International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Exploring Vegetative Indices for Yield Prediction using Sentinel 2 Data – A Study in a Select Region of Karnataka

Geetha M*[1], Asha Gowda Karegowda[2], Nandeesha R[3], Nagaraj B V[4]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract: India is an agrarian economy and largest share of population depend on agriculture. Though there are mechanisms to approximately estimate crop yield by means of controlled experiments or past data, the reliability is limited. As a matter of fact, the crop yield is based on estimates that may suffer from multiple bias. In recent years, remote sensing images augmented with machine learning and deep learning techniques help us get the efficient crop yield statistics based on the crop on the field which help policymakers in devising better policies and governance. Remote sensing images of crops under study when subjected to machine learning techniques, one can classify the images into homogenous crop classes and record crop health / growth indicators such as Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI) and Soil Adjusted Vegetative Index (SAVI) based on the training dataset and image quality, which further leads to crop yield estimates with desired level of accuracy. The present study investigates the relationship between yield influencing parameters such as physical variables, soil, weather characteristics and vegetation indices. Using correlation and multiple regression analysis, most efficient parameters that best estimate the crop yield is determined using the satellite data obtained for a study region in central part of the state of Karnataka. It was found that one of the vegetation indices, Soil Adjusted Vegetative Index values can predict near to accurate crop yield values by the end of 81 days after transplantation of paddy where as NDVI and NDVI can give yield estimated only after 116 days of transplantation. Thus crop yield estimates using SAVI works better in terms of predicting paddy yield at least one month before (after 81 days of transplantation) actual harvesting i.e., after 120 days as practiced by farmers in the study area.

Keywords: *Crop Yield, NDVI, NDWI, SAVI, Correlation and Regression Analysis*

1. Introduction

In India, huge farming community, small land holding, disproportionate economic stability of farmers, varied patterns of farming and impulsive marketing practices of agricultural produce, render the govt./non-govt. policy making bodies vulnerable to fetch accurate yield data. Henceforth yield estimates rely on the agriculture department's data, which are not based on actual crop output from fields and rather depend on sample survey or past data. Thus there is a despairing need for accurate crop yield data.

2. Literature Review

Vegetation indices play an important role in predicting the crop yield. According to the studies, in cloud free conditions, satellite data provides growth and status of the crop on weekly basis by allowing researchers to draw inferences on parameters like biomass, plant density and yield(Campos et al., 2019; Punalekar et al., 2018, Clevers et al., 2017; Pasqualotto et al.,2019 and Battude et. al., 2016; Hunt et al., 2019).

One more study finds that the derivation of yield from satellite, soil and relief data is not insignificant because the yield of each plant depends on compound set of factors and it is different for every type of crop (Geisler 1988).

The actual yield of the field is dependent on the number of seeds, soil type and therefore soil fertility, water and nutrient supply, and duration of sunshine throughout the season (Evans& Fischer, 1999; Geisler, 1988)

It was found that the grain yield of cereals, for example, cannot be measured explicitly from satellite data, but by the methods which are based on biomass, chlorophyll content (Babar et al., 2006; Ren et al., 2008, Guo et al., 2018; Serrano et al., 2000).

The Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI) or chlorophyll content, which can be related to yield are commonly used vegetation indices for yield estimation models (Bognár et al., 2017; Marti et al., 2007, Barnes et al., 2000; Viña et al., 2011).

A meaningful correlation between crop yield and the vegetation index is not only a question of the suitable vegetation index, but also of the time of satellite image acquisition. It is not congruent in every phenological stage. For wheat, the phenological growth stages of stem elongation, heading and development of fruit until early ripening are stated to be suitable to derive spatial yield patterns from satellite data (Hack et al., 1992, Knoblauch et al., 2017; Marti et al., 2007).

The present study is intended to estimate crop yield of paddy using some of the well-known vegetative indices. The study also attempts to identify the most appropriate vegetative index among the chosen ones. The study involves spatial observation of 40 paddy plots across all growth phases of the crop starting from transplantation to the harvesting. Further with periodically

[1] Bapuji Institute of Engineering and Technology, Davanagere – 577004,, INDIA
ORCID ID: 0000-0002-8762-980X
[2] SSIT., Tumakuru – 572102, INDIA
[3] Dayananda Sagar College of Engineering, Bangalore-560078, India
4 KIAMS, Harihara,-577601 India
** Corresponding Author Email: bmgeethamallad027@gmail.com*

recorded observations of vegetative indices and the resultant yield during crop cutting experiment, the correlation analysis was carried out and the regression model was arrived, to estimate the crop yield with minimal inputs of highly predictive indices as found fit in the analysis.

3. Study Area

In this study, Davangere taluk of Karnataka state of India is selected as region of Interest/Study area, where paddy is the major crop. Pre-processed satellite images from Google Earth Engine for the area of interest was obtained and false color composite was obtained by band stacking of three bands B4,B3 and B2. Since the study area comes under four tiles, four different images must be mosaicked to cover required study area. After mosaicking, the study area was superimposed on it to clip the area of interest for further processing.

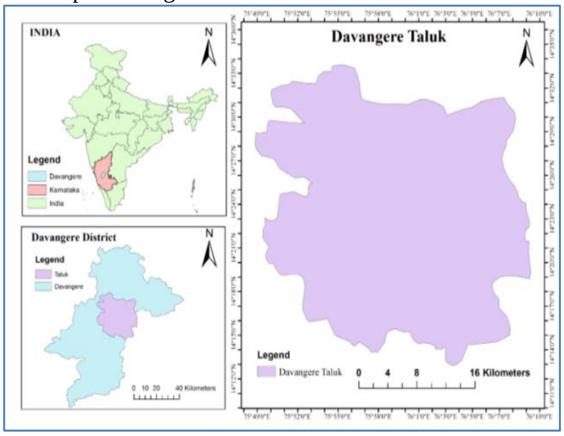

Figure 1. Study Area (Davangere Taluk)

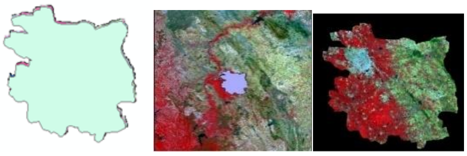

Figure 2. a) Shape file of study area b) Study area superimposed on Mosaic and FCC Image c) Clipped Image of study area

Figure 1 shows the study area boundaries and Figure 2a shows the shape file of the study area and 2b represents the overlying of study area over the satellite sentinel 2 pre-processed image and 2c shows the complete clipped area of the Area of Interest.

4. Methodology

The present study is intended to estimate crop yield of paddy using some of the well-known vegetative indices. 76 paddy plots under cultivation were chosen in the region of Davangere such that the plots are spread across all soil types found in the region of study. These study plots were also the plots identified by the agriculture department to monitor the crop yield at different geographical areas to estimate crop yield of the district. Such plots are called plots meant for crop cutting experiment (CCE) conducted by the department every season. The selected plots under paddy cultivation in summer crop were monitored for every 15 days (observations for every 5 days recorded towards tail-end/harvesting of the crop to accurately find the cut off duration of harvesting) to note three vegetative indices viz., NDVI, NDWI and SAVI from 17th January of 2021(date of transplantation) to 26th May of 2021 (date of harvesting) as a part of longitudinal study observations.

Accurate Crop yield prediction using satellite images has been a challenge, since it depends on multiple factors such as weather conditions, physical characteristics of plant and the soil, farming practices, micronutrients apart from derived vegetative indices out of satellite image processing viz., NDVI, NDWI, SAVI and many more. The impact of these characteristics on crop yield is not persistent for all geographical locations and hence need evaluative procedure to determine the set of variables which essentially impact the yield. Correlation and multiple regression analysis were applied in the present study to arrive at the combination of those parameters, which would result in best possible crop yield estimates. Longitudinal observations on vegetative indices were collected and subjected to data analysis to get the yield estimates.

5. Data Analysis, Results and Discussion

Farmers were contacted to collect details of plot of the study area about prevailing cropping practice, area under paddy, fertilizers/pesticides used and harvesting practices. Other variables such as soil type, soil profiling were also collected from the state agriculture department to study the relationship of all these variables with the crop yield. In addition to these variables, Sentinel 2 satellite image based vegetative indices viz., NDVI, NDWI and SAVI were also collected at regular intervals of time (fortnightly) starting from paddy transplantation to harvesting.

Normalized Difference Vegetation Index (NDVI) was calculated by using equation (1)

$$\boldsymbol{NDVI} = (\boldsymbol{NIR} - \boldsymbol{R})/(\boldsymbol{NIR} + \boldsymbol{R}) \tag{1}$$

Similarly the other two indices viz., NDWI and SAVI were also calculated using equation (2) and (3).

$$\boldsymbol{NDWI} \ = \ (\boldsymbol{NIR} - \boldsymbol{G})/\ (\boldsymbol{NIR} + \boldsymbol{G}) \tag{2}$$

$$\boldsymbol{SAVI} = \boldsymbol{1.5} * (\boldsymbol{NIR} - \boldsymbol{R})/(\boldsymbol{NIR} + \boldsymbol{R} + \boldsymbol{5000}) \tag{3}$$

Data collected was further subjected to correlation & multivariate regression analyses. Yield estimates of these plots were collected from crop cutting experiments of the state agricultural department. The data so gathered was used for correlation analysis and it was found that there is no statistically significant correlation between crop yield and the other variables (Crop age, Variety, source of irrigation (SOI), fortnightly collected NDVI/NDWI/SAVI values from Jan17 to May 26 of 2021 and soil characteristics such as pH, OC, Nitrogen, Phosphorous, Potassium and other micro nutrients). It is evident from the analysis that the correlation between crop yield and the other parameters considered yields statistically insignificant correlations as the crop yield is a compounded effect of many other factors like weather condition, rainfall, wind speed, farmers' effort & care, agricultural practice, potential of the variety and optimum harvesting time which are not a part of this study. Hence change in the pattern/extent of correlation during the period of study was noted.

However, the correlation between crop yield and the vegetative index NDVI indicate a drop in the extent of correlation at the ripening stage of grains/towards harvesting of the crop (coefficient of correlation r = 0.155 as recorded on April 21 and r = 0.009 as recorded on April 26th of 2021), this can be attributed to the change in crop colour from green to light brown (dried up) crop when observed from the satellite (canopy view). The detailed correlation analysis to ascertain correlation between NDVI values and other variables is as shown in Table1.

5.1 Correlation analysis

Karl-Pearson's correlation coefficient (r) is used to study the correlation between crop yield of paddy (as collected from crop cutting experiments) and other variables concerned to the growth of paddy (Crop age, Variety, source of irrigation (SOI), soil characteristics such as pH, OC, Nitrogen, Phosphorous, Potassium and other micro nutrients and vegetative indices (NDVI, NDWI and SAVI) collected at different time points across the growth phase of paddy after transplantation. Since the correlation between these variables and the crop yield is just partial (as the crop yield is a compounded effect of many other variables such as weather, humidity, wind speed, fertility of soil, potential of seed variety, farmers' care, application of timely fertilizers, pesticides, optimum harvesting time etc., which are not a part of this study) in most of the cases, insignificant correlation is observed. In all such cases, especially when tracking the time series of NDVI, NDWI and SAVI, change in the nature/extent of correlation is observed (rather than its statistical significance) to relate to the change in physical characteristics (crop colour / crop aging) of the crop under study.

Correlation between two variables is tested with the following null hypothesis using t-test at 5% level of significance.

H0: There is no significant correlation between two variables under study (Crop yield and other variables taken one at a time) ……..…(I)

If p-value corresponding to the correlation coefficient is < 0.05, the null hypothesis stated in (I) gets rejected stating that there is a statistically significant correlation between the variables considered. Otherwise if p-value corresponding to the correlation coefficient is ≥ 0.05, the resulting correlation is statistically insignificant (insig.).

Table 1 Correlation of Crop Yield with NDVI & Other Study Parameters

Correlating variables with Crop Yield	*Correlation with Yield*		*Statistical Significance*
	Correlation Coefficient (r)	*Sig.*	
Crop age (Days)	-.181	.118	Insig.
Variety	.136	.243	Insig.
SOI*	.091	.436	Insig.
Jan17**	-.004	.975	Insig.
NDVI Feb1	-.047	.686	Insig.
NDVIFeb16	-.165	.155	Insig.
NDVIMar2	.052	.653	Insig.
NDVIMar17	.112	.336	Insig.
NDVIApr1	.023	.844	Insig.
NDVIApl16	.096	.411	Insig.
NDVIApl21	.155	.181	Insig.
NDVIApr26	.009	.939	Insig.
NDVIMay11	.087	.456	Insig.
NDVIMay26	.085	.465	Insig.
pH	-.017	.882	Insig.
EC	.133	.254	Insig.
N	.059	.611	Insig.
P	.093	.423	Insig.
K	.096	.408	Insig.
S	.091	.436	Insig.
Zn	.064	.584	Insig.
Fe	-.031	.792	Insig.
Cu	.009	.936	Insig.
Mn	.155	.180	Insig.
B	-.094	.419	Insig.

*Source of irrigation

**NDVI values collected from Jan17 to May 26 of 2021in the study area

Table 1 exhibits the correlation between crop yield of paddy and the other variables under consideration. It may be noted that correlation between crop yield and multiple other parameters is insignificant almost for all variables under study. This may be attributed to the fact that crop yield depends on compounded effect of all the variables under study. However some NDVI values viz., NDVIFeb16, NDVIApl21, NDVIApr26, NDVIMay11 are retained in regression model (Table 2 &3) to predict crop yield.

5.2 Regression Model to predict crop yield

Analysis of variance (ANOVA) is a statistical analysis tool which splits an observed mean variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, whereas the random factors do not. In the study, ANOVA test is used to determine the influence that independent variables (different vegetative indices) have on the dependent variable (crop yield) in arriving at a regression (predictive) model.

The regression model uses the following null hypothesis

H0: Regression model is not a good fit between independent (or predictor variables in the present study viz., NDVI/NDWI/SAVI collected periodically) and dependent variable (i.e., crop yield). ..……..(II)

The null hypothesis stated in (I) is tested for its validity by calculating F-ratio, which is a test statistic / test procedure used to find the goodness of fit of regression model between the variables under consideration. Usually it is tested at 5% level of significance (or tested at 95% confidence level).

F-Statistic/ F-Ratio is found to draw statistical inference using Analysis of Variance (ANOVA) as

$F = \frac{Mean\ regression\ sum\ of\ squares}{Mean\ error\ sum\ of\ squares}$ at desired level of significance (α), which is 5% for the present study.

It is calculated as a ratio of mean regression sum of squares and mean error sum of squares with respective degrees of freedom (Degrees of freedom: Number of independent components in a given dataset with one estimable parameter). The degrees of freedom for 'n' samples will be (n-1).

If F-ratio is close to 1 (or corresponding p value >0.05), there is no true variance existing between the groups and hence (H0 stands accepted) the regression model do not fit between the variables considered. Otherwise if F-ratio is away from 1 (or corresponding p value <0.05), there is a true variance existing between the variables under study and hence (H0 stands rejected) the regression model fits between the variables considered.

After the regression model is found fit to the data, the regression coefficients (unstandardized coefficients 'b' and Beta) were obtained as a part of regression analysis. Since Beta gives the regression model by considering units of measurement of independent variables to be same, it is always practiced to use coefficients under 'b' to explain the linear regression between dependent (crop yield) and other independent variables (NDVI/NDWI/SAVI collected at different time points of the crop growth (paddy). The following linear regression model is stated based on the coefficients so obtained.

To further test the significance of the derived coefficients corresponding to the independent variables the following null hypothesis is tested at 5% level of significance using t-test.

H0: There is no significance of independent variable (i.e., unstandardized coefficient 'b' corresponding to NDVI or NDWI or SAVI at specific time points) in explaining the dependent variable Y (i.e., crop yield). ………(III)

If the p-value (corresponding to the coefficient b1for example) is less than 0.05, then we reject the null hypothesis stated in (II) and conclude that an independent variable b1 is statistically significant in explaining the dependent variable Y

$Y= a + b_1x_1+b_2x_2+b_3x_3…..+bnxn$ (1)

Where,
Y is the dependent variable (crop yield)
b1 is the coefficient corresponding to independent variable x1
b2 is the coefficient corresponding to independent variable x2 and so on.

5.3 Regression Analysis using NDVI

Table2 shows the Regression analysis applied with stepwise elimination of inefficient independent variable through iterative process yielded the befitting regression model with sig. value 0.003<0.01 (refer to the rule of rejecting the hypothesis stated in (II)), indicating a good fit for regression model. The corresponding regression model involved predicting variables to be NDVI values recorded on Feb16, April21, April26 and May11 i.e., NDVI values recorded on day after transplantation of paddy would contribute in estimating the crop yield of paddy.

Model Summary - NDVI

Table 2 ANOVA& Regression Model to predict crop yield using NDVI

ANOVA for regression model fit					
Model	*Sum of Squares*	*DF*	*MSS*	*F*	*Sig.*
Regression	141.9	4	35.4	4.44	.003
Residual	567.3	71	7.9		
Total	709.2	75			

Table 3 Regression coefficients using NDVI

Model	*Unstandardized Coefficients*		*Unstandardized Coefficients*		*Sig.*
	b	*SE*	*Beta*	*t*	
(Constant)	15.1	2.3		6.360	.000
Feb16	-8.8	3.8	-.246	-2.259	.027
Apl21	30.9	8.1	.872	3.773	.000
Apr26	-40.7	11.5	-1.027	-3.530	.001
May11	17.2	7.6	.383	2.248	.028

Note from Table3 that for all unstandardized coefficients "b", the corresponding p-values (sig.) are less than 0.05 we reject the null hypothesis (III) stated using observed NDVI values and conclude that there is a statistically significant relationship between the predictor variable (NDVI observations collected on February 16, April 21, April 26 and May 11and the response variable i.e., crop yield. In other words the predictor variable significantly explains the response variable i.e., crop yield.

The resultant regression equation using the relation in equation (1) and the coefficients noted from Table3 was

Crop Yield = 15.123-8.808(NDVI of Feb16) +30.929*(NDVI on Apr21)-40.762*(NDVI on Apr26) +17.277*(NDVI on May11).

Or

Crop Yield = 15.123-8.808(NDVI of 16thday) +30.929*(NDVI on 81st day)-40.762*(NDVI on 86th day) +17.277*(NDVI on 115th day).

5.4 Correlation & Regression analysis using NDWI

Similar study with NDWI values collected on the same time points was subjected to correlation and regression analysis.
Table 4 exhibits the correlation between crop yield of paddy and the other variables under consideration and longitudinal data collected using NDWI. It can be noted that correlation between crop yield and multiple other parameters is insignificant almost for all variables under study. This may be attributed to the fact that crop yield depends on compounded effect of all the variables under study. However some NDWI values viz., NDWIFeb16, NDWIMay11 are retained in regression model (Table % & 6) to predict crop yield.

Table 4 Correlation of crop yield with NDWI& other study variables

Correlating variables with Yield	*Correlation with Yield*		*Statistical Significance*
	Correlation Coefficient	*p-value*	
Days	-.181	.118	Insig.
Variety	.136	.243	Insig.
SOI*	.091	.436	Insig.
NDWIJan17**	-.142	.220	Insig.
NDWIFeb1	-.180	.120	Insig.
NDWIFeb16	.205	.076	Insig.
NDWIMar2	-.023	.842	Insig.
NDWIMar17	-.026	.826	Insig.
NDWIApr1	-.047	.689	Insig.
NDWIApl16	-.235*	.041	Insig.
NDWIApl21	-.183	.114	Insig.
NDWIApr26	-.182	.115	Insig.
NDWIMay11	-.271*	.018	Sig.
NDWIMay26	-.138	.236	Insig.
pH	-.017	.882	Insig.
EC	.133	.254	Insig.
OCN	.059	.611	Insig.
P	.093	.423	Insig.
K	.096	.408	Insig.
S	.091	.436	Insig.
Zn	.064	.584	Insig.
Fe	-.031	.792	Insig.
Cu	.009	.936	Insig.
Mn	.155	.180	Insig.
B	-.094	.419	Insig.

*Source of irrigation
**NDWI values collected from Jan17 to May 26 of 2021in the study area

Model Summary –NDWI
The regression model was significant with p-value = 0.013 <0.05 as shown in Table5.

Table 5 ANOVA & Regression Model to predict crop yield using NDWI

ANOVA for Regression Model Fit					
Model	*Sum of Squares*	*DF*	*MSS*	*F*	*Sig.*
Regression	78.9	2	39.4	4.574	.013
Residual	630.2	73	8.6		
Total	709.2	75			

Table 6 Regression coefficients using NDWI

Model	*Unstandardized Coefficients*		*Unstandardized Coefficients*		*Sig.*
	B	SE	Beta	t	
(Constant)	10.5	2.8		3.63	.001
Feb16	5.9	3.3	.194	1.75	.043
May11	-18.0	7.5	-.264	-2.38	.020

As evident from Table 6, for all unstandardized coefficients "b", the corresponding p-values are less than 0.05 and hence we reject the null hypothesis (II) stated using observed NDWI values and conclude that there is a statistical significance of a predictor variable (NDVI observations collected on February 16, April 21, April 26 and May 11 in explaining the response variable i.e., crop yield.

Accordingly obtained the regression equation using (1) and regression coefficients of Table 6 is as follows.

Crop Yield = 10.506+5.909(NDWIFeb16) - 18.059(NDWIMay11)

Or

Crop Yield = 10.506+5.909(NDWI on Day16) -18.059(NDWI on Day115)

5.5 Correlation & Regression analysis using SAVI

Soil adjusted Vegetative Index (SAVI) values collected on the same time points (as that of NDVI and NDWI) were subjected to correlation and regression analysis and the results obtained were as follows.

Table 5 shows correlation between crop yield and other variables under study. It is observed that between SAVI recorded on April 16th and that on April21stof 2021, there is a change in correlation coefficient from r = 0.177 to r = 0. 227 attributing to the change in colour of the crop when observed from the satellite. Though correlation between crop yield and the other variables found insignificant in many cases (as the crop yield may not significantly depend only on one variable under study and that it may be due to the compounded effect of multiple variables), some SAVI observations ate retained by the regression model (Table 8 & 9) i.e., SAVI values read on February1, February 16 and April 21 of 2021.

Regression model was significant with p-value = 0.032<0.05 as shown in Table 8.

Table 7 Correlation analysis of crop yield with SAVI& other study variables

Correlating variables with Yield	*Correlation with Yield*		*Statistical Significance*
	Correlation Coefficient	**Sig.**	
Days	-.181	.118	Insig.
Variety	.136	.243	Insig.
SOI*	.091	.436	Insig.
Jan17**	.031	.792	Insig.
SAVIFeb1	.003	.979	Insig.
SAVIFeb16	-.171	.141	Insig.
SAVIMar2	.074	.524	Insig.
SAVIMar17	.066	.572	Insig.
SAVIApr1	.103	.375	Insig.
SAVIApl16	.177	.126	Insig.
SAVIApl21	.227*	.049	Sig.
SAVIApr26	.178	.123	Insig.
SAVIMay11	.222	.054	Insig.
SAVIMay26	.176	.128	Insig.
pH	-.017	.882	Insig.
EC	.133	.254	Insig.
OCN	.059	.611	Insig.
P	.093	.423	Insig.
K	.096	.408	Insig.
S	.091	.436	Insig.
Zn	.064	.584	Insig.
Fe	-.031	.792	Insig.
Cu	.009	.936	Insig.
Mn	.155	.180	Insig.
B	-.094	.419	Insig.

*Source of irrigation

**SAVI values collected from Jan17 to May 26 of 2021in the study area

Continuing with one more vegetative index SAVI (Table8), the regression analysis yielded regression model being fit with p-value = 0.032<0.05. The variables involved were SAVI values collected on February1, 16 and April 21.

Model Summary - SAVI

Table 8 ANOVA and regression model to predict crop yield using SAVI

ANOVA for Regression Model Fit					
Model	*Sum of Squares*	*DF*	*MSS*	*F*	*Sig.*
Regression	81.2	3	27.0	3.10	.032
Residual	627.9	72	8.7		
Total	709.2	75			

Table 9 Regression Coefficients using SAVI

Model	*Unstandardized Coefficients*		*Unstandardized Coefficients*		*Sig.*
	B	*Std. Error*	*Beta*	*t*	
(Constant)	11.4	2.5		4.43	.000
Feb1	21.0	12.1	.252	1.73	.088
Frb16	-20.9	9.4	-.319	-2.22	.029
Apl21	16.6	7.2	.262	2.30	.024

The resultant model with the help of (1) and regression coefficients from Table 9 wad as follows.

Crop Yield = 11.483+21.042(SAVIFeb1) -20.996(SAVIFeb16) +16.623(SAVIApr21)

Or

Crop Yield = 11.483+21.042(SAVI on Day1) -20.996(SAVI on Day16) +16.623(SAVI on Day81)

6. Results & Discussion

6.1Comparative evaluation of regression models

The factors evident from correlation analysis followed by regression modelling confirm that physical characteristics and micronutrients do not explicitly influence the crop yield whereas the reflectance values based on satellite images and resulting vegetative indices do help in estimating crop yield towards the end of crop maturity (i.e., about 86 days to 117 days for paddy after transplantation). All vegetative indices considered in the study show similar trend of decrease/increase in their values (decrease in NDVI, SAVI and increase in NDWI) after 86 days of transplantation.

It is also indicative from Table2 and Table3, yield prediction models using NDVI and NDWI use 4 reflectance values from February 16(Crop age of 25 days after transplantation) to May11 (110 days) and 2 reflectance values of February16 (25 days) and May11 (110 days) respectively. However prediction model using SAVI uses 3 reflectance values of February1 (10 days), February16 (25 days) and April21 (90 days) as shown in table- 5 which are away from average cut off harvesting day of 100 days and hence be considered as an efficient yield predictor model when compared to that using NDVI and NDWI.
The chart in Figure 3 shows that comparative evaluation of different vegetative indices under study against a time series with 15 days of periodicity set across the paddy growth phases starting from transplantation till harvesting as observed in the study area. One can observe that all indices converge similarly between the dates Apr 26 and May 11 and the same can be found in change/decline in the extent of correlation in Table10.

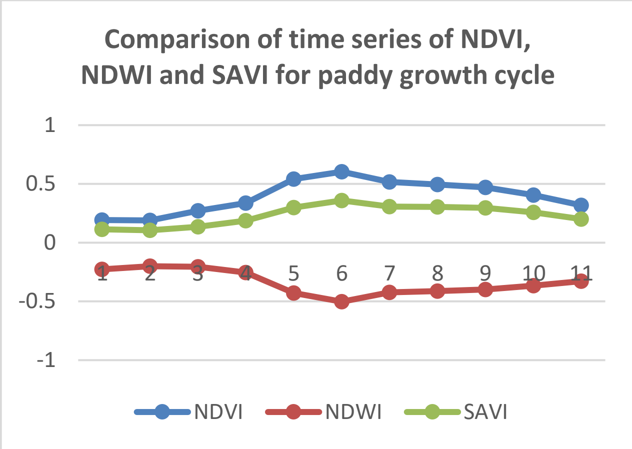

Figure 3. Time series data of NDVI, NDWI and SAVI

Table10 Comparative Evaluation of vegetative indices

	Jan17	***Feb16***	***Mar2***	***Apr1***	***Apr16***	***Apr21***	***Apr26***	***May11***	***May26***
Days	-13	16	31	61	76	81	86	101	117
NDVI	0.2	0.3	0.3	0.6	0.5	0.5	0.4	0.4	0.3
NDWI	-0.2	-0.2	-0.3	-0.5	-0.4	-0.4	-0.4	-0.3	-0.3
SAVI	0.1	0.1	0.2	0.4	0.3	0.3	0.3	0.3	0.2

Figure 3 is a graphical representation of the data exhibited in Table-7 to note the pattern of change in all the three indices after crossing April 26th of 2021, that is between 81 and 101 days of growth after transplantation of paddy signifying the fact that NDVI and SAVI are decreasing (attributing to the change in appearance of crop colour from green to brown when observed from the satellite) and NDWI increasing (indicating the change of watery land to a dry land when the paddy is getting ready for harvesting). Thus it may be concluded that the optimum duration for harvesting paddy is after 100 days of transplantation after confirming the dryness quality/readiness of the crop to store after harvesting.

7. Conclusion and Future Work

Though the crop yield of paddy depends on various factors such as weather, soil type, micronutrients, farmer's timely care, rainfall and many more, it can be precisely estimated only after harvesting, which is impossible in Indian context as there is no unique regulatory body that administers crop harvesting or takes stock of agricultural produce after harvesting. But various agricultural policy making bodies need current year yield of the crop to arrive at better policies/decisions for the next year. The present study attempted to estimate the crop yield using satellite images using vegetative indices such as NDVI, NDWI and SAVI. The change in the correlation between crop yield based on CCE and any of the vegetative index under study indicates suitable time of harvesting after a minimum of 100 days after transplantation of paddy. Further, multiple regression analysis carried out using each of these indices gave a suitable model to estimate crop yield just by knowing 2 to 4 values of these indices from their corresponding time series.
Using SAVI, the number of data points to estimate crop yield were only three, with third SAVI reading collected on 90th day of the crop which is at least a month away from average harvesting duration of 120+ days (practiced in the area of interest by the farmers after judging the optimum dryness of paddy grains) and hence SAVI is considered to be the most suitable vegetative index to estimate the yield of paddy in the area of interest under study.

However model using NDVI uses 4 values ranging from 16th day to 115th day of the crop and the model using NDWI uses only two values, one on Feb 16th and the other on 115th day of the crop, which is the harvesting duration of most of the farmers and hence crop yield prediction beforehand is not possible.

Further studies in this direction with better accuracy/ease can be arrived at by using high resolution satellite images and the appropriate vegetative index among multiple such indices available. The study may be extended with more CCE plots across the district and larger geographical area.

Acknowledgement

We thankfully acknowledge the support extended by the Agricultural Department, the Formers and District Officials of Davangere District in gathering the relevant data for this research work.

Conflicts of interest

The authors declare no conflicts of interest

References

Journal reference

[1] Babar, M. A., van Ginkel, M., Klatt, A. R., Prasad, B., & Reynolds, M. P. The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation.Euphytica, 2006; 150(1–2), 155–172. https://doi.org/10.1007/s10681-006-9104-9

[2] Battude,M.,AlBitar,A.,Morin,D.,Cros,J., Huc, M., Marais Sicre,C.,Le Dantec,V.,& Demarez,V. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. Remote Sensing Of Environment,184 ,668–681. https://doi.org/10.1016/j.rse.2016.07.030

[3] Bognár,P.,Kern,A.,Pásztor,S.,Lichtenberger,J., Koronczay,D.,& Ferencz,C. Yield estimation and forecasting for winter wheat in Hungary using time series of MODIS data. International Journal of Remote Sensing,2017; 38(11), 3394–3414. https://doi.org/10.1080/01431161.2017.1295482

[4] Campos,I.,González Gómez,L., Villodre,J., Calera,M., Campoy,J.,Jiménez, N.,Plaza,C.,Sánchez- Prieto,S. & Calera A, Mapping within-field variability in wheat yield and biomassusing remote sensing vegetation indices Precision Agriculture, 2019; 20(2), 214–236. https://doi.org/10.1007/ s11119-018-9596-z

[5] Clevers,J.,Kooistra,L.,& vanden Brande, M Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop, Remote Sensing, 2017.;9(5), 405. https://doi.org/10.3390/ rs9050405

[6] Geisler, G. Pflanzenbau.Paul Parey. Evans, L.T., & Fischer,R.A.Yield potential: Its definition, measurement,and significance. Crop Science, 1999;39(6),1544–1551. https://doi.org/10.2135/cropsci1999.3961544x

[7] Guo, C., Zhang, L., Zhou, X., Zhu, Y., Cao, W., Qiu, X., Cheng, T., &Tian, Y. Integrating remote sensing information with crop model to monitor wheat growth and yield based on simulation zone partitioning. Precision Agriculture, 2018; 19(1), 55–78. https://doi.org/10.1007/s11119-017-9498-5

[8] Hack, H., Bleiholder, H., Buhr, L., Meier,U., Schnock-Fricke,U.Weber,E.,& Witzenberger, A. Auniform code for phenological growth stages of mono-and dicotyledonous plants—Extended BBCHscale,general-.Nachrichtenblatt Deutscher P flanzen schutzdienst,1992; 44(12),265–270.

[9] Hunt,M.L.,Blackburn,G.A.,Carrasco,L.,Redhead,J. W.,& Rowland,C.SHigh resolution wheat yield mapping using Sentinel-2. Remote Sensing of Environment, 2019; 233, 111410. https://doi. org/10.1016/J.RSE.2019.111410

[10] Journal, Knoblauch, C., Watson, C., Berendonk, C., Becker, R., Wrage-Mönnig, N., &Wichern, F. Relation- ship between remote sensing data, plant biomass and soil nitrogen dynamics in intensively managed grasslands under controlled conditions. Sensors,2017;17(7),1483. https://doi.org/10.3390/s17071483

[11] Marti,J., Bort,J., Slafer,G.A.,& Araus,J. L. Can wheat yield be assessed by early measure- ments of Normalized Difference Vegetation Index. Annals of Applied Biology, 2007;150(2),253–257. https://doi.org/10.1111/j.1744-7348.2007.00126.x

[12] Pasqualotto,N.,Delegido,J.,VanWittenberghe,S.,Rinaldi,M.,&Moreno ,J. Multi-cropgreen LAI estimation with a new simple Sentinel-2 LAI Index (SeLI). Sensors.(2019).https://doi.org/10.3390/ s19040904.

[13] Punalekar,S.M.,Verhoef,A.,Quaife,T.L.,Humphries,D.,Bermingham, L.,&Reynolds,C.K.Application of Sentinel-2 A data for pasture biomass monitoring using a physically based radioactive transfer model. Remote Sensing of Environment,2018;218,207–220.https://doi.org/10.1016/J.RSE.2018.09.028

[14] Serrano,L.,Filella,I.,&Pen,J. Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. Crop Science, 2000; 40, 723–731.

Conference reference

[15] Barnes,E.M.,Clarke,T.R.,Richards,S.E.,Colaizzi,P.D.,Haberland,J,Kostrzewski,M.,Waller,P., Choi,C. Riley, E.,& Thompson,T. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. In Proceedings of the 5th International Conference on Precision Agriculture and Other Resource Management. Bloomington, Madison, USA. 2000.

[16] Baruth, B., Royer, A., Klisch, A., & Genovese, G. The use of remote sensing within the MARS crop yield monitoring system of the European Commission. Proceedings of ISPRS,2008; 27, 935–940

[17] Arumugam, P.; Chemura, A.; Schauberger, B.; Gornott, C. Remote Sensing Based Yield Estimation of Rice (Oryza Sativa L.) Using Gradient Boosted Regression in India. Remote Sensing, MDPI. 2021, 13, 2379. https://doi.org/10.3390/rs13122379

[18] Bose, P., Kasabov, N.K., "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series", IEEE transactions on geo-science and remote sensing, vol. 54, no. 11, pp. 6563-6573, 2016.

[19] Gandhi, N., Armstrong, L.J., Petkar, O., "Proposed Decision Support System (DSS) for Indian Rice Crop Yield Prediction", IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development, pp. 13-18, 2016.

[20] X.Zhou, H.B.Zheng et. al. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery, Volume 130, pp 246-255, 2017.

Authors Biography

Geetha M, DOB: 03-03-1978, Bellary, Karnataka, Post graduated in MCA from Gulbarga University Gulbarga. Working as Assistant Professor, having teaching experience of 20 years. She has published two book chapters and three research papers in International Journals.

Email:asha_gowda@sit.ac.in

Dr. Asha Gowda KaregowdaDOB: 15-09-1971, Tumkur,Karnataka, received M.CA degree from Bangalore University, Karnataka, and M.Phil. Degree from Madurai Kamraj University, Tamil Nadu, Ph.D in Computer Science from Visvesvaraya Technological University, Belgaum, Karnataka. She is currently working as Associate Professor, in the Department of Master of Computer Applications, Siddaganga Institute of Technology, Tumkur, Karnataka since 1998. She is a board of study member and BOE member for few of the deemed and autonomous institutes in Karnataka. Her area of interest includes data mining, data analytics, data science, data structures, machine learning and bio-inspired computing. She has authored a few books, books chapters, published around 90+ papers in International Conferences & Journals and has one patent.

Email:nagaraj.bevinmarad@kiams.ac.in

Dr. Nagaraj B V DOB: 12-11-1973, Hospet, Karnataka, Awarded Ph.D from the University of Mysore. Presently working as Associate Professor at KirloskarInstitute of Advanced Management Studies, Harihar. He has a diverse background of both industry and academia. He was awarded doctorate in the area of agricultural value chains and supply chain systems. Basically being a statistician he has got experience in teaching Statistical Methods, Inferential Statistics, Operations Research & Bio Statistics. Besides teaching, his interest area is data analytics and research in the field of Agricultural marketing and innovation. He has published eight Research papers in reputed Journals of National and International repute.

Email:nandeesha-cvl@dayanandasa

Nandeesha DOB: 06-07-1960,Tumkur, having teaching experience of 37 years and post graduated in M.Sc Geology and Ph.D Hydrogeology from Bangalore University, Bangalore. He has published/presented 39 research papers in National/ International Journals/Seminars. He is guiding 3 Ph.D students under Visvesvaraiah Technological University, Belaguam. His research funded projects are DST and VGST amounting to Rs. 65 lakhs on Groundwater Quality studies and Aquifer characterizations using Remote Sensing and GIS techniques.