A linear approach for wheat yield prediction by using different spectral vegetation indices

Yunus Kaya[*1] , **Nizar Polat** [1]

[1]*Harran University, Geomatics Engineering Department, Türkiye*

Abstract

Yield prediction before harvest is one of the important issues in terms of managing agricultural policies and making the right decisions for the future. Using remote sensing techniques in yield estimation studies is one of the important steps for many countries to reach their 21st-century agricultural targets. The aim of this study is to develop a wheat yield model using Landsat-8 and Sentinel-2 satellite data. In this study, the development stages of winter wheat were examined with the help of satellite images obtained between the years 2015-2018 of a selected region in Sanliurfa, Turkey, and it was aimed to predict the yields for other years by establishing a yield estimation model. The yield estimation model was established with the help of Normalized Difference Vegetation Index (NDVI), Soil-adjusted Vegetation Index (SAVI), Green Normalized Difference Vegetation Index (GNDVI) and Modified Soil-adjusted Vegetation Index (MSAVI) obtained from remote sensing satellite images. Linear regression analysis was established between calculated NDVI, SAVI, GNDVI, MSAVI indices, and actual yield values on the pre-flowering, flowering stage, and post-flowering stage. As a result of the study, the highest correlation coefficient was found in the flowering stage between the vegetation indices values and the actual yield values. The values of NDVI, SAVI, GNDVI, and MSAVI and correlation coefficients are obtained in the flowering stage were 0.82, 0.80, 0.86, and 0.87, respectively. With the established model, yield values in 2019 were tried to be estimated for three different fields. The highest correlations were seen in the flowering stage for MSAVI and GNDVI, pre-flowering stage for NDVI and post-flowering stage for SAVI. This clearly shows that the satellite images can be used in yield estimation studies with a remarkable correlation between vegetation indices and actual yield values.

1. Introduction

Nutrition is the basis of human survival and development [1]. Increasing population [2], air pollution, and reduction of cultivated lands have serious impacts on cereal crops [3] and food security has become a serious problem worldwide. Especially, for agriculture depended on economies like Turkey [4], yields prediction is vital for determining a sustainable supply-demand statistic [5]. In addition, considering in a timely manner predicting yields a few months earlier before harvest provides great advantages for securing national demand, organizing food transport [6,7], forecasting agricultural imports and exports, regulating grain markets, and managing plantations [8,9]. Previously, yield prediction studies were based on field measurements and observations. This traditional method is often subjective, costly, and prone to large errors [10]. Today, thanks to the

developments in remote sensing techniques, it is possible to get more reliable results in agricultural applications by using satellite data. Aside from other possibilities, the satellite systems present high spatial resolution data and the multispectral information about the plants [11]. These possibilities allow us to analyze growth stages and physical conditions. Moreover, compared to terrestrial observations, satellite systems are more economical, more consistent, and more holistic. Pre-harvest yield prediction studies are not only advantageous in predicting yields for specific crops but also in controlling possible disease and pest control during the growing period [12]. Pre-harvest yield prediction and crop monitoring are a common issue for many countries [13]. The use of remote sensing in yield prediction studies is a very active research area [14,15]. There are many studies in the literature to predict pre-harvest yield. The known

earlier studies on the yield prediction model are AgRISTARS [16] and the Large Area Crop Inventory Experiment (LACIE) [17]. The Monitoring Agriculture with Remote Sensing (MARS) project in Europe and the National Agricultural Statistics Service (NASS) service of the US Ministry of Agriculture are play important role in remote sensing studies [18]. [19,20] reported that vegetation indices gave successful results in predicting yield before harvest. [21] used the AgroMetShell model developed by FAO and predict the wheat yield with, a correlation of $R^2 = 0.9067$ between predict and actual yield. [22] used the BBCH scale to monitor the developmental stages of sunflower and wheat plants. [23] used Sentinel-2 data and vegetation indices to monitor the phenological stages of the sunflower plant. Among the VIs, they obtained the best forecast with NDVI ($R^2=0,74$) when three months before harvest of sunflower. [24,25] reported that NDVI based pre-harvest yield prediction includes a 10% error rate. [26] examined the performance of Landsat-8 and Sentinel-2 satellites in their efficiency estimation studies. In Skakun's study, yield estimation was made using only Landsat-8, only Sentinel-2 and NDVI value obtained from the combined model and R^2 values were found to be 0.64, 0.88, and 0.90, respectively. [27] compared the NDVI and SAVI results and they indicated that SAVI performed better than NDVI.

In this study, the success of NDVI, SAVI, GNDVI and MSAVI values obtained from Landsat-8 and Sentinel-2 satellite data provided free of charge for a selected region, unlike previous studies, was examined. The phenological stages of the wheat plant were examined with the vegetation indices obtained from satellite images. A linear regression model was established between the vegetation indices obtained for pre-flowering, flowering, and post-flowering stage, and yield values. The yield prediction model was applied for 2018-2019 data and a model performance accuracy was made.

2. Method

2.1. Study area and data

2.1.1. Study area

The study was conducted in a planting area dominated by winter wheat agricultural city, Şanlıurfa (38°45'E to 38°53'E and 37°06'N to 37°14'N), in the south-eastern part of Turkey (Fig. 1). The region covers approximately 1633 km². The prevailing planting pattern is dominated by an intensive dual-cropping system based on winter wheat, including barley, cotton, maize, isot pepper, lentil [28]. Şanlıurfa has a hot and dry summer climate, and a rainy and mild winter climate. The city, which is close to the Equator in terms of mathematical location, has a continental climate. The region is suitable for wheat agriculture Due to enough rainfall, groundwater, and suitable temperature.
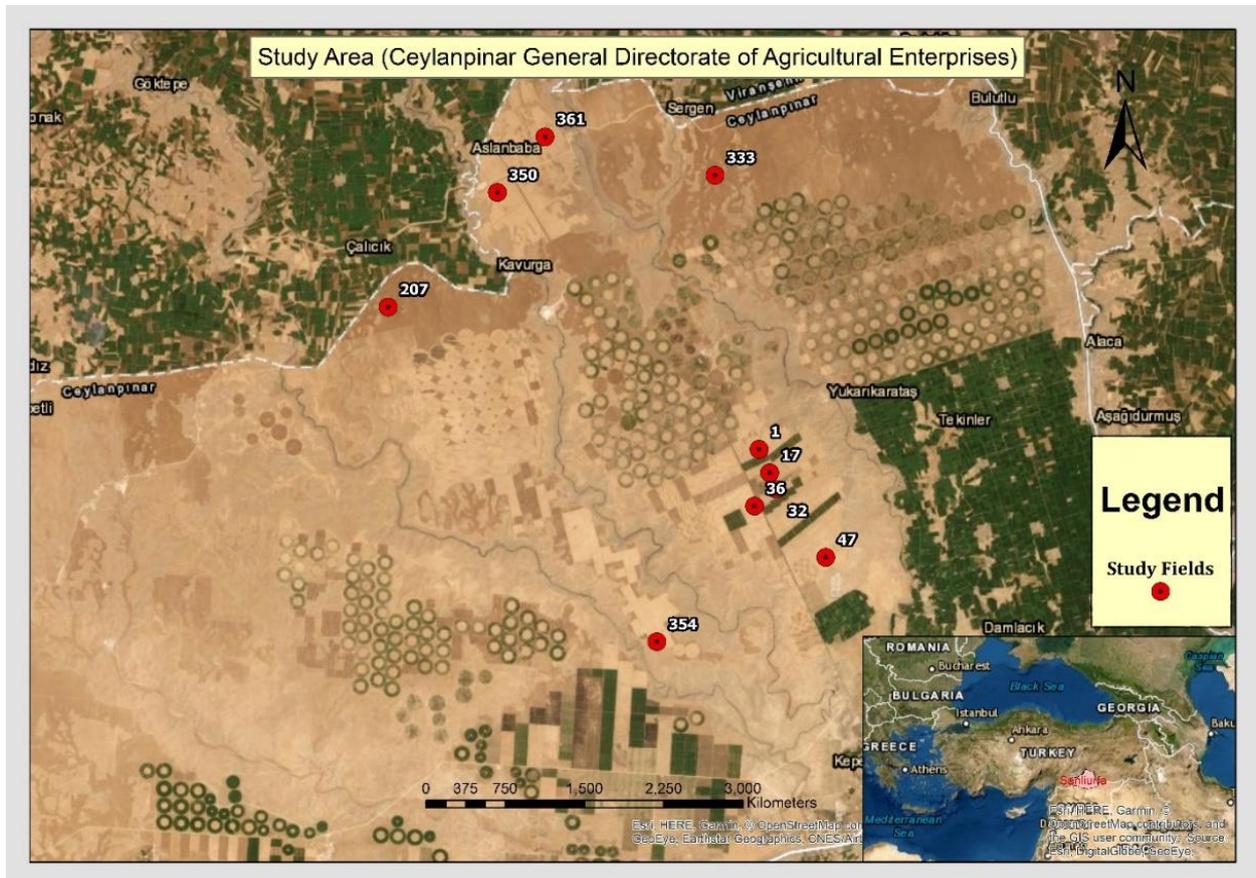
Figure 1. Location of the study area and of the sample plots in the TIGEM of south-eastern Turkey

2.1.2. Satellite data

We used Landsat-8 and Sentinel-2 data to monitor wheat fields. Because Landsat-8 and Sentinel-2 satellites have different temporal resolution [29], and they scan the same region at different times. Therefore, it is advantageous to use two satellite data in combination to monitor changes in plant growth. Landsat-8 data is shared free of charge by a collaboration of the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS). Landsat-8 satellite was launched into space on February 11, 2013. It orbits at an altitude of 705 kilometers and completes one tour in 16 days. Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) sensors were used on the Landsat-8 satellite [30,31]. The Sentinel-2 satellite mission was developed by the European Space Agency (ESA) and is shared free of charge on the internet. The Sentinel-2 mission consists of two polar-orbiting satellites phased at 180 degrees to each other placed in the same orbit. It scans between 56° south and 84° north latitudes. Its temporal resolution is 10 days at the equator. However, since there are 2 satellites in the system, the same place is displayed every 5 days. It has a temporal resolution of 3 days at 45° latitude. Band properties used in the study are shown in Table 1.

In the study, the vegetation indices were created using satellite images obtained between 2015-2019 (3 seasons for the model and 1 season for the test). Satellite images obtained in certain periods from sowing to harvest were downloaded for each season. For 2015-16, 2016-17 and 2017-18 seasons, 8 Landsat-8 and 33 Sentinel-2A satellite images were used from the sowing time (December) to the harvest time (June) of wheat, where the study area was cloudless (Table 2).

2.1.3. Reference data

The yield values of the study area were provided by the General Directorate of Agricultural Enterprises (TIGEM). Since it is forbidden to fly UAVs and measure land in the study area, the field study has not been conducted. Yield values and information on the land were obtained from the institution. In the study, 5 irrigated agricultural lands and 5 dry agricultural lands were used. Wheat plant is not generally cultivated every year [32]. Therefore, wheat cultivation was not made in some fields in some seasons. These fields have been removed from the data set. The parcel yields of the data set used in the study are shown in Table 3.

Table 1. Feature of Landsat-8 and Sentinel-2

	Landsat 8 OLI-TIRS			Sentinel-2		
Scene size (km)	185 x 180			290		
Radiometric Resolution (bit)	12			8		
Band Name	**Band Count**	**Wavelength (μm)**	**Spatial Resolution (m)**	**Band Count**	**Wavelength (μm)**	**Spatial Resolution (m)**
Blue	Band 2	0.45-0.51	30	Band 2	0.490	10
Green	Band 3	0.53-0.59	30	Band 3	0.560	10
Red	Band 4	0.64-0.67	30	Band 4	0.665	10
Near Infrared	Band 5	0.85-0.88	30	Band 8	0.842	10

Table 2. Dates of satellite data (* refers Landsat-8 images)

	2015-2016	2016-2017	2017-2018
December	05.12.2015 19.12.2015 * 25.12.2015	09.12.2016	04.12.2017 08.12.2017 * 14.12.2017
January		08.01.2017 18.01.2017	09.01.2018 * 13.01.2018
February		07.02.2017 17.02.2017 23.02.2017 *	02.02.2018
March	08.03.2016 * 24.03.2016		19.03.2018
April	23.04.2016	28.04.2017	08.04.2018 13.04.2018 23.04.2018
May	03.05.2016	8.05.2017 14.05.2017 * 30.05.2017 *	18.05.2018 23.05.2018
June	02.06.2016 12.06.2016 22.06.2016 28.06.2016 *	07.06.2017 17.06.2017 27.06.2017	07.06.2018 12.06.2018 17.06.2018 22.06.2018 27.06.2018

2.2. Methodology

Health status, growth etc. has a direct impact on yield. Using remote sensing methods, the health status of the plant, its growth rate and the rate of foliage can be determined quickly. The satellite data used for this must go through certain phases. In this study, yield estimation was made using Sentinel-2 and Landsat-8 satellite images. This topic can best be treated under three headings: detection of phenological stage, yield estimation model and application of yield estimation model to test data. Firstly, we applied pre-processing operations on satellite images. The workspace was in two different scenes in Sentinel-2 satellite data. For this reason, Sentinel-2 images were mosaic processed in ENVI software. Secondly, since Sentinel-2 (10 m) and Landsat-8 (30 m) satellite images have different spatial resolution, Landsat-8 images were resampled to 10 m. In the next step, vegetation indices were then created. NDVI and SAVI were used to determine the phenological stages. Then 4 different indices values were determined for each field in the study area. A yield estimation model was established for three stages between index values and yield values. Finally, the model was applied to the same region for the 2018-2019 season. Model performance accuracy was performed for the obtained results (Fig. 2).

2.2.1. Decision of phenological stage

Plants are living things that show continuous development from the sowing period to the harvest period. Therefore, it does not make sense to wait for the harvest period to decide about the health and development of plants. BBCH-scale was used to determine the growth stages of the wheat plant from sowing to harvest [33] (Table 4).

Due to the temporal and spatial resolution of satellite data and cloud conditions, all stages of plant development could not be monitored. Therefore, the growing stages of the wheat plant can be examined in 4 basic stages.

Germination and seedling stage: Germination begins with water intake by a wheat kernel, which ends the postharvest sleeping period. At the germination stage, enough temperature and humidity are needed for wheat seeds to germinate. The optimum temperature at the germination stage of wheat seeds is between 12 ºC and 25 ºC. Under favorable conditions, seedling emergence usually takes place within seven days. The seedling stage begins with the emergence of the first leaf and the top of the plant usually becomes apparent after the third leaf.

Tillering and stem elongation: Sprouts appear shortly after crown formation and crown root system develops. The crown root system provides nutrients and water to the plant during the growing season. After this stage, the wheat plant starts in the stem elongation stage. Most short season wheat produces 7-8 leaves on the stem before stem elongation.

Heading and flowering: This stage is the stage when the wheat ear comes out of the stem completely. After this stage, the plant begins to grow and bloom. During flowering, high temperature and drought reduce the grain yield.

Ripening and senescence: The senescence stage begins after the flowering stage. Ripening takes place in four stages: milk, soft dough, hard dough and finally mature. Meanwhile, the wheat plant turns straw color, and the grain becomes very hard.

Flowering phase is very important for yield estimation of wheat plant. During the flowering phase, the chlorophyll content of the wheat plant reaches its highest level and reaches its greenest form. In the study, by examining the NDVI and SAVI indexes, the flowering phase was determined for each season, and the previous satellite image was evaluated as the pre-flowering phase and the next satellite image post-flowering phase.

Table 3. Wheat Yield (kg/daa) (* refers fallowing field)

CEYLANPINAR TIGEM WHEAT YIELDS (kg/daa)									
Irrigated Agricultural Fields					Dry Agricultural Fields				
Cropland Number / Season	1	17	32	36	47	207	333	350	354	361
2015-2016	496.9	*	*	*	*	285.6	359.9	*	161.0	*
2016-2017	*	627.0	411.0	397.0	532.8	*	*	167.3	*	249.6
2017-2018	506.2	558.4	403.1	371.8	445.0	125.1	46.6	*	37.3	*

Table 4. Phenological Stage of Wheat in the BBCH-Scale (Meier 2001)

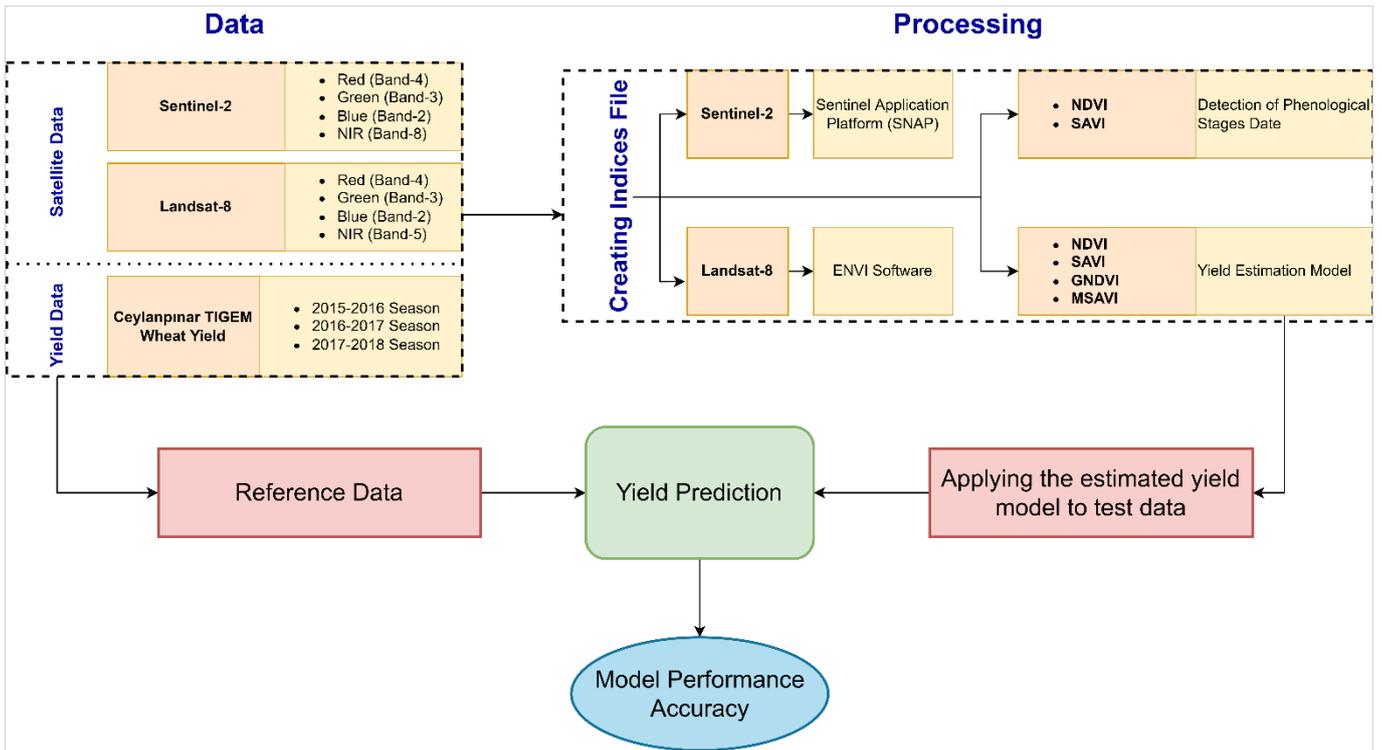| BBCH-Scale	Principal Growth Stage	Stage
0-10	Stage 0	Germination
10-20	Stage 1	Leaf development
20-30	Stage 2	Tillering
30-40	Stage 3	Stem elongation
40-50	Stage 4	Booting
50-60	Stage 5	Inflorescence emergence, heading
60-70	Stage 6	Flowering, anthesis
70-80	Stage 7	Development of fruit
80-90	Stage 8	Ripening
90-100	Stage 9	Senescence

Figure 2. Flow-process diagram

2.2.2. *Vegetation indices*

Vegetation density is one of the important indicators in the plant, soil, and climate formation [34,35]. In vegetation determination studies, the original spectral bands on satellites may not always be enough to determine the characteristic spectral features of the vegetation in the target region [36]. In this case, vegetation indices, which are frequently used in remote sensing studies, help to understand the characteristic features of vegetation [37,38].

The NDVI designed to monitor biomass [39,40] and makes use of the reflection near-infrared (NIR) and red spectrum [41]. The NIR is between 0.76 nm and 0.90 nm in the TM sensor and 0.77-0.90 in the ETM sensor. The red spectrum is between 0.63-0.69 in the TM and ETM sensors. High NDVI values refer to green and healthy plants, low NDVI values indicate that the plant is not green or low amounts of vegetation. Unlike the NDVI, GNDVI uses green wavelength instead of visible red

wavelength. GNDVI is useful in calculating photosynthesis rates and monitoring plant stress. SAVI is basically like NDVI. However, in SAVI, unlike NDVI, the effect of land areas is more. Therefore, the soil correction (L) is used in SAVI calculations according to the vegetation density in the field. [42] reported that the L parameter should take a value close to 0 or very close to 0 in regions with dense vegetation, and a value close to 1 or 1 in areas where vegetation is sparse. The soil correction is expressed with a variable L, for which we select 0.50. Changes occur in the L factor in the SAVI value according to the density of the vegetation. When the L factor is 0, the SAVI value is equal to the NDVI value. However, the soil correction factor (L) must vary according to the amount of vegetation available in order to obtain the optimum setting for the effect of the soil. Thus, the modified SAVI (MSAVI) index is obtained [43] (Table 5).

Table 5. Equations of Vegetation Indices

Index	Equation	References
NDVI	$\dfrac{NIR - Red}{NIR + Red}$	[44]
GNDVI	$\dfrac{NIR - Green}{NIR + Green}$	[45]
SAVI	$\dfrac{NIR - Red}{NIR + Red + L} x(1 + L)$	[46]
MSAVI	$\dfrac{2xNIR + 1 - \sqrt{(2xNIR + 1)^2 - 8x(NIR - Red)}}{2}$	[47]

$$Yield\ prediction\ = a\ x\ Vegetation\ Index + b \qquad (1)$$

The peak date of the vegetation indices values (NDVI, GNDVI, SAVI and MSAVI) during the year is the heading and flowering stage. While establishing the yield model in the study, pre-flowering, flowering stage and post-flowering reflectance values were evaluated separately and the relationship between each stage and yield was investigated. The average values of each field were matched with the amount of yield in the season and linear regression analysis was performed between these two variables. The yield prediction model is established by the equation 1.

3. Results

First step of the yield prediction is to determine the phenological stages of the wheat. Hence a continuous remote sensing data set is required. In this section, determination of phenological stages and establishment of yield estimation model will be explained, respectively. Finally, the obtained yield prediction model was tested for the season of 2018-2019 and the model accuracy was determined.

3.1. Detection of phenological stages

The development processes of plants can be determined numerically with the vegetation indices obtained from satellite images, and it is possible to comment on the analyzed field with the help of the graphics. NDVI and SAVI values were obtained from all available satellite images in the process from sowing to harvest for the investigation of phenological stages. The changes in NDVI and SAVI values were examined for each season. The peaks of the index values were determined as the flowering stage, the previous date pre-flowering stage, and the next date the post-flowering stage. (Table 6).

Table 6 shows that the flowering stage is towards the end of April in the 2015-2016 season and at the beginning of May in the 2016-2017 season. In the 2017-2018 season, the flowering stage occurred in the middle of March (19 March).

In the yield prediction model, the indices value pre-flowering, flowering and post-flowering period were used. NDVI, GNDVI, SAVI and MSAVI values for three seasons in three periods are shown in Table 7.

3.2. Yield estimation model

We used simple linear regression model, that is, a model with a single regressor x that has a relationship with a response y that is a straight line. Simple linear regression model is given in equation 2.

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2)$$

where the intercept β_0 and the slope β_1 are unknown constants and ε is a random error component. The error does not depend on the value of other error. Correlation coefficient is calculated by equation 3.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \qquad (3)$$

Linear regression analysis was performed between NDVI, GNDVI, SAVI and MSAVI values and actual yield values in wheat cultivated areas. A yield estimation model has been established to determine yield values. Yield equations, correlation coefficients and determination coefficients are shown in Table 8(a).

Table 8(a) shows that when all indices are examined, the correlation coefficient obtained during the flowering period is higher than the other stages. The highest correlation coefficient was determined in MSAVI (0.87). Correlation coefficients for GNDVI, NDVI and SAVI were 0.86, 0.82 and 0.80, respectively.

3.3. Model performance accuracy

The yield estimation model obtained from experiments shown in Fig. 3. Position number 1 in Fig. 3 is in the same position as the reference data used in the yield estimation model. Field numbers 2 and 3 are approximately 72 km from field number 1.

NDVI, GNDVI, SAVI and MSAVI indices were generated from remote sensed Landsat-8 and Sentinel-2A satellite data at certain time intervals from planting to harvest. The obtained index values were determined pre-flowering, flowering stage, and post-flowering, and the indices values at these dates were applied to the equation obtained in the yield model. The yield values of the cultivated fields have been tried to be estimated by using all indices values for every three periods and the results are shown in Table 8(b, c and d) for test areas 1, 2, and 3, respectively. Model accuracy was calculated with equation 4.

Table 8(b) shows that the most suitable result out of four indexes is during flowering with MSAVI. GNDVI and MSAVI also achieved 95.88% and 99.99% accuracy at the flowering stage, respectively. In NDVI, the highest accuracy was achieved pre-flowering (95.80%). In addition, high accuracy was obtained in other periods (88.76% for flowering and 87.98% for post-flowering). SAVI is the only index with more than 90% accuracy in all three periods. Although the highest accuracy was achieved after flowering (98.62%), high accuracies were also obtained pre-flowering and during the flowering phase (97.11% for pre-flowering and 93.47% for the flowering period) (Fig. 4).

$$\text{Model Accuracy} = \left(1 - \frac{|\text{ Estimated Yield} - \text{Actual Yield}|}{\text{Actual Yield}}\right) * 100 \qquad (4)$$

Table 6. Dates of Phenological Stages

| Season	Pre-Flowering Date	Flowering Date	Post-Flowering Date
2015-2016	24.03.2016	23.04.2016	03.05.2016
2016-2017	28.04.2017	08.05.2017	14.05.2017
2017-2018	02.02.2018	19.03.2018	18.05.2018
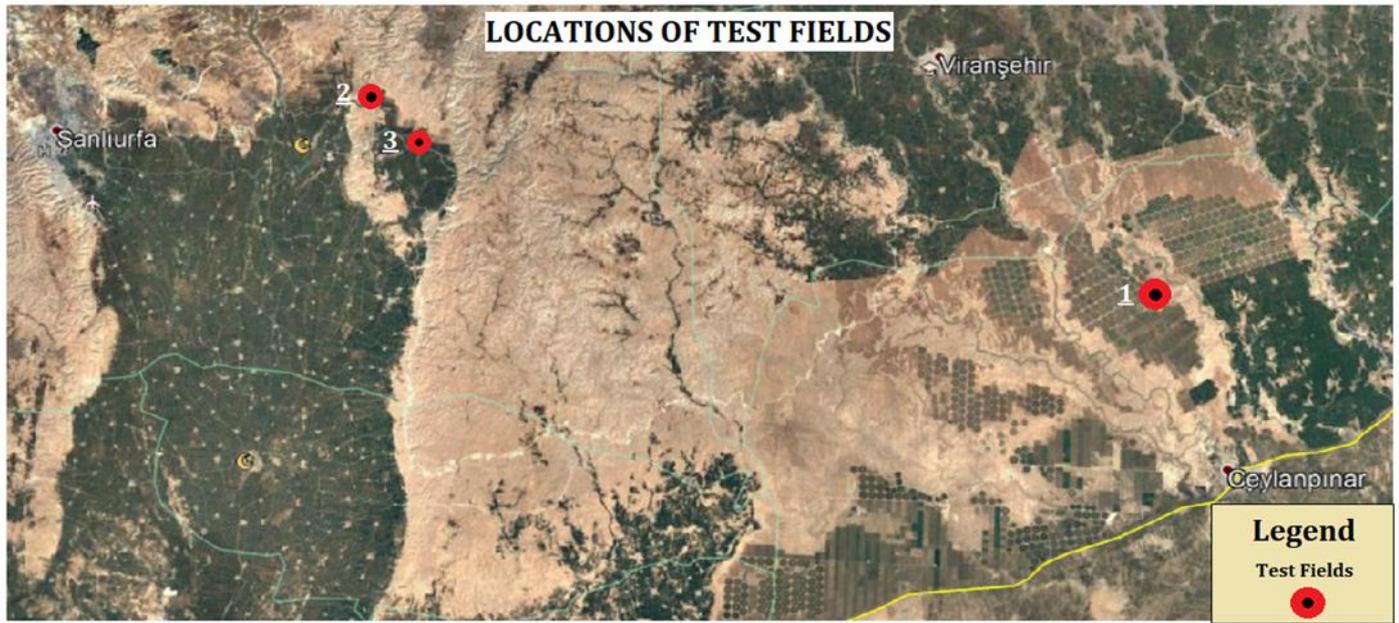
Figure 3. Locations of Test Fields

Table 7. Indices Values for Pre-flowering, flowering and post-flowering seasons (a) NDVI, (b) GNDVI, (c) SAVI, (d) MSAVI

Season	Field Number	(a) NDVI			(b) GNDVI			(c) SAVI			(d) MSAVI		
		Pre-Flowering	Flowering	Post-Flowering	Pre-Flowering	Flowering	Post-Flowering	Pre-Flowering	Flowering	Post-Flowering	Pre-Flowering	Flowering	Post-Flowering
2015-2016	1	0.64	0.81	0.41	0.55	0.7	0.37	0.51	0.59	0.27	0.45	0.63	0.24
	350	0.65	0.67	0.51	0.58	0.59	0.47	0.54	0.64	0.34	0.49	0.48	0.31
	354	0.37	0.4	0.29	0.32	0.3	0.25	0.29	0.3	0.13	0.67	0.09	0.08
	361	0.66	0.75	0.6	0.58	0.67	0.54	0.53	0.6	0.44	0.48	0.6	0.43
2016-2017	17	0.56	0.72	0.43	0.51	0.63	0.56	0.43	0.65	0.44	0.42	0.5	0.4
	32	0.62	0.74	0.44	0.55	0.64	0.57	0.46	0.66	0.45	0.46	0.51	0.42
	36	0.66	0.73	0.43	0.58	0.63	0.54	0.49	0.65	0.4	0.48	0.51	0.39
	47	0.61	0.74	0.44	0.55	0.65	0.58	0.46	0.66	0.48	0.46	0.53	0.43
	207	0.49	0.52	0.28	0.45	0.5	0.39	0.36	0.41	0.23	0.34	0.34	0.18
	333	0.47	0.63	0.31	0.44	0.56	0.44	0.37	0.46	0.24	0.35	0.4	0.24
2017-2018	1	0.48	0.81	0.5	0.42	0.68	0.45	0.29	0.59	0.31	0.27	0.62	0.28
	17	0.49	0.81	0.49	0.44	0.68	0.36	0.3	0.64	0.3	0.28	0.68	0.18
	32	0.47	0.82	0.44	0.41	0.69	0.41	0.28	0.6	0.26	0.25	0.64	0.24
	36	0.39	0.75	0.37	0.38	0.46	0.42	0.25	0.46	0.28	0.23	0.35	0.26
	47	0.52	0.8	0.46	0.44	0.67	0.43	0.31	0.57	0.28	0.28	0.6	0.25
	350	0.47	0.69	0.34	0.46	0.35	0.6	0.3	0.49	0.19	0.27	0.17	0.48
	354	0.24	0.34	0.12	0.3	0.27	0.28	0.09	0.12	0.06	0.11	0.06	0.08
	361	0.37	0.49	0.27	0.41	0.36	0.56	0.23	0.42	0.18	0.21	0.15	0.41

Table 8. (a) Yield equations (y refers yield value; x refers indices value) and Model performance accuracy calculations for phonological stages per vegetation index: (b) Test field 1, (c) Test field 2, (d) Test field 3
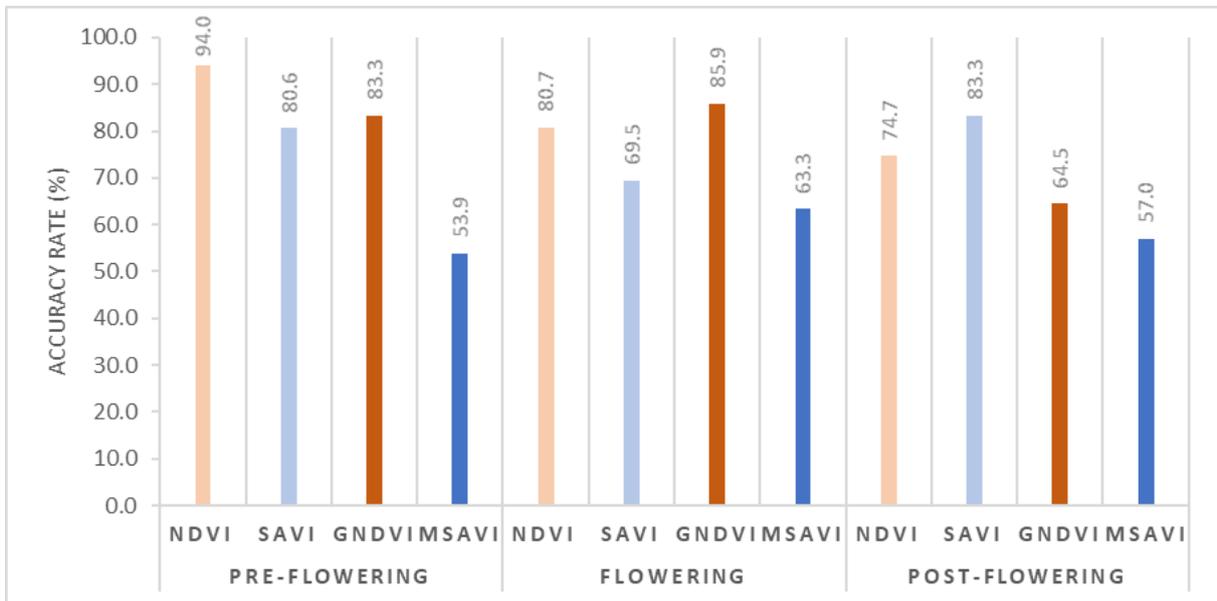
	(a)					(b) Test Field 1			(c) Test Field 2			(d) Test Field 3	
Indices	Stage	Equations	Correlation coefficients (R)	Determination coefficient (R²)	Index Value	Estimated Yield (kg/daa)	Actual Yield (kg/daa)	Index Value	Estimated Yield (kg/daa)	Actual Yield (kg/daa)	Index Value	Estimated Yield (kg/daa)	Actual Yield (kg/daa)
NDVI	Pre-Flowering	y= 894.5x-111.36	0.59	0.35	0.47	310.33	297.3	0.6	420.87	450	0.86	657.91	700
	Flowering	y= 1006x-340.5	0.82	0.67	0.67	334.96	297.3	0.71	373.76	450	0.9	564.9	700
	Post-Flowering	y= 1159x-114.11	0.73	0.53	0.39	337.9	297.3	0.52	488.57	450	0.86	876.84	700
SAVI	Pre-Flowering	y= 698.68x+92.319	0.48	0.23	0.28	288.95	297.3	0.34	326.38	450	0.68	563.93	700
	Flowering	y= 987.73x-178.09	0.8	0.64	0.46	279.09	297.3	0.48	293.55	450	0.67	486.16	700
	Post-Flowering	y= 1147.5x+8.0299	0.74	0.55	0.25	293.27	297.3	0.29	340.8	450	0.71	817.02	700
GNDVI	Pre-Flowering	y=980.68x-113.13	0.48	0.23	0.46	340.78	297.3	0.41	288.95	450	0.71	583.15	700
	Flowering	y=1044.6x-239.74	0.86	0.74	0.5	285.54	297.3	0.63	418.36	450	0.81	601.16	700
	Post-Flowering	y=356.08x+180.76	0.64	0.4	0.41	326.75	297.3	0.46	342.78	450	0.76	451.38	700
MSAVI	Pre-Flowering	y=301.09x+234.53	0.71	0.5	0.32	331.31	297.3	0.24	305.29	450	0.48	377.55	700
	Flowering	y=771.23x+7.5578	0.87	0.75	0.38	297.32	297.3	0.48	377.75	450	0.57	443.3	700
	Post-Flowering	y=325.98x+247.37	0.64	0.41	0.22	318.15	297.3	0.28	338.64	450	0.47	398.95	700
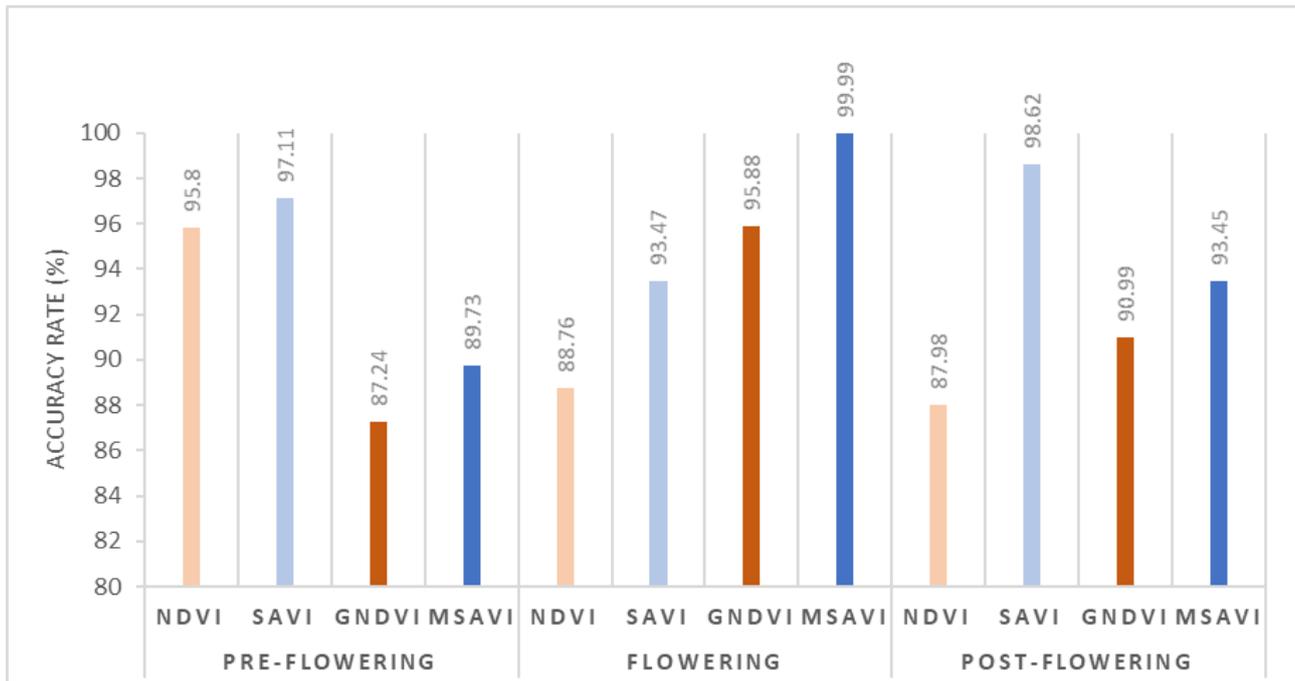
Figure 4. Model Performance Accuracy Analysis for all phenological stages per index for test field 1
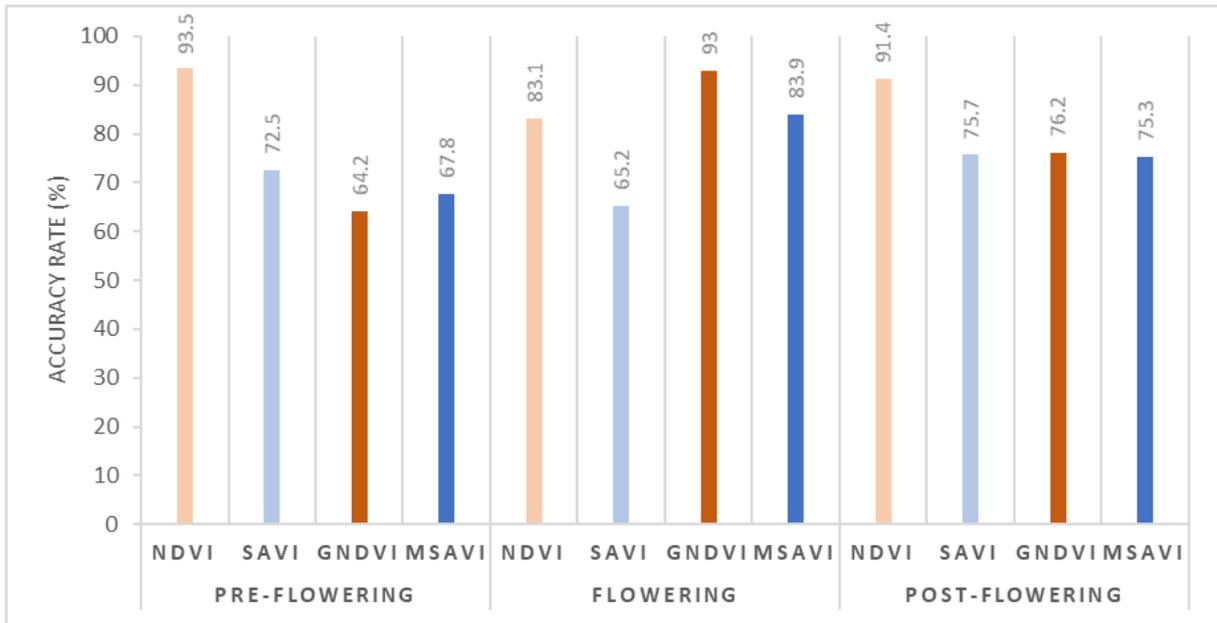
Figure 5. Accuracy Analysis for all phenological stages per index for test field 2

Figure 6. Accuracy Analysis for all phenological stages per index for test field 3

Table 8(c) shows that the highest accuracy is achieved with NDVI pre-flowering. GNDVI and MSAVI achieved 93% and 83.9% accuracy at the flowering stage, respectively. Although the highest accuracy was obtained pre-flowering (93.5%) in NDVI, high accuracy was also obtained in other stages. The highest accuracy in SAVI was obtained post-flowering (75.7%) (Fig. 5).

Table 8(d) shows, as Table 8(c), that the highest accuracy is achieved with NDVI (94%) pre-flowering. The second highest accuracy was obtained with GNDVI in the flowering stage (85.9%). While 83.3% accuracy was obtained post-flowering with SAVI, 63.3% accuracy was obtained at the flowering stage with MSAVI (Fig. 6).

When Table 8(b, c and d) are evaluated together, GNDVI and MSAVI gave more accurate results in the flowering stage compared to the other stages. In addition, compared to other stages, the highest accuracy was obtained pre-flowering in NDVI while it was obtained post-flowering in SAVI.

4. Conclusion

The main aim of this study was to establish a yield estimation model by using four widely used spectral indices. Since the life cycle of a vegetation starts with sowing, passes from several stages, and ends with harvest, firstly a phenological stages determination process was applied to get yearly proper dates for spectral indices. In the next step, a linear regression model was established between vegetation indices and actual yield values for pre-flowering, flowering, and post-flowering stages. The highest correlation of the generated model was observed in MSAVI with a coefficient of r = 0.87. The other correlation coefficients were calculated as GNDVI (r = 0.86), NDVI (r = 0.82) and SAVI (r = 0.80). The generated yield estimation model was applied to get predict actual yield of 3 different test field and compared with its actual value. According to the index-based results, the NDVI was the most successful

index to predict yield with an 87.77% mean accuracy value. GNDVI, SAVI, and MSAVI had 82.36%, 81.78%, and 76.04% mean accuracy value, respectively. Besides, in a phenological stages manner, NDVI was more successful in Pre-Flowering stage with a 94.43 % mean accuracy rate. In Flowering stage, GNDVI reach the best mean accuracy rate with 91.59%. In Post-Flowering stage the SAVI showed the highest mean accuracy rate 85.87%.

Although this comprehensive study reached notable results yet includes limitations. Firstly, the resolution of the satellite images has direct effect on the results. Secondly the generated model could not apply different fields due to the model is depending on the reference data. This means, for a region which has similar conditions the model should be re generated.

In future studies, it is planned to develop a wider range of yield estimation models from satellite images with higher spatial resolution and images obtained with hyperspectral cameras. In addition, further studies need to be carried out in order to validate field measurements.

Acknowledgement

Author contributions

Yunus Kaya: Methodology, Data curation, Writing-Original draft preparation, Software **Nizar Polat:** Conceptualization, Reviewing and Editing

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Wang, Y., Xu, X., Huang, L., Yang, G., Fan, L., Wei, P. & Chen, G. (2019). An improved CASA model for estimating winter wheat yield from remote sensing images. Remote Sensing, 11(9), 1088.
2. Selim, S., & Demir, N. (2019). Detection of ecological networks and connectivity with analyzing their effects on sustainable urban development. International Journal of Engineering and Geosciences, 4(2), 63-70.
3. Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. Science, 327(5967), 812-818.
4. Uyan, M. (2019). Comparison Of Different Interpolation Techniques in Determining of Agricultural Soil Index on Land Consolidation Projects. International Journal of Engineering and Geosciences, 4(1), 28-35.
5. Knox, J. W., Haro-Monteagudo, D., Hess, T., & Morris, J. (2018). Forecasting Changes in Agricultural Irrigation Demand to Support a Regional Integrated Water Resources Management Strategy. Advances in Chemical Pollution, Environmental Management and Protection, 3, 171-213.
6. Bastiaanssen, W. G. M. & Ali, S. (2003). A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agriculture, Ecosystems and Environment, 94(3), 321–340.
7. Apaydin, C., & Abdikan, S. (2021). Fındık bahçelerinin Sentinel-2 verileri kullanılarak piksel tabanlı sınıflandırma yöntemleriyle belirlenmesi. Geomatik, 6(2), 107-114.
8. Li, H., Chen, Z., Liu, G., Jiang, Z. & Huang, C. (2017). Improving Winter Wheat Yield Estimation from the CERES-Wheat Model to Assimilate Leaf Area Index with Different Assimilation Methods and Spatio-Temporal Scales. Remote Sensing, 9(3), 190.
9. Lipper, L., Thornton, P., Campbell, B. M., Baedeker, T., Braimoh, A., Bwalya, M., Caron, P., Cattaneo, A., Garrity, D., Henry, K., Hottle, R., Jackson, L., Jarvis, A., Kossam, F., Mann, W., McCarthy, N., Meybeck, A., Neufeldt, H., Remington, T., Sen, P. T., Sessa, R., Shula, R., Tibu, A. & Torquebiau, E. F. (2014). Climate-smart agriculture for food security. Nature Climate Change, 4(1068-1072).
10. Reynolds, C. A., Yitayew, M., Slack, D. C., Hutchinson, C. F., Huete, A. & Petersen, M. S. (2000). Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. International Journal of Remote Sensing, 21(18), 3487–3508.
11. Liu, W. T., & Kogan, F. (2002). Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. International Journal of Remote Sensing, 23(6), 1161–1179.
12. Pinter, P. J., Hatfield, J. L., Schepers, J. S., Barnes, E. M., Moran, M. S., Daughtry, C. S. T., & Upchurch, D. R. (2003). Remote sensing for crop management. Photogrammetric Engineering and Remote Sensing, 69(6), 647-664.
13. Fernandez-Ordonez, Y. M., & Soria-Ruiz, J. (2017). Maize crop yield estimation with remote sensing and empirical models. In International Geoscience and Remote Sensing Symposium (IGARSS), Institute of Electrical and Electronics Engineers Inc., 3035-3038.
14. Salazar, L., Kogan, F. & Roytman, L. (2007). Use of remote sensing data for estimation of winter wheat yield in the United States. International Journal of Remote Sensing, 28(17), 3795–3811.
15. Ahmad, I., Saeed, U., Fahad, M., Ullah, A., ur Rahman, M. H., Ahmad, A. & Judge, J. (2018). Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan. Journal of the Indian Society of Remote Sensing, 46(10), 1701-1711.
16. Ferguson, M.C. (1982). Evaluation of Trends in Yield Models: Agristars Supporting Research. December, SR J1-04157, JSC-17428.
17. Tucker, C. J., Holben, B. N., Elgin, J. H. & McMurtrey, J. E. (1981). Remote sensing of total dry-matter accumulation in winter wheat. Remote Sensing of Environment, 11, 171-189.
18. Craig, M. E. (2001). A resource sharing approach to crop identification and estimation. In ASPRS 2001 Proceedings of the 2001 Annual Conference.

19. Raun, W. R., Solie, J. B., Johnson, G. V., Stone, M. L., Lukina, E. V., Thomason, W. E., & Schepers, J. S. (2001). In-Season Prediction of Potential Grain Yield in Winter Wheat Using Canopy Reflectance. Agronomy Journal, 93(1), 131–138.

20. Ren, J. Q., Chen, Z. X., Zhou, Q. B., & Tang, H. J. (2010). LAI-based regional winter wheat yield estimation by remote sensing. Chinese Journal of Applied Ecology, 21(11), 2883-2888.

21. Şimşek, O., & Çakmak, B. (2012). Agrometshell modeli ile buğdayda geleceğe dönük senaryolar ve risk analizi. Tarım Bilimleri Dergisi, 18(3), 187-196.

22. Narin, O. G., Noyan, O. F., & Abdikan, S. (2021). Monitoring Vegetative Stages of Sunflower and Wheat Crops with Sentinel-2 Images According to BBCH-Scale. Journal of Agricultural Faculty of Gaziosmanpasa University, 38(1), 46-52.

23. Narin, O. G., & Abdikan, S. (2020). Monitoring of phenological stage and yield estimation of sunflower plant using Sentinel-2 satellite images. Geocarto International, 1-15.

24. 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.

25. Mkhabela, M. S., 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.

26. Skakun, S., Vermote, E., Roger, J. C. & Franch, B. (2017). Combined Use of Landsat-8 and Sentinel-2A Images for Winter Crop Mapping and Winter Wheat Yield Assessment at Regional Scale. AIMS Geosciences, 3(2), 163–186.

27. Gontia, N. K., & Tiwari, K. N. (2011). Yield Estimation Model and Water Productivity of Wheat Crop (Triticum aestivum) in an Irrigation Command Using Remote Sensing and GIS. Journal of the Indian Society of Remote Sensing, 39(1), 27–37.

28. Benek, S. (2006). Şanlıurfa ilinin tarımsal yapısı, sorunları ve çözüm önerileri. Turkish Journal of Geographical Sciences, 4(1), 67-91.

29. Yağmur, N., Tanık, A., Tuzcu, A., Musaoğlu, N., Erten, E., & Bilgilioglu, B. (2020). Oppurtunities provided by remote sensing data for watershed management: Example of Konya closed basin. International Journal of Engineering and Geosciences, 5(3), 120-129.

30. Zabci, C. (2021). Çok bantlı Landsat 8-OLI ve Sentinel-2A MSI uydu görüntülerinin karşılaştırmalı jeoloji uygulaması: Örnek çalışma alanı olarak Doğu Anadolu Fayı boyunca Palu–Hazar Gölü Bölgesi (Elazığ, Türkiye). Geomatik, 6(3), 238-246.

31. Ahady, A. B., & Kaplan, G. (2022). Classification comparison of Landsat-8 and Sentinel-2 data in Google Earth Engine, study case of the city of Kabul. International Journal of Engineering and Geosciences, 7(1), 24-31.

32. Tanaka, D. L. (1989). Spring Wheat Plant Parameters as Affected by Fallow Methods in the Northern Great Plains. Soil Science Society of America Journal, 53(5), 1506–1511.

33. Meier, U. (2001). Growth stages of mono- and dicotyledonous plants. BBCH Monograph.

34. Gündeş, S., & Peştemalci, V. (2008). Türkiye'nin Bitki Örtüsü Değişiminin NOAA Uydu Verileri ile Belirlenmesi. Ç. Ü. Fen Bilimleri Enstitüsü, 17(6).

35. Kundu, A., Dwivedi, S., & Dutta, D. (2016). Monitoring the vegetation health over India during contrasting monsoon years using satellite remote sensing indices. Arabian Journal of Geosciences, 9(2), 1–15.

36. Üstüner, M., Şanli, F. B., & Abdikan, S. (2016). Bitki örtüsü indekslerinin tarımsal ürün deseni tespitindeki etkisinin araştırılması. In Proceedings of VI. RS&GIS Symposium (VI. UZAL&CBS Sempozyumu), Adana, Turkey.

37. Aboelghar, M., Ali, A. R., & Arafat, S. (2014). Spectral wheat yield prediction modeling using SPOT satellite imagery and leaf area index. Arabian Journal of Geosciences, 7(2), 465–474.

38. Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. Preventive Veterinary Medicine, 11(3–4), 185-200.

39. Konda, V. G. R. K., Chejarla, V. R., Mandla, V. R., Voleti, V., & Chokkavarapu, N. (2018). Vegetation damage assessment due to Hudhud cyclone based on NDVI using Landsat-8 satellite imagery. Arabian Journal of Geosciences, 11(2), 1–11.

40. Peng, W., Wang, J., Zhang, J., & Zhang, Y. (2020). Soil moisture estimation in the transition zone from the Chengdu Plain region to the Longmen Mountains by field measurements and LANDSAT 8 OLI/TIRS-derived indices. Arabian Journal of Geosciences, 13(4), 1–13.

41. Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2), 127-150.

42. Ray, T. R. (1994). Vegetation indices in Remote Sensing. A FAQ on Vegetation in Remote Sensing.

43. Richardson, A. J., & Wiegand, C. L. (1977). Distinguishing vegetation from soil background information. Photogrammetric Engineering and Remote Sensing, 43(12), 1541-1552.

44. Rouse, W., Haas, R. H., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS-1 Symposium, 309-329.

45. Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. Remote Sensing of Environment, 58(3), 289–298.

46. Huete, A. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295-309.

47. Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. Remote Sensing of Environment, 48(2), 119-126.