10.1080/01431161.2012.657366

Towers, P. C., Strever, A., & Poblete-Echeverría, C. (2019). Comparison of Vegetation Indices for Leaf Area Index Estimation in Vertical Shoot Positioned Vine Canopies with and without Grenbiule Hail-Protection Netting. Remote Sensing (Basel, Switzerland), 11(9), 1073–. https://doi.org/10.3390/rs11091073

Vescovo, L., Wohlfahrt, G., Balzarolo, M., Pilloni, S., Sottocornola, M., Rodeghiero, M., & Gianelle, D. (2012). New spectral vegetation indices based on the near-infrared shoulder wavelengths for remote detection of grassland phytomass. International Journal of Remote Sensing, 33(7), 2178–2195. https://doi.org/10.1080/01431161.2011.607195

Viña, A., Henebry, G. M., & Gitelson, A. A. (2004). Satellite monitoring of vegetation dynamics: Sensitivity enhancement by the wide dynamic range vegetation index. Geophysical Research Letters, 31(4), L04503–n/a. https://doi.org/10.1029/2003GL019034

Wang, D., Chen, Y., Wang, M., Quan, J., & Jiang, T. (2016). A New Neighboring Pixels Method for Reducing Aerosol Effects on the NDVI Images. Remote Sensing (Basel, Switzerland), 8(6), 489–. https://doi.org/10.3390/rs8060489

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48(4), 817–838. https://doi.org/10.2307/1912934

Wilton, M. (2021). Crop Yield Estimation Using NDVI: A Comparison of Various NDVI Metrics.

Wood, S. N. (2006). Generalized additive models: an introduction with R . Chapman & Hall/CRC.

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. Journal of Sensors, 2017, 1–17. https://doi.org/10.1155/2017/1353691

Zaigham Abbas Naqvi, S. M., Awais, M., Khan, F. S., Afzal, U., Naz, N., & Khan, M. I. (2021). Unmanned air vehicle based high resolution imagery for chlorophyll estimation using spectrally modified vegetation indices in vertical hierarchy of citrus grove. Remote Sensing Applications, 23. https://doi.org/10.1016/j.rsase.2021.100596

Zhang, X., & Zhang, Q. (2016). Monitoring interannual variation in global crop yield using long-term AVHRR and MODIS observations. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 191–205. https://doi.org/10.1016/j.isprsjprs.2016.02.010