 doi:10.20944/preprints202209.0416.v1

Sentinel-2 accurately estimated wheat yield in a semi-arid region compared with Landsat-8.

Gaylan R. Faqe Ibrahim [1,2]* Azad Rasul [1] **and** Haidi Abdullah [3,]

[1] Geography Department, Faculty of Arts, Soran University, Kurdistan Region, Soran, 44008, Iraq; gailan.faqe@soran.edu.iq and azad.rasul@soran.edu.iq
[2] Department of Geography, College of Human Sciences, University of Halabja, Halabja, 46006, Iraq; gailan.faqe@soran.edu.iq
[3] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands; abdullah34748@alumni.itc.nl
* Correspondence: gailan.faqe@soran.edu.iq; Tel.: (009647504618417;

Abstract:

Wheat and barley are among the primary food resources of the world population; therefore, their growth and observation are essential in farms to enhance food security worldwide. On top of that, careful observation of the product is essential to find solutions for the issues faced during their production and to reduce the impacts of weather changes. With the advancement of Remote Sensing technology, the observation and estimation process has increased. In this study, numbers of spectral vegetation indices was used along with canopy biophysical properties (LAI) and biochemical properties (chlorophyll), there calculated from (Landsat 8 and Sentinel-2) satellite data. The wheat and barley samples were collected before were be ready for harvest, and a relation with the vegetarian indices was established using the Multi-Linear Regression module, in which the equations used in predicting the harvest were developed and used to create a graph for expected harvest. The result indicated that there is a strong relationship between the vegetation indices of Sentinel-2 and Landsat images and the actual grain yield with R^2 of 0.77 and 0.71, respectively. The results show that the strongest correlation is observed between the LAI data obtained from Sentinel data and cereal yield data, with an R^2 0.68, and the highest correlation for the indices of Landsat images is observed in the NDWI with R^2 0.59 and the lowest degree of error was in the root mean square error (RMSE) for the Sentinel-2 and Landsat 8 with 0.57 and 1.54. In addition, this study also showed that the least relationship for grain yield prediction was observed between the NDRI for Sentinel-2 (R^2 0.1) and SAVI for Landsat image (R^2 0.47).

Keywords: GEE; Landsat 8 OLI; Multi-linear regression; Remote Sensing; Vegetation indices; Wheat and barley.

1. Introduction:

The increase in population and other economic and social factors have a connection to the reduction of food resources. In the past half century, the human population has increased from three billion to more than six billion. A report from the United Nations food and agriculture organization in 2009 states that human population will increase by more than 30% by the year 2050, this indicates that in order to fulfil nutritional requirements the production of food must increase by 70% (Kamilaris et al., 2017a). Thus on a regional and international level, the production of various types of nutrients is essential, by highlighting the type of grain which will have a leading role in the nutrients of advanced countries around the world, through the careful combination of the best available agricultural activity and changing the old way of doing things and the use of modern advanced technology (Khalil & Abdullaev, 2021). Fundamental steps to fixing the food security issue are by estimating the products so that it makes it possible to accurately estimate the amount of product yield before harvesting (You et al., 2017). Estimating the yield value of a cereal grains for a specific season is essential and must be prepared as soon as possible as it allows to carefully plan the retail and quota of cereal by farmers and businesses. The prediction of crop yield can be done for specific locations, based on the crops farms, according to their region or even on a global scale (Panek & Gozdowski, 2021).

Estimation of yield can be defined as the beforehand evaluation of how much yield will be obtained in a specific crop before the final harvest time, while the real yield is the number of crops we gain after the harvest. There are many ways of predicting crop yield to us such as the traditional way of predicting, evaluating, observing and measuring the yield state by experts all throughout the growing season. There are two main methods of predicting the yield of cereals which are remote sensing (RS) and crop simulation models (Basso et al., 2013). Thus the appearance of digital technology and smart agriculture were among the important factors in enhancing the surveillance of agriculture, the production of cereal and its estimation (Khalil & Abdullaev, 2021). In these recent years, with the advancement of remote sensing technology, it has been easier to predict the yield of cereals, by way of the differentiating the location, temporal, and brightness data all around the world, to an acceptable degree (Qiao et al., 2021). The notes from remote sensing shed light on some of the changes that relate to the earth, soil, plant cover and the difference between plant health(Kamilaris et al., 2017b). There has been an increase in the importance of paying attention to and using the data collected from satellites to monitor and predict the yield of cereal crops this is caused by their ability to produce data by location coverage, real time coverage and objectived at product growth (Adeniyi et al., 2020).

Not having a good understanding of how spatial and temporal methods are used for estimating the crop yield is one of the main obstacles faced by Revenue Improvement Managers in a specific crop such as using water, energy and fertilizers to increase the yield (Franz et al., 2020). Using remote sensing data for the process of farming and crop production is very common, especially based on Predictive empirical models, the estimation of cereal yield can be done efficiently and quantitatively (Bose et al., 2016). Researchers have tried to obtain useful information regarding agricultural fields using satellites, which it's estimated would be of great benefit to farmers and food production policy researchers through which it estimates a correct yield for each field per year even this is correct for a piece of land within a field without taking into account the growth of the crop (Sibley et al., 2014).

In recent years, there has been a rapid development in the types and criteria of remote sensing, with the emergence of several aircraft, drones and satellite bases that have been used in various field methods and helped to fill the scientific gaps in this field. In general, remote sensing observations and data have brought about many variables that have led to the understanding of altitude, soil and vegetation cover in terms of crop variety and health, these were all used as initial data to determine the information in a preliminary and relative manner for several complex crops (Franz et al., 2020). For the remote sensing approach to be more useful to fill the gap in this area it requires two important features: (1) additional ground-based calibration, necessary to reduce the time and expense involved in data acquisition production over several regions and years, (2) Determination of the exact level of the field and the degree of proximity to reality (Sibley et al., 2014). Despite all these advances, one of the major challenges that still exists is the collection of ground data required to calibrate and validate remote sensing logarithms at large spatial and temporal scales (Paliwal & Jain, 2020). Most available cereal yield models require a variety of farm data that are not always available or accurate, especially in developing countries (Prasad et al., 2007).

In general, there are two main types of strategies used to estimate grain yield based on remote sensing data. The first model is based on crop growth models, combining remotely sensed data with an agrometeorological model or a biophysical model. Another commonly used method is to experimentally link remote sensing data to crop yields at the local or regional scale. Such correlations are always explored based on the use of some indicators generated from remotely sensed images (Ferencz et al., 2004). Different models have been developed to predict grain yield using remotely sensed data, the most common being the regression model, to develop empirical relationships between normalized difference vegetation index (NDVI) measurements and grain yield Derived from satellites at different times (Huang et al., 2013). NDVI, a product derived from multimodal satellite data, is used to estimate health and monitor vegetation changes, so that higher NDVI indicates more greenery coverage, while lower NDVI indicates loss of growth and viability of the crop (Bose et al., 2016). Vegetation index is one of the main variables used in modeling the relationship between remote sensing data and grain yield. Various plant

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2022 doi:10.20944/preprints202209.0416.v1

diversity indicators have been developed from visible satellite sensors that can provide a lot of information about plant health and biodiversity (Xu et al., 2011). There is a relationship between vegetation indicators (NDVI) and seasonal initial yield, because of this relationship vegetation density can be used as an indirect measure of initial grain yield through clear formation activity of agricultural plants during a certain period before harvest (Adeniyi et al., 2020). The main aim of this study is to predict wheat and barley yield and investigate the relationship between remote sensing derived vegetation indices to experiment with as many vegetation indices as possible with the aim of improving grain yield prediction. The result of the paper will contribute effectively by adding more knowledge to the field to the ongoing attempt to improve prediction techniques and it helps to efficiently predict crop data by using various vegetation indices.

2. Methodology:

2.1. Study area:

The study area is located in the northeastern part of Iraq in Erbil province, between 44° 1" 8' 44° longitude 29'15"44° and 05" 40' 36° latitude 17' 23' 36° East, as it borders Soran district on the north and northeast, and on the south the area is bounded by Koysinjeq district to the center of Erbil province as shown in Figure 1. Although the study area is far from any water bodies, the impact of some of them on the climate of the region is significant, the Mediterranean Sea being the most obvious, which directly affects the climate of the region. The climate of the region is characterized by semi-arid climate, BSh according to Köppen classification (Köppen, 2011). Summers are hot and dry, and winters are cold and rainy. Rainfall is inadequate for the period between October and November, averaging 543 mm annually (Hussein et al., 2017). According to Beurring's classification, the soil of the study area consists of four types of soil such as lithic mixture with limestone, brown soils of medium and deep thickness, covered with Bakhtyary dust, deep chestnut soil, rocky and sloping soil and cracked and rocky soil. (Buringh, 1960).

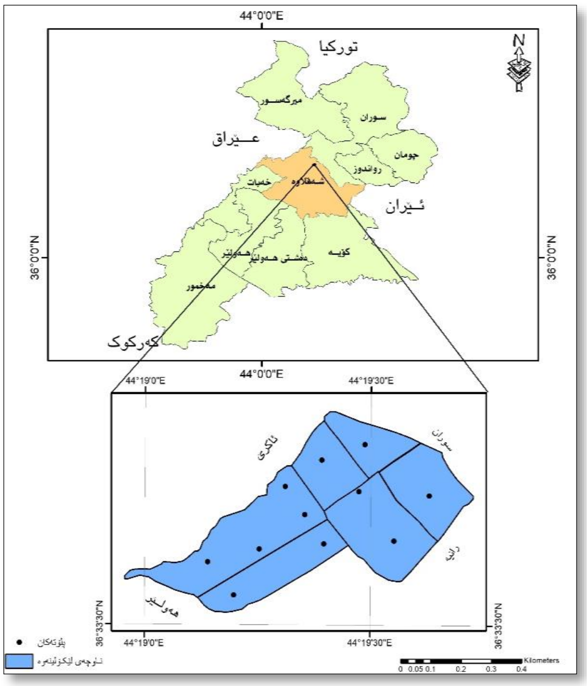


Figure1: illustration of the location of the study area, (Harir plain)

2.2. Data:

In this study, satellite data and field survey data were used to map and predict grain yield through linear regression models. This study is based on Landsat 8 OLI and Sentinel-2A satellite data, processed through Google Earth Engine (GEE) and R Program. All the processing steps and data used are shown in Figure 2. the details of these steps are described in detail in the following subsections.

Landsat 8 OLI and Sentinel-2A satellite image data:

Remote sensing images were acquired by the Operational Land Imager (OLI) onboard (Landsat 8), as well as the MSI-Multi-Spectral onboard (Sentinel-2A). In this study, 4 different time images of Landsat 8 OLI, and 9 Sentinel-2A have been obtained for the study area, from early January to mid-June (2022), as shown in Table 1.

Table 1: Experimental images for this study

No	Sensor	Dates of Pass	Spatial Resolution
1	Landsat 8(OLI)	31-01-2022 16-02-2022 05-04-2022 21-04-2022 08-06-2022	30m

2	Sentinel-2	06-01-2022 26-01-2022 15-02-2022 16-04-2022 21-04-2022 21-05-2022 26-05-2022 31-05-2022 10-06-2022	10m 20m 60m

The Landsat 8 OLI image contains 9wavelength bands with a locational resolution of 30m and 15m for the panchromatic band, namely blue, green, red and near infrared (NIR) and shortwave infrared (SWIR) bands were used. While the Sentinel-2A images include 13 bands of light at different locational resolutions (10, 20, 60 m). Blue, green, red, near infrared (NIR), short wave infrared (SWIR) and vegetation red edge bands were used in this study (Table 2). After testing the images, it was found that the most appropriate date for processing the forecast model is 21-04-2022 for Landsat 8 OLI image and 21-05-2022 for Sentinel-2A image. This is because the highest correlation is shown on this date. In addition, most previous studies suggested that the best date for taking images to predict grain yield is about one to two months before the start of harvest (Li et al., 2021).

Table 2: Details of Landsat 8 OLI and Sentinel-2 image satellites

Details of Landsat 8 (OLI) satellite Images			
Band Number	Spectral range µm	Spatial Resolution (m)	Band Name
1	0.435-0.451	30	Coastal/Aerosol
2	0.452-0.512	30	Blue
3	0.533-0.590	30	Green
4	0.636-0.673	30	Red
5	0.851-0.879	30	NIR
6	1.566-1.651	30	SWIR-1
7	2.107-2.294	30	SWIR-2
8	0.503-0.676	15	Pan
9	1.363-1384	30	Cirrus
10	10.60-11.19	100	TIR-1
11	11.50-12.51	100	TIR-2
Details of Sentinel-2 satellite Images			
Band Number	Spectral range µm	Spatial Resolution (m)	Band Name

1	0.443	60	Coastal/Aerosol
2	0.490	10	Blue
3	0.560	10	Green
4	0.665	10	Red
5	0.705	20	Vegetation Red Edge
6	0.740	20	Vegetation Red Edge
7	0.783	20	Vegetation Red Edge
8	0.842	10	NIR
8A	0.865	20	Vegetation Red Edge
9	0.945	60	water vapour
10	1375	60	SWIR-Cirrus
11	1610	20	SWIR
12	2.190	20	SWIR

2.3. Choosing the Vegetation Indices:

Vegetation indices are changes in multiple brightness bands and wavelengths that allow the monitoring of leaf canopies and photosynthesis activities (Zhao et al., 2020a). In this study, the ability of satellites to predict wheat and barley production was assessed using bands from Landsat 8 and Sentinel-2A, 9 different bands were chosen. Their effect on the estimation process of wheat and barley farms in the study area were determined in Table 3. These indices are generally known for their ability to make use of the spectral properties in the visible spectrum and near-red region. The NDVI index is a very important parameter for models locally, regionally, and globally, for example, biogeochemical models (Huete et al., 2002). The Simple Ratio Index (SR) is sensitive to green plant surfaces, it has a strong connection to the Leaf Area Index as well as the biomass of the leaves(Chen, 1996a). As for Enhanced Vegetation Index (EVI), it is responsive to the structure of leaf canopies, for instance, the Leaf Area Index (LAI), type of leaf canopy, and plant properties (Huete et al., 2002). Ratio Vegetation index (RVI) is very capable of showing the structure of chlorophyll and nitrogen indirectly based on the leaves and the surfaces of the leaves, this also makes it more capable of evaluating ant damage (Ye Tan, 2019). Also, SAVI detects and decreases any changes due to the environment and the earth as well as their impact (Huete, 1988). Although, NDWI senses the change in water ratio in plants, however, it is less sensitive to atmospheric changes compared to NDVI (Gao, 1996). It is also worth noting that GRVI detects changes in plant canopies and determines growth stages. GRVI is arguably more effective for this purpose than NDVI (Ballester et al., 2019). NDRE index is a sensitive index for monitoring chlorophyll levels (Boiarskii & Hasegawa, 2019). Chlorophyll and Leaf Area Index were also used, due to these two parameters have a direct effect on yield.

Table 3: Mathematical formulas of used vegetation indices

Index	Equation	Reference
Enhanced Vegetation Index	$\boldsymbol{EVI} = 2.5 \times \frac{(NIR-RED)}{(NIR+6\times RED-7.5\times BLUE+1)}$	(Liu & Huete, 1995)
Normalized Difference Vegetation Index	$\boldsymbol{NDVI} = \frac{(NIR-RED)}{(NIR+RED)}$	(TOWNSHEND & JUSTICE, 1986)
Normalized Difference Water Index	$\boldsymbol{NDWI} = \frac{(\mathrm{NIR}-\mathrm{MIR})}{(\mathrm{NIR}+\mathrm{MIR})}$	(McFEETERS, 1996)
Soil Adjusted Vegetation Index	$\boldsymbol{SAVI} = \frac{(\mathrm{NIR}-\mathrm{R})}{(\mathrm{NIR}+\mathrm{R}+0.5\times(1.0+0.5))}$	(Huete, 1988)
Simple Ratio	$\boldsymbol{SR} = \frac{\mathrm{NIR}}{\mathrm{Red}}$	(Chen, 1996b)
Ratio Vegetation Index	$\boldsymbol{RVI} = \frac{\mathrm{RED}}{\mathrm{NIR}}$	(Pearson et al., 1972)
Green Ratio Vegetation Index	$\boldsymbol{GRVI} = \frac{(\mathbf{Green}-\mathbf{Red})}{(\mathbf{Green}+\mathbf{Red})}$	(Motohka et al., 2010)
Normalized Difference Red Edge	$\boldsymbol{NDRI} = \frac{(\mathrm{NIR}-\mathrm{RedEdge})}{(\mathrm{NIR}+\mathrm{RedEdge})}$	(Thompson et al., 2019)
Cropping Management Factor Index	$\boldsymbol{CMFI} = \frac{\mathrm{Red}}{(\mathrm{NIR}+\mathrm{Red})}$	(Lin et al., 2010)
Chlorophyll	$\mathbf{Chlorophyll} = \frac{\mathrm{Vegetation\ Red\ Edge}}{(\mathrm{Green})} - 1$	
Leaf Area Index	$\boldsymbol{LAI} = \frac{\left(\frac{0.69-\mathrm{SAVI}}{0.59}\right)}{(0.91)}$	(Sousa et al., 2019)

2.4. Data collection and wheat and barley yield prediction models:

In mid-October 2022, the first field visit was conducted to determine the geographical location (latitude and longitude) of the study area. Location information is recorded by a handheld Global Positioning System (GPS) manufactured by (Garmin) model (62St). In general, the study area is divided into 11 plots (Figure 1). In each plot, 5 geographical locations of Leaf area index were measured with the Viticanopy app and a SPAD instrument has been used to measure chlorophyll, so that 6 LAI and chlorophyll were measured in each plot. In addition, wheat and barley yields are collected. Then the mean yield is obtained for each plot to represent the product because the accuracy of the GPS used was about 3 m. The data collected by GPS were used to generate the prediction model equations and evaluate the accuracy of the prediction model. In order to establish a relationship between vegetation indices and actual aggregate production values. Multiple -Linear Regression model was used for prediction, which mainly allows to explain the relationship between the independent variables and the tested dependent variables (Shastry, 2017). The Multiple –Linear Regression model is the most common form of Linear Regression used as a predictive tool, it allows us to infer the relationship between many independent variables (X1, X2 Xk) and the tested dependent variable (Y). The coefficient of determination (R^2) explains the relative variation of the dependent variable. In other words, it is a measure of the adequacy of the model. The following equation is used to express this relationship.

$\times 1$ ----1

Y- Dependent variable (explained variable),
X1, X2 . . . . Xk- independent variables (explanatory variables),
βo, β1, β2 . . . . βk- Parameters of Eq.
ε- random component (rest of the model).

Landsat 8 and Sentinel 2 imagery was used to develop this model, which resulted in an empirical equation for predicting wheat and barley yields. The most appropriate vegetation indices for prediction were identified in which indices were strongly correlated with yield. A standard set of measures was also used to evaluate the linear regression model for prediction such as the Root Mean Square Error (RMSE), the Mean Squared Logarithmic Error (MSLE), the Mean Square Error (MSE), and the Root Mean Square Logarithmic Error (RMSLE).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(log\ (X_i+1) - log(y_i+1))^2} \quad \text{--------------------2}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i + \hat{Y}i\right)^2 \quad \text{---3}$$

$$\text{MSLE}= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i + 1) - log\ (\hat{Y}i + 1\right)^2 \quad \text{--------------------------4}$$

$$\text{RMSLE}= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(log\ (X_i+1) - log\left(\hat{y}+1\right)^2} \quad \text{----------------------5}$$

In another step to provide an insight into the variance of the farm's yield level distribution, the maps of predicted crops were classified into five distinct zones based on their productivity and greenness co-pillars. These areas included very high yield (18 kg, hectare and above), high yield (16-18 kg, hectare), medium yield (14-16 kg, hectare), low yield (10-12 kg, hectare) and very low area below (10 kg, hectare).

3. Results

Food security in Iraq in general and the Kurdistan Region, in particular, is constantly threatened due to various natural and human impacts on grain production. Therefore, forecasting cereal yields can help plan makers and decision makers to optimize agricultural management and food security under various environmental conditions

In general, satellite vegetation indices are appropriate estimators of plant environmental parameters and grain yield determinations. This study focused on predicting grain yields (wheat and barley) before the harvest process, as well as mapping them. For this purpose, an area planted with wheat and barley containing 539 pixels (Landsat 8 OLI) and 4429 pixels (Sentinel-2) was resampled to 30 m to test several vegetation indices, in order to achieve the objectives of the research, in addition to analyzing and treating satellite images, fieldwork has been conducted and data on the actual production has been collected as shown in Table 4.

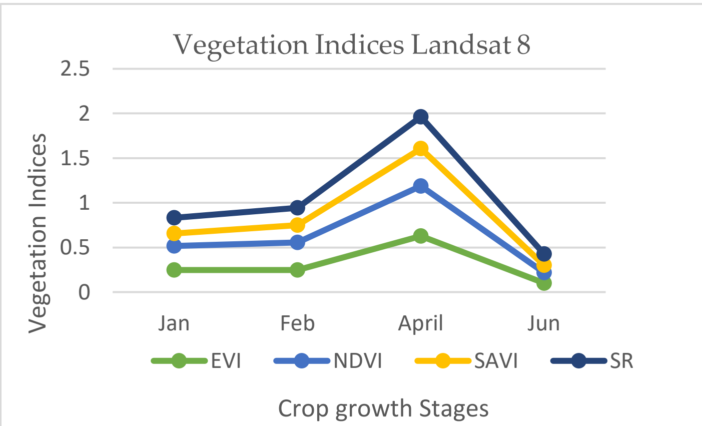
Table 4: Statistical analysis of actual farm data

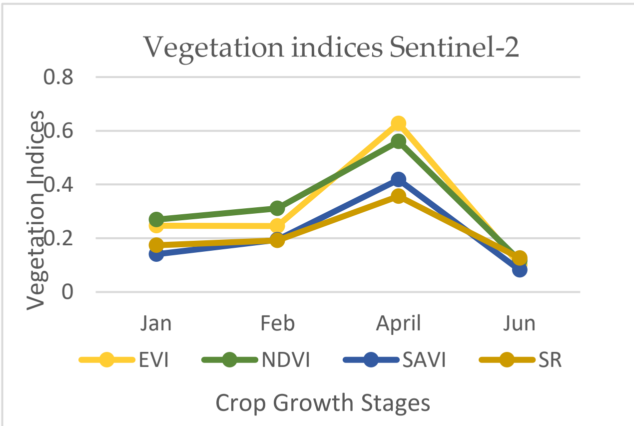
Number of Samples		24
Actual Yield (kg ha^{-1})	Minimum	7.6
	Maximum	18.4
	Mean	11.8
	Std. Deviation	2.9
	Std. Error	0.9

3.1. Temporal profile of vegetation index

The relationship between the estimated and actual data of crop yield, in the growth and early growth stages of plants (the first 70 days), was low, as seen in (Figure 2). This relationship, parallel to the growth of plants, has increased until the 90-100 day age. In the last stages of growth (blooming at 115-130 days) the value reaches its peak (21-April-2022). The stages after blooming saw a decline in terms of vegetation indices until an all time low at the fully developed stage (harvesting). Temporal changes in vegetation indices in this study are parallel to that of previous studies (Nagy et al., 2021; Qader et al., 2018a; Rahman & Robson, 2020). This could be a result of the inefficient reflection of plantation surfaces in the early stages of growth due to ground surfaces. On the other hand, at the late stages of development, the leaves decay and yellow, hence the lack of chlorophyll causes insufficient reflection. The value of vegetation indices always directly impacts the estimation of crops, therefore, in this study, the most accurate prediction methods have been adopted (21/April/2022) from Landsat images and (21-may-2022) from Sentinel images.

Figure 2: Temporal profiles of vegetation and grain growth indices





Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2022 doi:10.20944/preprints202209.0416.v1

3.2. Crop production prediction model:

This study aimed to find a relationship between vegetation parameters and actual crop production to estimate crop yield. For this purpose, 7 indices from Landsat 8 and 11 from Sentinel-2 images have been extracted and Red, NIR, Blue, and MIR bands of Landsat 8 have been utilised. Moreover, Green, Blue, MIR, Red Edge, NIR, Red bands of Sentinel-2 have been used to present the relationship between each vegetation index and the actual counterparts in crops throughout their growth using Multi-linear and Simple Linear Regression Analysis. Table 5 and 6 presents the estimated equations for making a crop prediction map. It is worth noting that each equation needs to be dealt with according to its unique sensor and location. The last stages of development (115-130 days) are found to be the optimal time for estimation of produce, for which two equations are suggested for each individual index. Through algorithmic categorisation, the chosen indices from Landsat 8 images were deemed efficient for prediction. This could be due to the resolution of the location, each pixel representing 30 m, which restricts the control over all the study area plantation surfaces. For the indices from Sentinel-2 images, it was shown that the actual product has a good relationship with the indices except for NDRI.

Table 5. The adequate equations utilized Multi-linear regression

No	Vegetation Index	Landsat 8(OLI)	R^2	Sentinel-2	R^2
1	All Vegetation Indices	Yield(kg ha)= 0.7065* indices + 3.4267	70.7	Yield(kg ha)= 0.8066* indices + 2.2229	76.8

Table 6. The adequate equations were utilized for the prediction of crop yield in the study area (simple linear regression).

No	Vegetation Index	Landsat 8(OLI)	R^2	Sentinel-2	R^2
1	NDVI	Yield(kg ha)=59.6993*NDVI-21.892	0.54	Yield(kg ha)=15.776*NDVI+7.1184	0.58
2	NDWI	Yield(kg ha)=-74.363*NDWI-26.308	0.59	Yield(kg ha)=-24.955*NDWI+3.1207	0.59
3	SAVI	Yield(kg ha)=62.396*SAVI-14.562	0.47	Yield(kg ha)=19.896*SAVI+6.8489	0.59
4	CMFI	Yield(kg ha)=-736.72*MCFI+32.355	0.54	Yield(kg ha)=-22.659*MCFI+18.195	0.46
5	EVI	Yield (kg ha35.334*EVI-10.649	0.51	Yield(kg ha)= 15.775*EVI-7.1676	0.59
6	SR	Yield(kg ha)=52.367*SR-7.1368	0.52	Yield(kg ha)=33.633*SR+5.1946	0.62
7	RVI	Yield(kg ha)=-7.3688*RVI+32.352	0.54	Yield(kg ha)=-1.3006*RVI+19.054	0.55
8	GRVI	-------------------------------------		Yield(kg ha)=125.5*GRVI+116.5	0.56

 doi:10.20944/preprints202209.0416.v1

9	NDRI	-----------------------------------		Yield(kg ha)= -9.0698*NDRI+4.336	0.1
10	LAI			Yield(kg ha)= 1.0307x + 0.4723	0.67
11	Chlorophyll			Yield(kg ha)= 0.541chlorophyll + 5.392	0.54

3.3. Predicting the crop production:

Mult-ilinear regression was used to evaluate the vegetation indices with the measured land yield data for both Sentinel and Landsat. The results showed that there is a strong correlation between the measured actual yield data and predicted yield data of the vegetation indices (Figure 3). High R^2 (0.70) and low RMSE (1.31 kg-ha) for Landsat 8, as well as for Sentinel-2 high R^2 (0.77) and low RMSE (1.27 kg-ha), which revealed a close match between measured and predicted grain yield. In addition, the study demonstrated the relationship of each coefficient with measured crop yield to demonstrate the role of different coefficients in predicting grain yield.

Figure 3: correlation between predicted yield data with actual yield data using Multi-Linear regression

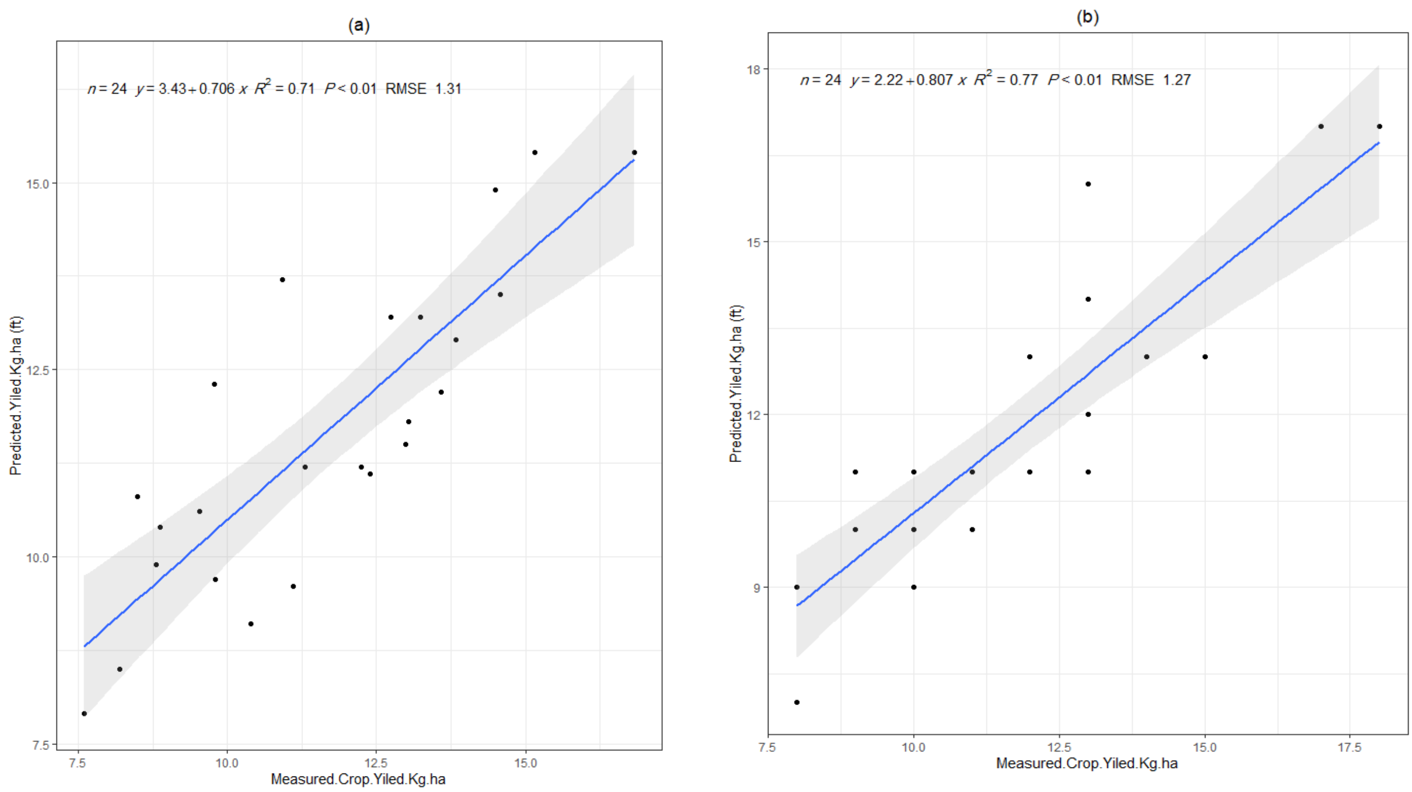


Table 7 and Figure 4,5,6 and 7 show the accuracy of the obtained empirical models (yield prediction) against the actual yield, where the significance level of the model is <0.001. The results show a strong correlation between the estimated images from Sentinel-2 and the actual yield of crops. The highest degree of accuracy was for the LAI index for which the R^2 value was 0.68. The lowest degree of error was in MRSE which was 0.57 kg-hectors. At the same time, the Sentinel Leaf Area Index data are strongly correlated with the SPAD Leaf Area Index data (field data). This shows that this parameter is a suitable measure for predicting grain yield (wheat and barley. The indexes SAVI, EVI and NDWI are of R^2 0.59 value respectively. Then, NDVI, GRVI, RVI and chlorophyll are 0.58, 0.56, 0.55 and 0.54, respectively, and finally, MCFI is 0.46, except for NDRE being 0.1 and proving useless for predicting any sort of actual yield.

Table7: Vegetation index and their relationship with the actual Crop yield.

No	Vegetation Index	Sentinel-2						
		R^2	Std. Dev.	RMSE	MSE	MSLE	RMSLE	Sig.(1-tauled)
1	NDVI	0.58	0.5	1.75	3.08	0.02	0.14	<0.001
2	NDWI	0.59	0.43	2.72	2.98	0.01	0.14	<0.001
3	SAVI	0.59	0.49	1.74	4.04	0.02	0.14	<0.001
4	MCFI	0.46	0.17	1.99	3.96	0.02	0.16	<0.001
5	EVI	0.59	0.71	1.72	2.98	0.01	0.14	<0.001
6	SR	0.62	2.87	1.67	2.80	0.01	0.13	<0.001
7	RVI	0.55	0.44	1.82	3.33	0.02	0.14	<0.001
8	GRVI	0.56	0.05	1.80	3.24	0.02	0.14	<0.001
9	NDRI	0.1	0.03	2.71	7.34	0.04	0.20	<0.001
10	LAI	0.67		0.56	0.32	0.003	0.55	<0.001
11	Chlorophyll	0.54		1.65	2.74	0.02	0.16	<0.001
No	Vegetation Index	Landsat 8(OLI)						
		R^2	Std. Dev.	RMSE	MSE	MSLE	RMSLE	Sig.(1-tauled)
1	NDVI	0.54	0.26	1.65	2.75	0.01	0.13	<0.001
2	NDWI	0.59	0.22	1.54	2.39	0.01	0.12	<0.001
3	SAVI	047	0.21	1.75	3.09	0.02	0.14	<0.001
4	MCFI	0.54	0.02	1.63	2.66	0.01	0.13	<0.001

5	EVI	0.51	0.4	1.63	2.66	0.01	0.13	<0.001
6	SR	0.52	0.27	1.67	2.79	0.01	0.13	<0.001
7	RVI	0.54	0.21	1.73	3.02	0.02	0.14	<0.001

The accuracy of images from Landsat 8 is relatively less than those from Sentinel-2, in a way that the R^2 value was the highest for NDWI at 0.59 and the lowest error was an RMSE of 1.54 kg hectares. The indices NDVI, CMFI, and RVI had R^2 values of 0.54 and error ranges of RMSE at 1.65, 0.63, and 0.73, respectively. The SR and EVI indices yielded an R^2 values of 0.52 and 0.51 and RMSE values of 1.67 and 1.63, respectively. The SAVI index had an R^2 value of 0.47 and RMSE of 1.75 kg hectares.

Figure 4: Maps of predicted crop yield using Landsat 8

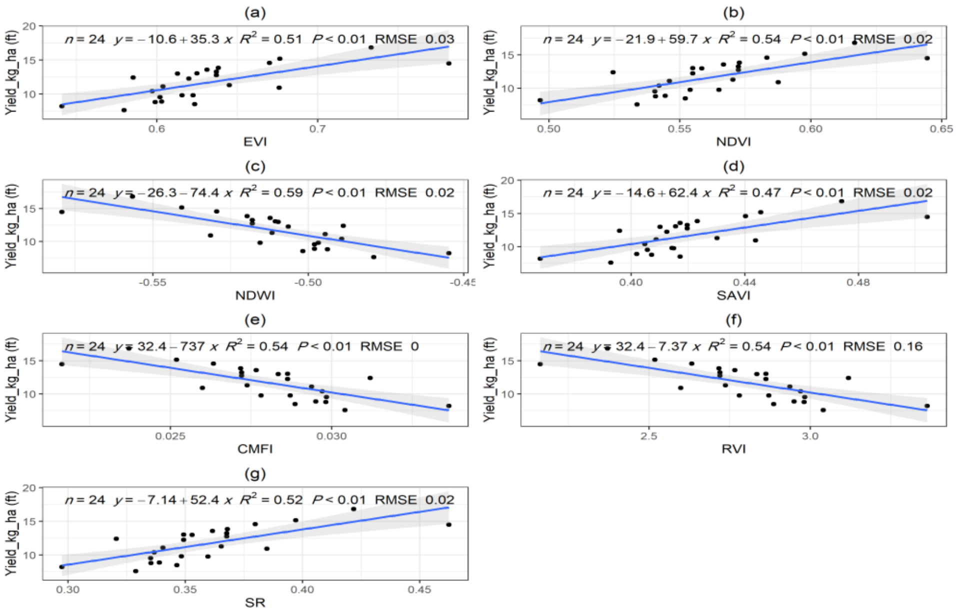
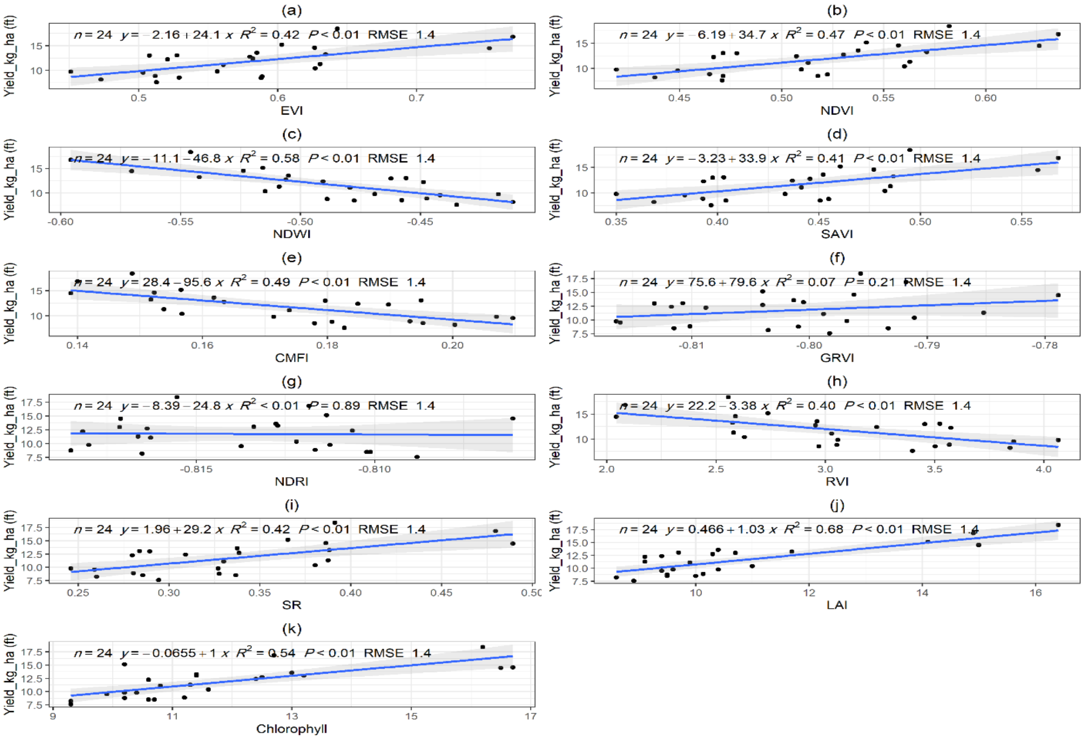


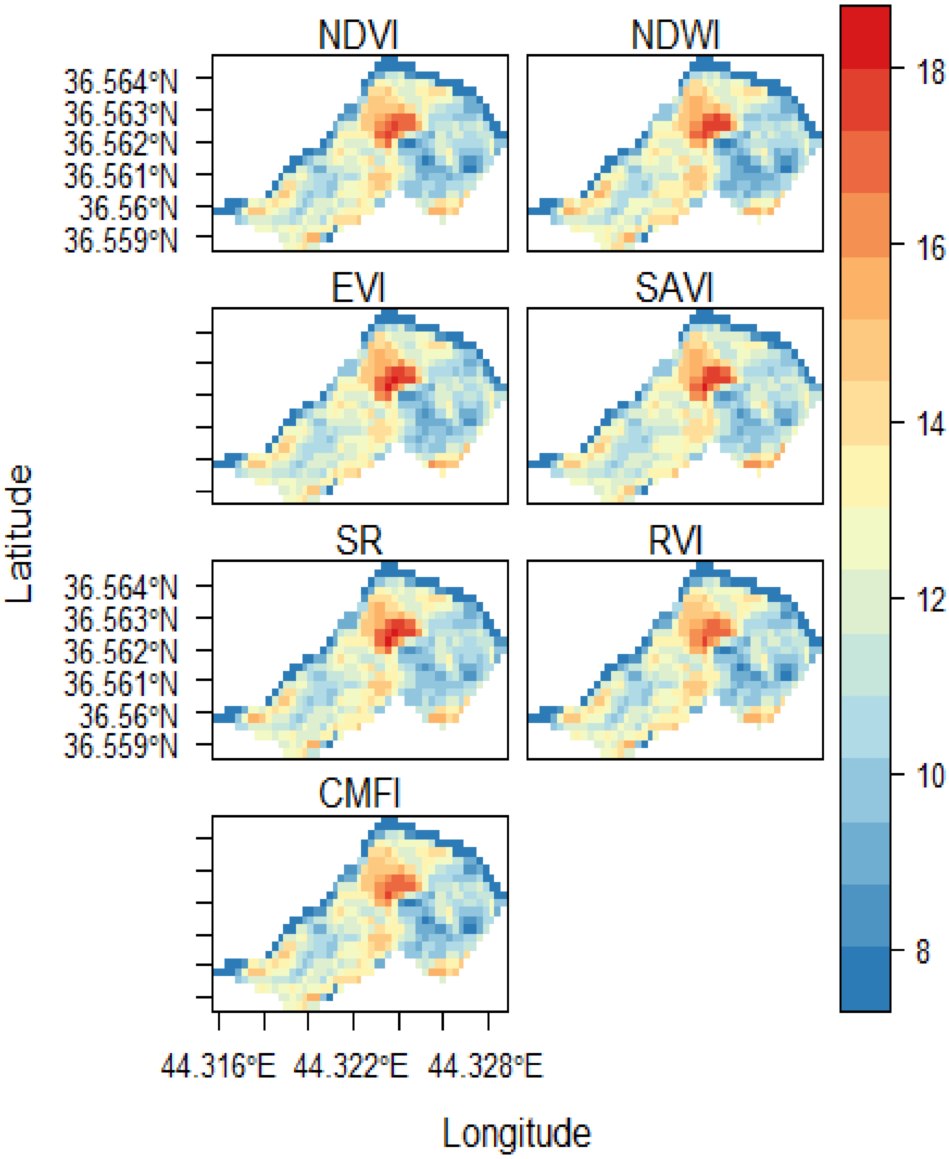
Figure 5: Maps of predicted crop yield using Sentinel-2 and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 September 2022 doi:10.20944/preprints202209.0416.v1



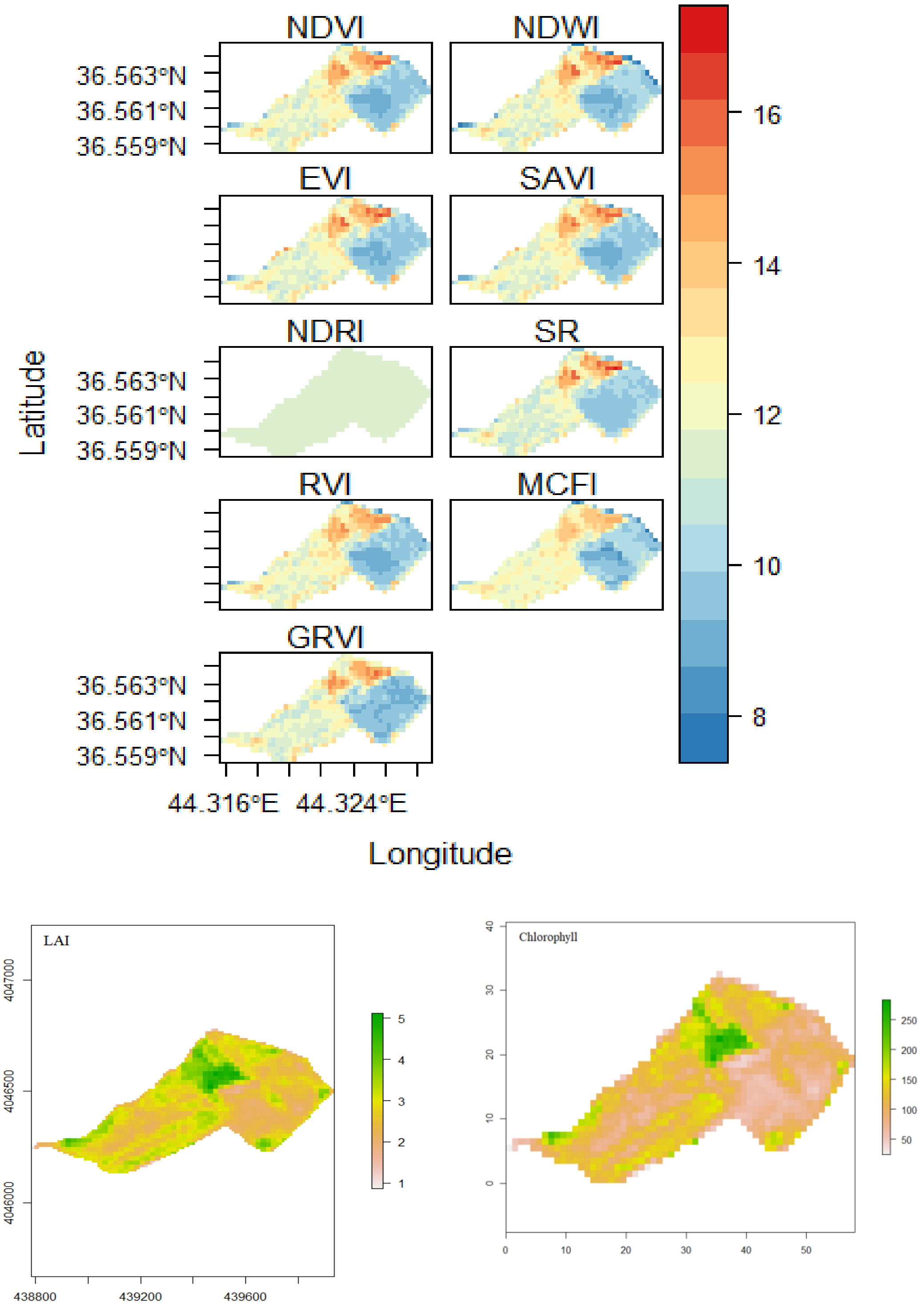
doi:10.20944/preprints202209.0416.v1

Figure 6: Maps of predicted crop yield for Landsat-8 OLI



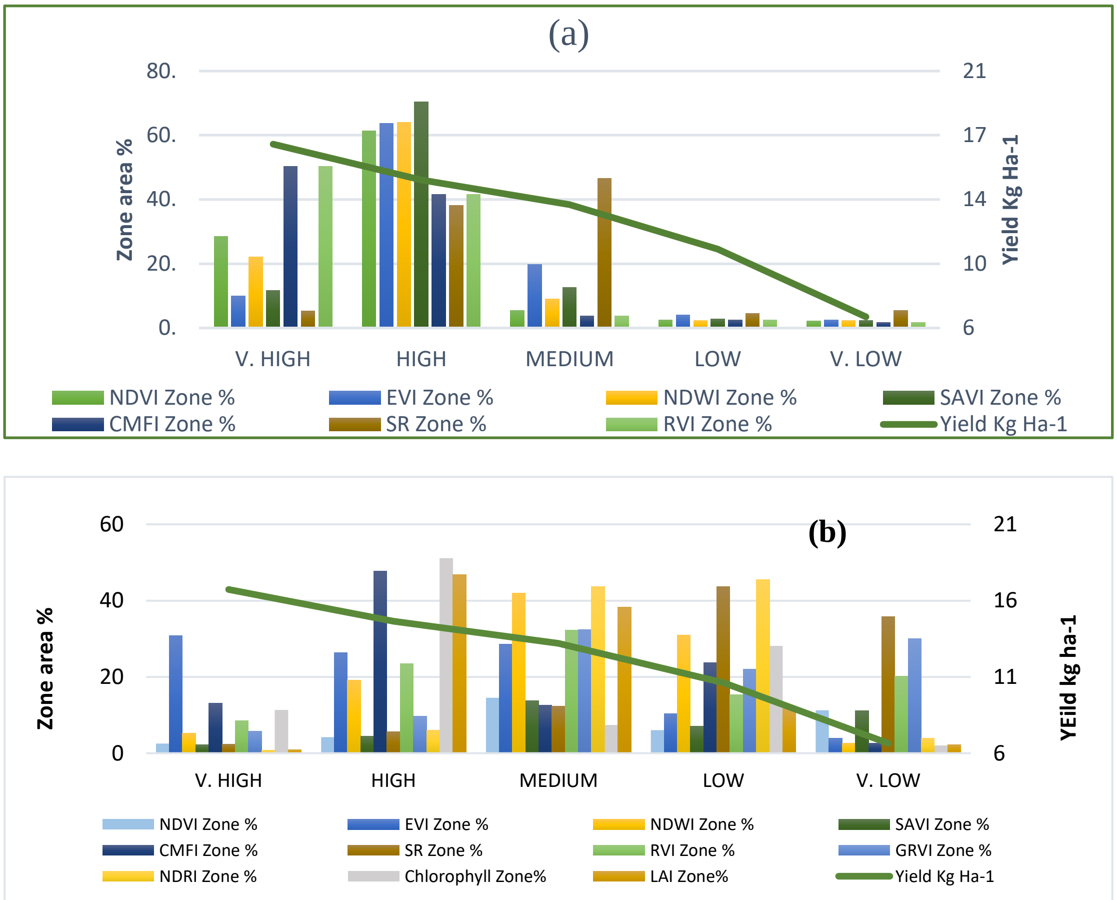
 doi:10.20944/preprints202209.0416.v1

Figure 7: Maps of predicted crop yield for Sentinel-2



In addition, Figure 8 shows the ratio of change in the area of each index compared to the satellite images obtained. In areas with a lot of production (18kgs/hectare), there are huge differences between the indices, in such a way that CMFI and RVI make up 50% of the area in Landsat 8 images. The same indices decrease to about 15% with Sentinel-2 images, in a way that CMFI is roughly 13% of the total area, meanwhile, EVI comes out to 31% of the area. Predicting through Sentinel-2 in dense areas yields opposing results to Landsat 8 results for the same index. Therefore, it can be derived that the combination of these two satellites, NDVI, SR, and SAVI from Landsat 8 and the index CMFI, SR, NDVI and NDWI from Sentinel-2, is efficient for estimation purposes. Future research should attempt to test several other methods for grain prediction and take several grain growing seasons to experiment with different data and methods

Figure-8: The difference in productivity classes of (a) Landsat 8 and (b) Sentinel-2 yield maps



Discussion:

Satellite data have been widely used in agriculture, such as crop monitoring, area estimation, and yield forecasting. (Kumhálová & Matějková, 2017) in their study relied on satellite images with medium spatial resolution (Landsat satellite data) and very high resolution images (QuickBird and WorldView-2 satellites), with artificial crop sensors (GreenSeeker). Each of (Skakun et al., 2019), (Zhao et al., 2020b), (Řezník et al., 2020) and (ADİBAN et al., 2021) used (Landsat 8) and (Sentinel-2) data to map and estimate wheat and barley yields in conjunction with field-based data. (Mkhabela et al., 2011) and (Qader et al., 2018b) also used low spatial accuracy multi-temporal MODIS satellite data to predict wheat and barley yields through NDVI and EVI. Despite this, the Landsat images face certain obstacles when it comes to the prediction of crops since Landsat images generally tend to be spaced out in 15 day periods and are often compromised due to cloud coverage. Therefore, for prediction and monitoring purposes, it needs to be combined with Sentinel data. Since Sentinel is a newly orbited satellite, many studies have taken advantage of its predictive abilities for crops and other farming applications, and it has proven more effective than other multispectral Landsat images.

In this study, seven vegetation indices for Landsat and eleven for sentinel images were utilized to analyze the relationships of vegetation indices with Crop yield for estimating wheat and barley production in the study site. According to the results of previous studies, the results of this study show a strong relationship between vegetation indices and yield of grain data, as well as peak season with the highest value of vegetation indices are considered the best time to predict yields. In this study, 21-4-2022 for Landsat images, 1-5-2022 for Sentinel images is considered the best time to predict grain yield. Regarding the validation of the models adopted for wheat and barley yield prediction, it was found that the use of multi-regression for Sentinel 0.77% has much more accurate prediction results than the use of Landsat images 0.70%, this result is very consistent with the studies in Study (Skakun et al., 2017, 2018) which reported that Sentinel image coefficients have a higher correlation with actual yield than Landsat indicators. In addition, the accuracy of the prediction model is based on different compositions of visible spectral bands, NIR, MIR, SWIR, vegetation red edge, red and green. the best relationship was observed for leaf index area of 0.67% for Sentinel and NDWI for Landsat with 0.59%, which confirmed findings of earlier study (Manivasagam et al., 2021) which showed that the best result for grain yield prediction is LAI.

5. Conclusion

The main goal of this study was to estimate crop production based on vegetation indices. The life cycle of any crop begins with plantation, goes through a few development stages, and gets harvested in the end. Thus, a process to determine the Phonological status of plants was conducted, and appropriate spectral indices were then assigned accordingly. Then, a linear regression model was used for vegetation indices from Landsat 8, Sentinel-2 and real pre-harvesting data. In this study, 9 indices were utilized and it was found that there exists a strong relationship between the indices and real yield of crops such that the R^2 value for Landsat 8 was 70 and for Sentinel-2 was 77, excluding the NDRE index which had no significant relation. The highest apparent connection was shown in the SR index from Sentinel with the lowest RMSE level of 1.67, for which the highest value was from Landsat's NDWI. This study also produced various maps for crop scattering in the study area and differences were highlighted in crop production. The lowest amount of crops was detected on the outskirts of farms. A great difference is apparent due to the various indices, and the highest ranking area has been detected by CMFI and RVI of Landsat 8 and by EVI of Sentinel-2.

References:

Adeniyi, O. D., Szabo, A., Tamás, J., & Nagy, A. (2020). *Wheat Yield Forecasting Based on Landsat NDVI and SAVI Time Series*. https://doi.org/10.20944/preprints202007.0065.v1

 doi:10.20944/preprints202209.0416.v1

ADİBAN, R., Hossein pour, A., & Parchami-Araghi, F. (2021). Predicting Barley Harvest Time in Dryland Conditions Using Satellite Images. *Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi*, *31*, 655–661. https://doi.org/10.29133/yyutbd.909711

Ballester, C., Brinkhoff, J., Quayle, W. C., & Hornbuckle, J. (2019). Monitoring the Effects of Water Stress in Cotton Using the Green Red Vegetation Index and Red Edge Ratio. *Remote Sensing*, *11*(7), 873. https://doi.org/10.3390/rs11070873

Basso, B., Cammarano, D., & Carfagna, E. (2013). Review of crop yield forecasting methods and early warning systems. *Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics, FAO Headquarters, Rome, Italy*, *241*.

Bose, P., Kasabov, N. K., Bruzzone, L., & Hartono, R. N. (2016). Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(11), 6563–6573. https://doi.org/10.1109/TGRS.2016.2586602

Buringh, P. (1960). *Soils and soil conditions in Iraq*.

Chen, J. (1996a). Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Undefined*. https://www.semanticscholar.org/paper/Evaluation-of-Vegetation-Indices-and-a-Modified-for-Chen/854127df9b99c4f6cb1bd373e4793a9761f2de02

Chen, J. (1996b). *Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications*. https://doi.org/10.1080/07038992.1996.10855178

Ferencz, Cs., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai†, Gy., Timár, G., Molnár, G., Pásztor, SZ., Steinbach, P., Székely, B., Ferencz, O. E., & Ferencz-Árkos, I. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, *25*(20), 4113–4149. https://doi.org/10.1080/01431160410001698870

Franz, T. E., Pokal, S., Gibson, J. P., Zhou, Y., Gholizadeh, H., Tenorio, F. A., Rudnick, D., Heeren, D., McCabe, M., Ziliani, M., Jin, Z., Guan, K., Pan, M., Gates, J., & Wardlow, B. (2020). The role of topography, soil, and remotely sensed vegetation condition towards predicting crop yield. *Field Crops Research*, *252*, 107788. https://doi.org/10.1016/j.fcr.2020.107788

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Huang, J., Wang, X., Li, X., Tian, H., & Pan, Z. (2013). Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR. *PLOS ONE*, *8*(8), e70816. https://doi.org/10.1371/journal.pone.0070816

Huete, A. (1988). Huete, A. R. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment. *Remote Sensing of Environment*, *25*, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017a). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017b). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Khalil, Z. H., & Abdullaev, S. M. (2021). Neural network for grain yield predicting based multispectral satellite imagery: Comparative study. *Procedia Computer Science*, *186*, 269–278. https://doi.org/10.1016/j.procs.2021.04.146

Kumhálová, J., & Matějková, Š. (2017). Yield variability prediction by remote sensing sensors with different spatial resolution. *International Agrophysics*, *31*(2), 195–202.

Lin, C.-Y., Chuang, C.-W., Lin, W.-T., & Chou, W.-C. (2010). Vegetation recovery and landscape change assessment at Chiufenershan landslide area caused by Chichi earthquake in central Taiwan. *Natural Hazards*, *53*(1), 175–194. https://doi.org/10.1007/s11069-009-9421-0

Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 457–465. https://doi.org/10.1109/TGRS.1995.8746027

Manivasagam, V. S., Sadeh, Y., Kaplan, G., Bonfil, D. J., & Rozenstein, O. (2021). Studying the Feasibility of Assimilating Sentinel-2 and PlanetScope Imagery into the SAFY Crop Model to Predict Within-Field Wheat Yield. *Remote Sensing*, *13*(12), 2395. https://doi.org/10.3390/rs13122395

McFEETERS, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Motohka, T., Nasahara, K. N., Oguma, H., & Tsuchida, S. (2010). Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology. *Remote Sensing*, *2*(10), 2369–2387. https://doi.org/10.3390/rs2102369

Nagy, A., Szabó, A., Adeniyi, O. D., & Tamás, J. (2021). Wheat Yield Forecasting for the Tisza River Catchment Using Landsat 8 NDVI and SAVI Time Series and Reported Crop Statistics. *Agronomy*, *11*(4), 652. https://doi.org/10.3390/agronomy11040652

Paliwal, A., & Jain, M. (2020). The Accuracy of Self-Reported Crop Yield Estimates and Their Ability to Train Remote Sensing Algorithms. *Frontiers in Sustainable Food Systems*, *4*, 25. https://doi.org/10.3389/fsufs.2020.00025

Panek, E., & Gozdowski, D. (2021). Relationship between MODIS Derived NDVI and Yield of Cereals for Selected European Countries. *Agronomy*, *11*(2), 340. https://doi.org/10.3390/agronomy11020340

Pearson, R. L., Miller, L. D., U.S. International Biological Program, & International Symposium on Remote Sensing of Environment. (1972). *Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie, Pawnee National Grasslands, Colorado*. Dept. of Watershed Sciences, College of Forestry and Natural Resources, Colorado State University.

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

Qader, S. H., Dash, J., & Atkinson, P. M. (2018). Forecasting wheat and barley crop production in arid and semi-arid regions using remotely sensed primary productivity and crop phenology: A case study in Iraq. *The Science of the Total Environment*, *613–614*, 250–262. https://doi.org/10.1016/j.scitotenv.2017.09.057

Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Zhang, L., & Tian, Z. (2021). Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102436. https://doi.org/10.1016/j.jag.2021.102436

Rahman, M. M., & Robson, A. (2020). Integrating Landsat-8 and Sentinel-2 Time Series Data for Yield Prediction of Sugarcane Crops at the Block Level. *Remote Sensing*, *12*(8), 1313. https://doi.org/10.3390/rs12081313

Řezník, T., Pavelka, T., Herman, L., Lukas, V., Širůček, P., Leitgeb, Š., & Leitner, F. (2020). Prediction of Yield Productivity Zones from Landsat 8 and Sentinel-2A/B and Their Evaluation Using Farm Machinery Measurements. *Remote Sensing*, *12*(12), 1917. https://doi.org/10.3390/rs12121917

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

Skakun, S., Franch, B., Vermote, E., Roger, J.-C., Justice, C., Masek, J., & Murphy, E. (2018). Winter Wheat Yield Assessment Using Landsat 8 and Sentinel-2 Data. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 5964–5967. https://doi.org/10.1109/IGARSS.2018.8519134

Skakun, S., Vermote, E., Franch, B., Roger, J.-C., Kussul, N., Ju, J., & Masek, J. (2019). Winter Wheat Yield Assessment from Landsat 8 and Sentinel-2 Data: Incorporating Surface Reflectance, Through Phenological Fitting, into Regression Yield Models. *Remote Sensing*, *11*(15), 1768. https://doi.org/10.3390/rs11151768

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*. https://doi.org/10.3934/geosci.2017.2.163

Sousa, V., Salami, G., Silva, M., Silva, E., Júnior, J., & Alba, E. (2019). Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. *Geology, Ecology, and Landscapes*, *4*. https://doi.org/10.1080/24749508.2019.1608409

Thompson, C. N., Guo, W., Sharma, B., & Ritchie, G. L. (2019). Using Normalized Difference Red Edge Index to Assess Maturity in Cotton. *Crop Science*, *59*(5), 2167–2177. https://doi.org/10.2135/cropsci2019.04.0227

TOWNSHEND, J. R. G., & JUSTICE, C. O. (1986). Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing*, *7*(11), 1435–1445. https://doi.org/10.1080/01431168608948946

 doi:10.20944/preprints202209.0416.v1

Xu, W., Ren, X., & Smith, A. (2011). Remote sensing, crop yield estimation and agricultural vulnerability assessment: A case of Southern Alberta. *2011 19th International Conference on Geoinformatics*, 1–7. https://doi.org/10.1109/GeoInformatics.2011.5980692

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 4559–4565.

Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B., & Hammer, G. L. (2020a). Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sensing*, *12*(6), 1024. https://doi.org/10.3390/rs12061024

Zhao, Y., Potgieter, A., Zhang, M., Wu, B., & Hammer, G. (2020b). Remote sensing Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sensing*, *12*. https://doi.org/10.3390/rs12061024