ISSN: 2148-9173 Vol: 9 Issue:4 December 2022

International Journal of Environment and Geoinformatics (IJEGEO) is an international, multidisciplinary, peer reviewed, open access journal.

Wheat Yield Prediction with Machine Learning based on MODIS and Landsat NDVI Data at Field Scale

Murat Güven TUGAÇ, Ahmet Murat ÖZBAYOĞLU, Harun TORUNLAR, Erol KARAKURT

Abstracting and Indexing: TR DIZIN, DOAJ, Index Copernicus, OAJI, Scientific Indexing Services, International Scientific Indexing, Journal Factor, Google Scholar, Ulrich's Periodicals Directory, WorldCat, DRJI, ResearchBib, SOBIAD

IJEGEO

International Journal of Environment and Geoinformatics 9(4): 172-184 (2022)

Research Article

Wheat Yield Prediction with Machine Learning based on MODIS and Landsat NDVI Data at Field Scale

Murat Güven Tuğaç[1,*], Ahmet Murat Özbayoğlu[2], Harun Torunlar[1], Erol Karakurt[3]

[1] Soil, Fertilizer and Water Resources Central Research Institute, Ankara, TURKIYE
[2] Department of Computer Engineering, TOBB University of Economics and Technology, Ankara, TURKIYE
[3] Field Crops Central Research Institute, Ankara, TURKIYE

* Corresponding author: M. G. Tuğaç
E-mail: mgtugac@gmail.com

Received 14.06.2022
Accepted 25.11.2022

How to cite: Tuğaç et al. (2022). Wheat Yield Prediction with Machine Learning based on MODIS and Landsat NDVI Data at Field Scale. *International Journal of Environment and Geoinformatics (IJEGEO)*, 9(4):035-047 doi. 10.30897/ijegeo.1128985

Abstract

Accurate estimation of wheat yield using Remote Sensing-based models is critical in determining the effects of agricultural drought and sustainable food planning. In this study, Winter wheat yield was estimated for large fields and producer fields by applying Normalized Difference Vegetation Index (NDVI) based linear models (simple linear regression and multiple linear regression) and Machine Learning (ML) techniques (support vector machine_svm, multilayer perceptron_mlp, random forest_rf). In this study, depending on the ecological zone, crop sampling was carried out from 380 rainfed parcels where wheat was planted. On the basis of crop development periods (CDP), the highest correlation between NDVI and yield occurred during the flowering period. In this period, coefficient of determination (R2) was 63% in TIGEM fields and 50% in producer fields for MODIS data, and 61% and 65% for Landsat data, respectively. In TIGEM fields, the best prediction performance was obtained with the MLP model for MODIS (RMSE:0.23-0.65 t/ha) and Landsat (RMSE: 0.28-0.64 t/ha). On the other hand, the highest forecasting accuracy was acquired with the SVM model in producer fields. The RMSE values ranged from 0.74 to 0.80 t/ha for MODIS and 0.51 to 0.60 t/ha for Landsat 8. The error value obtained with MODIS was approximately 1.4 times higher than the Landsat 8 data in producer fields. For yield estimation, the best estimation can be made 4-6 weeks before the harvest. In regional yield estimations, satellite-based ML techniques outperformed linear models. ML models have shown that it can play an important role in crop yield prediction. In crop yield estimation, it is a priority to consider the impact of climate change and ecological differences on crop development.

Keywords: Crop yield prediction, Remote Sensing, Machine learning, NDVI, Wheat

Introduction

Wheat is among the most important basic nutrients in world food production (Shiferaw et al., 2013). While the demand for food products increases with the rapid population growth in the world, it is predicted that the rate of increase in global grain production will slow down (Fischer et al., 2014). The negative effects of climate change also make it difficult to meet the demand (Ray et al., 2019; Aydoğdu et. Al., 2020). Careful agricultural production planning and managing support policies are priority issues to ensure food safety. Accurate and timely estimation of wheat yield is very important considering national food security and its impact on a global scale (Han et al., 2020; Rahman, 2022). In this context, the management of wheat production, which is a basic ingredient in the nutrition of human populations, has strategic importance. Estimation of crop yield is important for determining and overcoming the negative effects of global warming on agricultural production due to climate change. Crop productivity is influenced by the interaction of Genotype, Environment and Management (G×E×M) (Cooper et al., 2021). In the estimation of crop yield; linear models, machine learning models and crop simulation models are applied. Remote Sensing technologies also play an important role in the estimation of efficiency. In order to increase the prediction accuracy, the model inputs, the accuracy of the data, as well as the spatial and temporal resolution of the data must match the scale of the data source. The Central Anatolian Region in Turkey is the main center where wheat production is widespread. While 52% of the agricultural areas in this region are cereal fields, approximately 60% of these areas are wheat fields (TUIK, 2021). The Central Anatolian Region has a semi-arid climate regime and is a region where dry agriculture is intense. In regions with semi-arid precipitation regime, temporal and spatial fluctuations occur in production and yield between years depending on precipitation (Teasdale and Cavigelli, 2017). This situation necessitates the timely monitoring and correct management of agricultural production from the very beginning.

Remote sensing techniques are used to monitor agricultural activities spatially and temporally and to make accurate forecasts (Atzberger, 2013). Satellite data can play an important role in obtaining information about crop type and crop development conditions from field scale to regional and national level. Crop yield is under the influence of many factors such as climate, soil, geographical location, variety characteristics, cultivation technique during the crop development period. Today, remote sensing-based models are preferred over traditional techniques in predicting the efficiency of these complex relationships. Vegetation indices (VI) derived from values obtained by reflection and absorption from visible and near-infrared (NIR) bands are widely used in monitoring crop growth and yield estimation by Remote Sensing (Huete et al., 2002).

Most of the studies in the literature indicate that the vegetation indices obtained from images with different spatial resolutions during crop development periods have been associated with yield and there have been some attempts to increase the estimation accuracy by using different modeling techniques (Rasmussen, 1997; Dempewolf et al., 2014; Johnson, 2016). Leaf area index (LAI) is an indicator of both the structure and development of vegetation. Therefore, it is possible to predict the wheat yield of a particular region with empirical models that correlate the maximum leaf area with the yield of NDVI (Kouadio et al., 2012). These empirical models are preferred because they require less data and are easy to apply on a regional scale. In the relationships established between VI and yield, different types of variables were used, such as periodic value (Lopresti, 2015), average value (Boken and Shaykewich, 2002; Mkhabela et al., 2011), cumulative value (Ren et al., 2008; Mashaba, 2017; Mirasi et al., 2019; Panek and Gozdowski, 2021) and maximum value (Becker-Reshef et al., 2010). It has been determined that seasonally integrated VIs can be predicted more accurately than a single period and have a high correlation with yield around the maximum time VI peaks (Guo et al., 2021; Ji et al., 2022).

Machine learning (ML), a branch of Artificial Intelligence, is a practical, data-driven approach that can provide better yield prediction based on various features by focusing on learning (Klompenburg et. al., 2020). Previous studies have indicated that ML techniques, along with linear approaches using differently calibrated models, can play an important role in crop yield estimation (Kaul et al. 2005; Ji et al. 2007; Jeong et al., 2016; Sayago and Bocco 2018; Cai et al. 2019; Han et al., 2020; Gomez et al., 2021). It can better explain the nonlinear relationships between ML algorithms and multiple data sources (Chlingaryan et al., 2018). Many sets of algorithms are available for prediction models (e.g. random forest, support vector regression, kernel machines and neural networks). Model performances generally increase when there is more available training data (Goodfellow et al., 2016). Low-resolution Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery is widely used due to its spatial, spectral and radiometric resolution for continuous monitoring of crop growth and for the creation of prediction models in large areas (regional, national scale). Although such images do not have high spatial resolution, it is possible to have cloud-free images due to the high repeating visit frequency of the sensor (1–2 days). On the other hand, low spatial resolution is not sufficient for estimations for producer fields with small field areas. High spatial resolution satellite images are used to overcome this problem on a field basis and to determine the difference in crop development within the field. Although the temporal resolution of Landsat images is low (16 days) in terms of spatial resolution (30 m), it allows yield estimations to be made at the scale of the producer field in determining the variability in crop development due to land characteristics and growing technique applications.

In this study, the creation of yield prediction models for winter wheat (Triticum aestivum L.) is aimed based on both the spatial resolution of satellite images and the field sizes of wheat production areas. In this context the relationships between yield and NDVI on the basis of crop development periods were investigated. LR, MLP, SVM and RF model performances for different field sizes were compared using NDVI data obtained from MODIS and Landsat satellite images.

Materials and Methods

Study Area

The study area is located in the Central Anatolian Region of Turkey. It covers the central and southern parts of Ankara province and the northern part of Konya province, which is between the Sakarya river in the west, the Kızılırmak river and the Salt Lake in the east (Figure 1). The elevation of the area varies between 610 m and 2065 m. There are 14 districts within the project area. The surface area of the project area is 22,900 km2. The study area has a semi-arid climate regime that typically characterizes the Central Anatolia Region, with an approximate annual average temperature of 12.0°C (MGM, 2021). While the annual average total precipitation varies between 330 and 390 mm, there are differences between years and seasonal distributions due to the scattered precipitation regime. The region is mainly composed of dry agricultural areas and grain cultivation is common. In grain fields, wheat and barley, rye and oat production are carried out. Other than grain, chickpeas, beans, lentils, sunflowers, safflower, sugar beet and corn are also produced throughout the region.

Remote Sensing Data

NDVI data, which is the main input of the crop yield estimation models, was obtained from MODIS and Landsat satellite images. NDVI, is derived from reflection in the red (R) and near-infrared bands (NIR) while representing the relationship of the product to dry matter deposition by monitoring vegetation growth conditions in large areas (Rouse et al., 1973; Mehta et al., 2021; Panchal et al., 2021). MODIS image has 250 m spatial resolution, 1-2 days temporal resolution and 36 bands (0.4 μm to 14.4 μm). MODIS data consists of 16-day composite images and is very useful for crop

monitoring in large areas by reducing the cloudiness effect. 288 images (MOD13Q1.h20v06) of the long-term time series (2001-2018) obtained from the MODIS TERRA platform (USGS, 2022) were used.
The Savitzky-Golay filter (Savitzky and Golay, 1964) was used to eliminate anomalies such as smearing errors and cloudiness that occur in the time series of MODIS images, and the TIMESAT program was used to apply the image correction technique (Jönsson and Eklundh, 2004). During 2001-2018, we used 240 images from Landsat-5 TM (7 bands, 0.45 to 12.5 μm), Landsat-7 ETM (8 bands, 0.45 to 12.5 μm) and Landsat-8 OLI & TIRS (11 bands, 0.43 to 12.51 μm) sensors with 177/33 and 178/33 World Reference System-2 (WRS-2) path/rows. Images are in the WGS 84 UTM Zone 36 coordinate system and the cloudiness is less than 30%.

Crop Yields

In yield estimation models, wheat yields obtained from large lands and producer fields were evaluated. Representing the large lands, within the study area, the dry farming yields of the fields belonging to the Polatlı, Altınova and Gözlü enterprises, which are affiliated to the General Directorate of Agricultural Enterprises (TIGEM) throughout the years 2001-2018 were used. TIGEM is a government institution that produces and distributes certified seeds as well as plant and animal production. The average field area of TİGEM enterprises is greater than 100 ha. The yield values of the producer fields were obtained from field studies in 2016, 2017 and 2018. Harvest data were obtained from agricultural areas where the fallow-sowing system was applied and from producer fields ranging from 1.4 ha to 25 ha. The yield averages of the fields from which the crop samples were taken, for the years 2016, 2017 and 2018, are 3.34 t/ha, 3.02 t/ha and 3.84 t/ha, respectively. Crop samples representing the sampled fields were taken from 1.0 m*0.5 m areas of three different regions. Sampling points were selected according to the random sampling method, considering the ecological regions created according to climate and topographic factors. In this context, sampling was carried out from 380 locations, 106 in 2016, 144 in 2017 and 130 in 2018. At the same time, the geographic coordinate locations of the sampling were recorded by GPS. Finally, after blending the crop material collected from the locations, the wheat grains were ±0.01 g. It was weighed with a precision electronic balance.

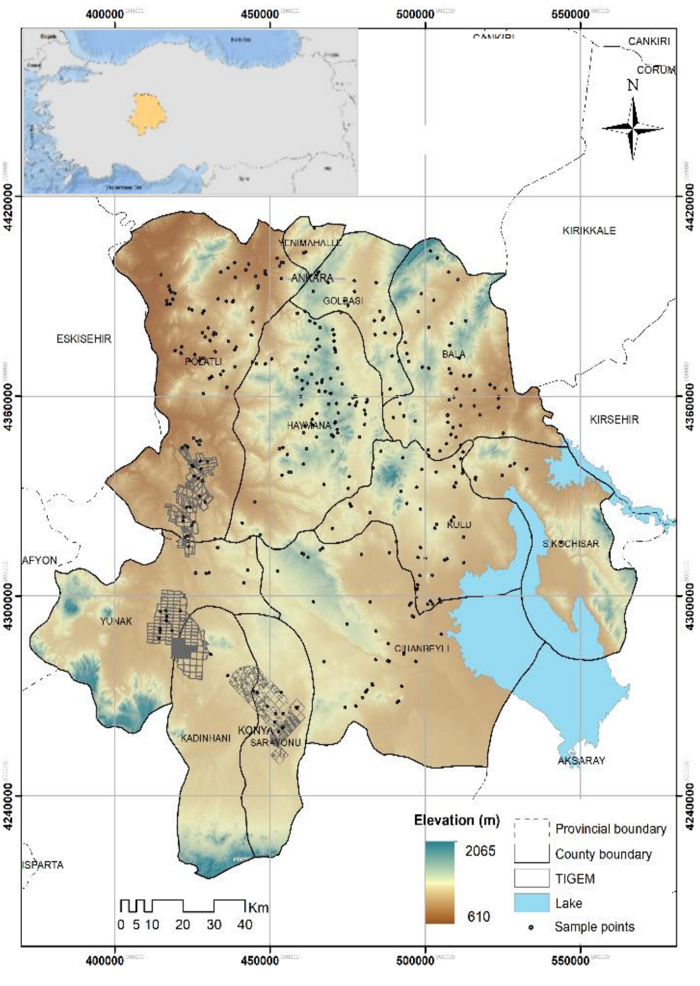

Fig.1 Location map of the study area.

Yield Estimation Models

The growing season for winter wheat starts in mid-October and continues until the end of July. Periodic vegetation indices (NDVI) were created first to determine the yield estimation model performances. Crop development varies over the years according to climatic conditions. In order to determine these differences, time series data were created and associated with the phenological periods of wheat. Since the effects of phenological periods on crop development are different, 12 vegetation periods were determined by comparing these periods with satellite image dates. Vegetation periods include 16 days of composite image. The beginning of each period is indicated as the day of the year (Table 1). These periods are; germination (DOY:289, 305), tillering (DOY:321, 337, 65), jointing (DOY:81.97), booting (DOY:113), heading/flowering (DOY:129, 145) and ripening (DOY:145,161).

Table 1- Periodic time intervals of satellite data

DOY	Date	DOY	Date
289	16 Oct - 31 Oct	81	22 Mar - 06 Apr
305	01 Nov -16 Nov	97	07 Apr - 22 Apr
321	17 Nov - 02 Dec	113	23 Apr - 08 May
337	03 Dec - 18 Dec	129	09 May - 24 May
353	19 Dec - 03 Jan	145	25 May - 09 Jun
65	06 Mar - 21 Mar	161	10 Jun - 25 Jun

Simple regression models were created to periodically determine the relationships between yield and NDVI. Depending on the phenological development of wheat, the input data were determined for the period when the maximum coverage and NDVI value were peaked, and prediction models were created accordingly. As model inputs; main period NDVI (NDVIp), maximum NDVI (NDVIm) and cumulative NDVI (NDVIc) values are used. Model performances were determined by creating linear and non-linear models with NDVI data and field yield values. In this context, as traditional approach, multiple regression model and ML models MLP, SVM and RF were applied.

Multiple Linear Regression (MLR)

Linear regression models are functional descriptions of the observed (dependent) value and influencing (independent) events. As being a traditional model, it has been extensively applied in yield estimation for many areas (Huang et al., 2013; Lopresti et al., 2015; Mashaba et al., 2016; Satir and Berberoglu, 2016). During the model creation, the correlation levels were determined by applying simple linear regression analysis between the dependent variable yield and VI as independent variables. For the models based on years, stepwise, which is a multiple regression model (MLR) for the main period NDVI, NDVIm and NDVIc values, was applied. In this model, the independent variable with the lowest significance level in the previous model was removed at each stage for the dependent variable, and the model with the highest significance level was obtained.

Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is a feed-forward neural network type using a supervised learning technique called backpropagation. Nonlinear relationships between the data within the corresponding datasets are processed functionally. The MLP model structure consists of input layers, hidden layers, and output layers, where each neuron is connected to all neurons in the next layer. The model is trained with the training data and the prediction performance is evaluated from the test data based on the error between the expected and determined output (Paudel et al., 2021). In order to use this model and data in the best way, a feature extraction process was performed first to find the particular features that best represent the problem in hand. The aim is to train the network based on the dataset using these features with the backpropagation learning algorithm. In order to achieve this, an iterative structure was followed. The model was trained using the training data on different network topologies, and then the performance of the predicted model was measured on the test data. In order to prevent overfitting k-fold Cross Validation (CV) is implemented.

Random Forest (RF)

Random Forest is a learning technique that combines a set of decision trees for classification or regression (Breiman, 2001). The RF is constructed by selecting each tree, a random set of variables, and samples of the dataset. It works by constructing many decision trees and extracting predictions by combining a set of baseline decisions in regression models. During the training process, each decision tree is trained using some random partition of the data. However, each decision tree also uses a random subset of the full feature set. As a result, each decision tree is different and their overall contribution in the decision process results in better generalization without explicit need for Cross Validation.

Support Vector Machine (SVM)

A support vector machine (SVM) is a machine learning algorithm that can be used for both binary classification and regression. SVM is a supervised non-parametric algorithm that partitions the data using kernels and separating the margins between the two as much as possible (Gunn, 1998). Kernel functions, along with other SVM hyperparameters are optimized such that the resulting model achieves the highest margin between the two classes. As a result, SVM is also called maximum margin classifier. During SVM regression, the input is mapped to a high-dimensional feature space using a kernel function and a linear regression model is created to minimize errors (Hearst et al., 1998). In this study, the Gaussian kernel function is used. Cross Validation also is not necessary in the SVM case, since it optimizes the margin between the different classes; hence it can overfit.

Model Validation

In the forecasting model, it is aimed to make yield predictions based on the crop development period. Yield estimation model outputs were produced at 250 m and 30 m spatial resolutions. Training and testing procedures were applied for LR, SVM, MLP and RF methods. Estimation results were compared to determine model performances. Each model was trained using 2 years of data and the remaining year was spared for testing. This was implemented for every year such that all the available data was tested by a 3-fold-cross-validation method. Model performances were evaluated through coefficient of determination (R^2), Root Mean Square Eror (RMSE) and *Mean Absolute Percentage Error* (*MAPE*) accuracy metrics.

$$R^2 = \frac{\left(\sum_{i=1}^{n} (y_i - \bar{y})(f_i - \bar{f})\right)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (f_i - \bar{f})^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2}$$

$$MAPE = 100 \times \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - f_i}{y_i}\right|$$

Here, n is the number of observations, y_i is the observed yield, f_i is the estimated yield, $\bar{y}$ and $\bar{f}$ are the mean values. The closer the R^2 value is to 1, the higher the prediction performance of the model. Smaller RMSE and MAPE values indicate less discrepancy between observed and predicted yield.

Results

Relationships between NDVI and yield during CDP

In this study, ndvi obtained from modis and landsat images was used as vegetation index. Relationships between wheat yield and NDVI during the CDP were monitored for TIGEM and producer fields. The effect of the phenological period was determined for the yield estimation model.

MODIS-NDVI and yield in TIGEM fields

During the development period, simple linear models were constructed between the average field NDVI data and the field yields for the TIGEM lands for 2001-2018.

When the vegetation indices are examined periodically with MODIS images, it is observed that while R^2 was 0.14 in the 321st period of crop emergence after planting, it increased to 0.28 with the onset of tillering. In the 65th period, for the crop R^2 was 0.31 and in the 113th period, the relationship ratio increased with the crop's booting period and R^2 rose to 0.48. Furthermore, in the 129th period which is the flowering period, R^2 reached the peak value of 0.63. The high correlation (R^2:0.54) continued in the 145th period and the index value decreased (R^2:0.34) as the crop entered the maturation period with the 161st period (Figure 2).

MODIS-NDVI and yield in producer fields

In field studies, the relationships between NDVI data obtained from MODIS and Landsat 8 satellite images and yield were determined during the CDP. Wheat yield values obtained from crop samples collected in field studies in 2016, 2017 and 2018 were used.

When the vegetation indices obtained from MODIS images are examined seasonally, it is observed that while there was an initial relationship of R^2:0.15 in the post-germination period of the crop, then it became R^2:0.32 in March during the vegetative development period. Then, it reached at the levels of R^2:0.43 result of the acceleration of crop growth with the booting period. It increased to R^2:0.50 during the flowering period (Figure 3). Due to the spatial resolution of the MODIS satellite image, the correlation levels decreased significantly for small fields.

There are several studies in the literature for wheat yield estimation using linear modeling through MODIS satellite imagery. Mashaba et al. (2017) found a high correlation (R^2: 0.73) between yield and NDVI during flowering, while Lopresti et al. (2015) North of Buenos Aires province, Argentina showed a good correlation (R^2:0.52) between wheat yield and MODIS-NDVI 1 month before harvest during flowering period. A similar result was given in a national study in the USA in the relationship between NDVI data obtained from 250 m spatial and 16 day temporal resolution MODIS images and yield (R^2: 0.51) for the period between the end of May and the beginning of June, when the vegetation reached its peak level (Jahnson, 2016). Panek and Gozdowski (2020) determined the varying correlation levels between yield and NDVI (R^2: 0.49- 0.71) in Central Europe during flowering and grain-filling periods. In the study conducted in Canada, (Mkhabela et al., 2011) a high correlation (R^2: 0.47-0.80) was determined between the yield of flowering and grain-filling periods covering July and early August for Wheat in semi-arid climate conditions and NDVI. In our study, the relationship of NDVI data obtained from MODIS satellite image with 250 m spatial resolution and 16-day temporal resolution with yield during flowering and grain-filling period was found to be R^2:0.49-0.62 in TIGEM fields and R^2:0.38-0.42 in producer fields. The relationships between the developmental periods of wheat and the yield were statistically significant ($p < 0.0001$).

Landsat-NDVI and yield in TIGEM fields

In the periodic monitoring of crop growth, regression models were created between the average field NDVI data of 2001-2018 for TIGEM fields and the average NDVI data obtained from LANDSAT 5, 7 and 8 images. When the vegetation indices were examined periodically, it is observed that while R^2 was 0.14 in the 321st period after planting, it increased to 0.22 with the onset of tillering. While R^2 was at the levels of 0.26 in the 65th period after the winter period, and 0.31 in the 97th period, it increased to the level of 0.51 in the 113th period with the stem lift period of the crop, and reached the level of 0.61, peaking in the 129th period, which is the flowering period of the crop. The high correlation (R^2:0.50) continued in the 145th period and decreased (R^2:0.29) rapidly at 6 months as the crop entered the maturation period in the 161st period (Figure 4).

Landsat-NDVI and yield in producer fields

Wheat yields collected in field studies and NDVI data obtained from LANDSAT 8 satellite images were compared during the CDP. While the initial relationship during the germination period of the crop indicated an R^2 of 0.10, it reached the levels of R^2:0.27 during the beginning of the vegetative period in March (65th semester). R^2 became 0.43 with the acceleration of crop growth during the stemming period. During the flowering period, these values reached their best levels with R^2:0.65. However, it was observed that the flowering period shifted to the 145th period and R^2 decreased to 0.26 in areas with high altitudes in the study area (Figure 5). In the study, the effect of land size and ecological differences was observed depending on the resolution of the image. It was observed that the correlation levels between yield and NDVI decreased significantly in the producer lands when the product passed from the flowering period to the maturation period, while high correlation levels continued in the TIGEM lands with similar ecology.

It has been shown that the temporal period when the NDVI value peaks from a Landsat image from March to June stands out as the primary parameter for evaluating wheat yield and is strongly correlated with yield. In similar studies with Landsat, strong correlations were observed between yield and NDVI. According to Jelínek et al. (2020), in field studies, a relationship at the level of R^2:0.64 was determined between Landsat 8 NDVI data and wheat yield. . In the study conducted in the Tisza basin in Hungary (Nagy et al., 2021), the strongest relationship (R^2: 0.64) was observed between the NDVI data obtained from Landsat 8 images and the yield for the May-early June period when the wheat reached its peak level.

NDVI values reach the highest value during the flowering period in connection with the increase in the chlorophyll content in the leaves during the wheat development periods. As the plant enters the maturation period (DOY 161), the NDVI value decreases with the decrease in moisture content due to the yellowing of the leaves, and a weak relationship with yield is observed. In the study area, the 129th and 145th periods have the highest coefficient of determination among the forecast models, reaching the maximum biomass value by the end of May, indicating that it is the best time to make a forecast. These periods include the flowering and grain filling periods within the phenological period of the crop. These results are consistent with studies reporting a high correlation between NDVI and yield during flowering and grain filling (Mkhabela et al., 2011, Mashaba et al., 2017, Panek and Gozdowski, 2020). The most suitable period for wheat yield estimation in the study area is flowering and grain filling periods, which corresponds to 4-6 weeks before harvest. The relationships between the developmental periods of wheat and the yield were statistically significant ($p < 0.0001$).

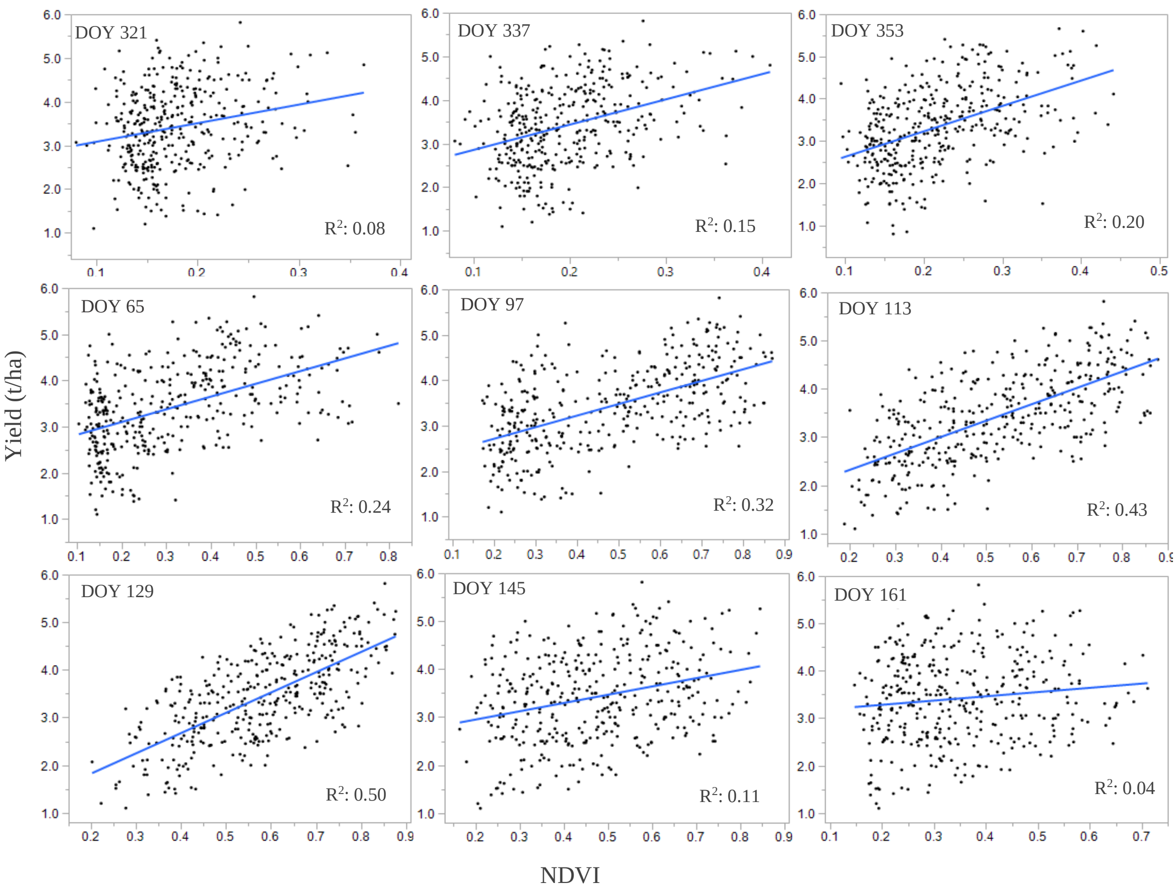

Fig. 2 Seasonal relationships between MODIS-NDVI and yield in TIGEM fields

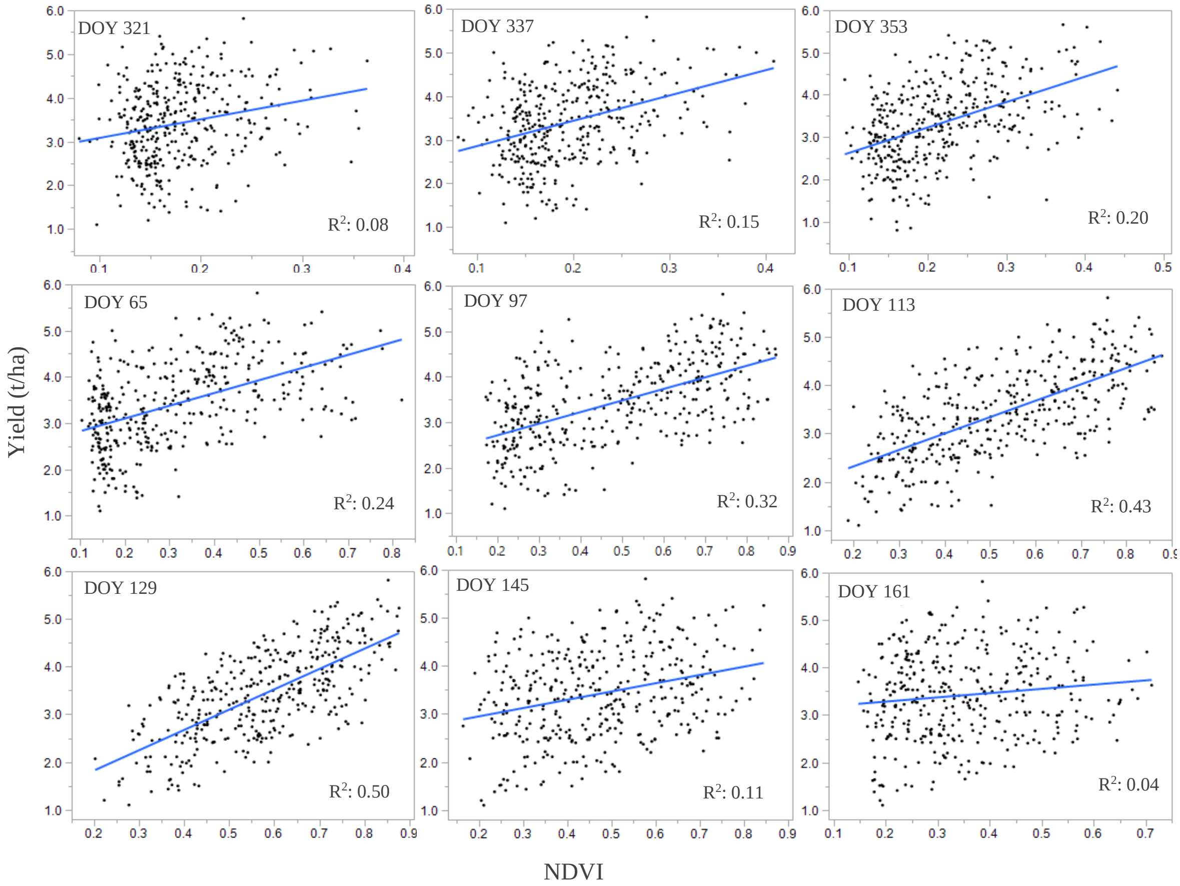

Fig. 3 Seasonal relationships between MODIS-NDVI and yield in producer fields.

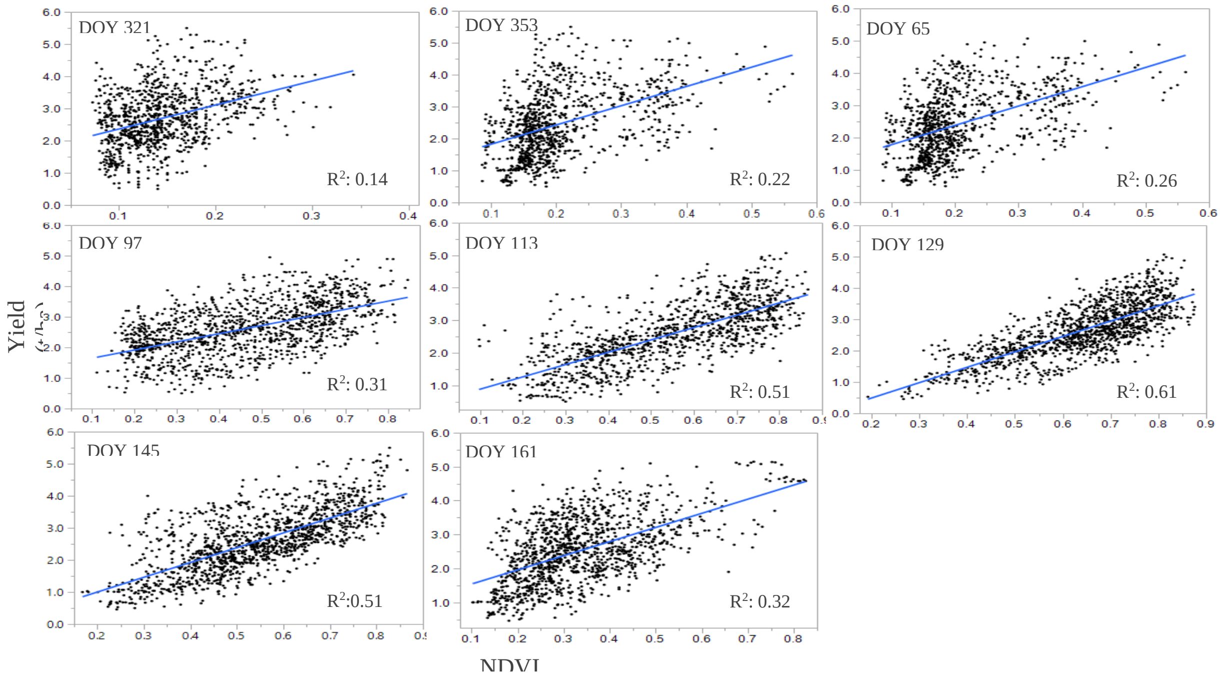

Fig. 4 Seasonal relationships between Landsat 8-NDVI and yield in TIGEM fields.

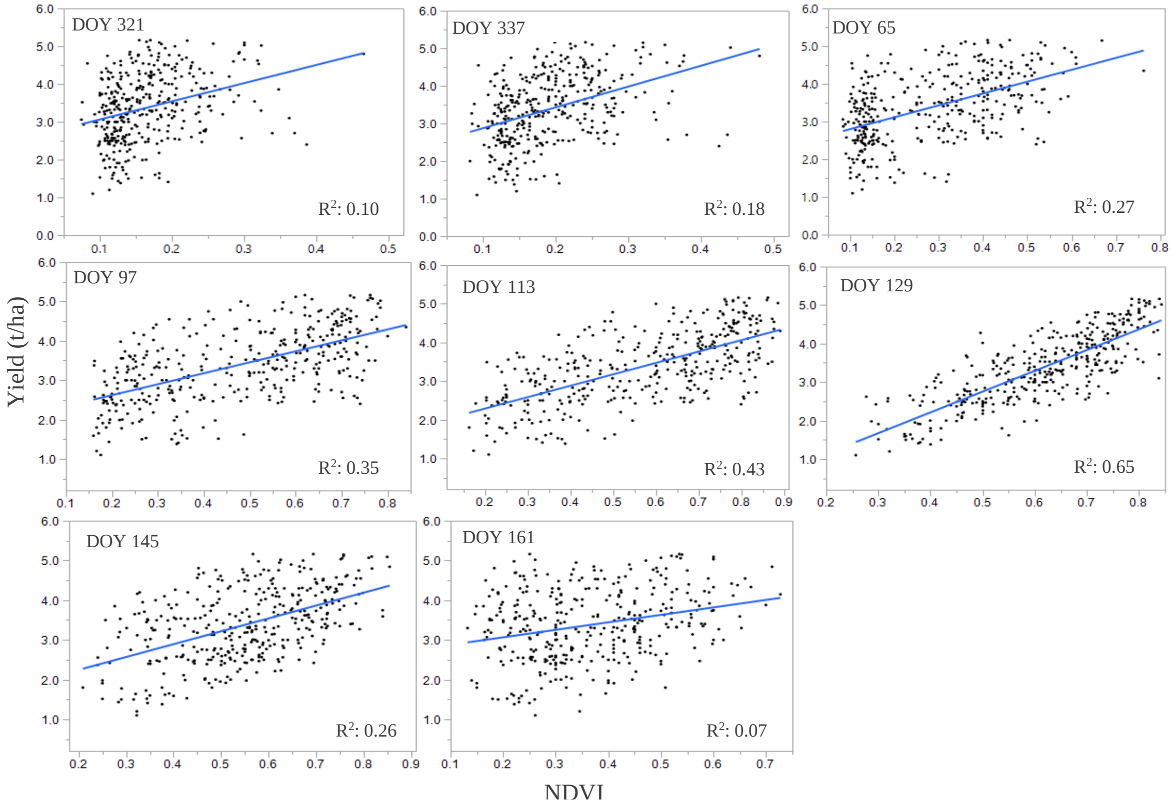

Fig. 5 Seasonal relationships between Landsat 8-NDVI and yield in producer fields.

Comparison of model performances

Model estimation performances were compared for fields with different areal sizes in rainfed agricultural areas where wheat is planted. In this context, NDVIp, NDVIm and NDVIc data obtained from Modis and Landsat satellite images were used as model inputs in large fields (TIGEM) and producer fields. NDVIp, NDVIm and NDVIc values between the end of tillering period and the milk production period were used as model variables within the development periods of wheat. Stepwise multiple regression model (MLR) was applied as linear model whereas MLP, RF and SVM techniques were applied as ML models. Forecasts were made at the end of May and beginning of June, four weeks before the harvest period.

Yield prediction with ML models in TIGEM fields

Efficiency values obtained from TIGEM fields and MLR, MLP, RF and SVM model performances for MODIS and Landsat NDVI data between 2001-2018 were compared. Yield estimation was made for each year and the data of the estimated year were not included in the model. Among the models created with the NDVI data obtained from the MODIS satellite image, the best estimation was made with the MLP model. The estimation performance of the MLP model over the years ranged from RMSE between 0.23-0.65 t/ha. The MLP model was followed by the SVM (RMSE of 0.29-0.81 t/ha), MLR (RMSE of 0.33-0.77 t/ha) and RF (RMSE of 0.39-0.84 t/ha) models (Figure 6a). In a similar fashion, among the models created with the NDVI data obtained from the Landsat satellite image, the best estimation was also made with the MLP model. The estimation performance made with the MLP model over the years ranged from RMSE between 0.28-0.64 t/ha. The MLP model was followed by the SVM (RMSE: 0.30-0.95 t/ha), MLR (RMSE: 0.33-0.95 t/ha) and RF (RMSE: 0.46-0.89 t/ha) models.

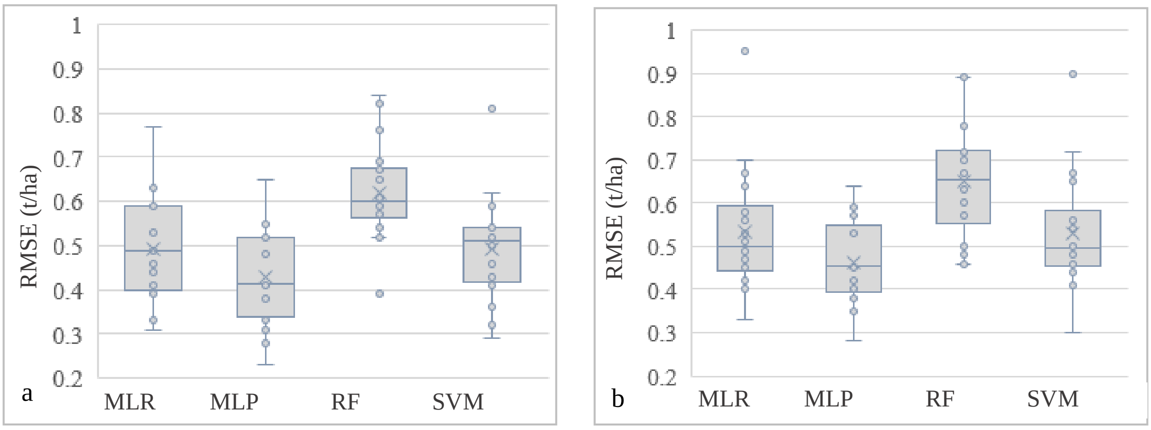

Fig. 6 Variation of model performances (RMSE, t/ha) for MODIS (a) and Landsat (b) data in TIGEM fields.

Temporal variation of model accuracies in TIGEM fields

When the NDVI time series are monitored on a yearly basis, it is seen that the RMSE value for MODIS data fell below 0.30 t/ha and reached 0.80 t/ha due to the extreme climatic conditions. While Landsat data decreased to around 0.30 t/ha, it exceeded 0.90 t/ha in years when the error value increased (Figure 7a). According to the long term average error values; RMSE values obtained with MODIS data for MLP, SVM, MLR and RF are 0.43 t/ha, 0.49 t/ha, 0.50 t/ha, 0.62 t/ha, respectively. Meanwhile, for Landsat data, RMSE values for MLP, SVM, MLR and RF are 0.46 t/ha, 0.52 t/ha, 0.54 t/ha, 0.65 t/ha, respectively (Figure 7b).

MLP and SVM models showed the best performance in both data sets. A better estimation was made with MODIS data compared against the Landsat data. The fact that MODIS has a high temporal resolution has a positive effect, while the temporal resolution of Landsat data is 16 days and cloudiness etc. during the critical periods of product development affected the results. Hence, the error values have increased in some of the years for these reasons. The amount of precipitation that the crop receives during the growing period is important in model performances. In the forecast years, 2002 and 2007 were dry years and 2010 and 2015 were humid years. It is observed that the RMSE value has increased (> 0.5 t/ha) in the semi-arid climatic conditions of the region during these years. It can be said that rather than the total seasonal precipitation, the extremely dry or rainy period during the April-June period, which is critical in the growing period of the crop, increases the error.

Within the scope of the study, the error values obtained from the yield estimation model results with MLR overlap with similar studies for MODIS (RMSE: 0.50 t/ha) and Landsat (RMSE: 0.54 t/ha) data. Becker-Reshef's et al. (2010) study in Ukraine, the model accuracy for MODIS was found to be RMSE: 0.44 t/ha. Kouadio et al. (2012) in their study using MODIS data, estimated the model performance at regional scale as RMSE: 0.57 t/ha. Lyle et al. (2013) in a study using Landsat 5 and 7 images in large farm areas in Australia, it was found that there was a good correlation between NDVI and yield, and the average RMSE for the models was 0.58 t/ha. Skakun et al. (2017) In regional scale yield estimation, ndvi data obtained from Landsat-8 satellite image and official statistical yield values were compared and model accuracy (RMSE 0.57 t/ha) was determined.

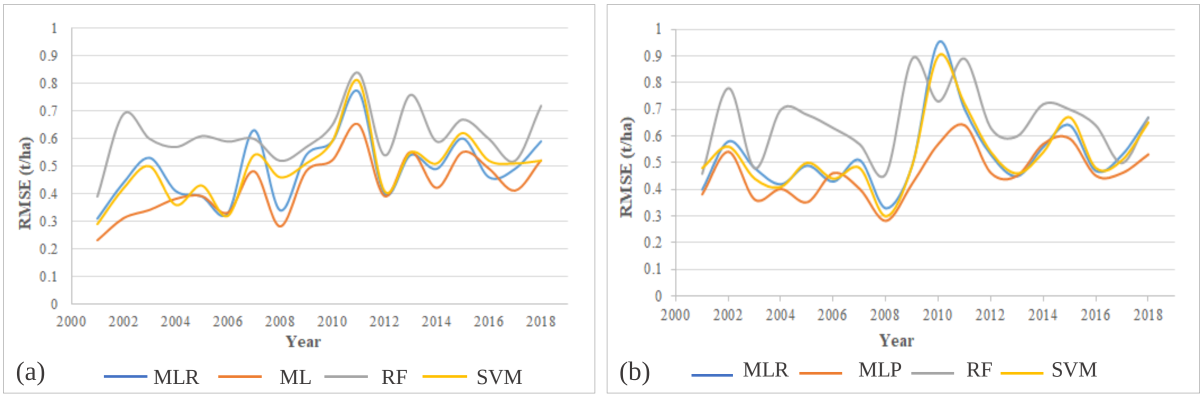

Fig. 7 Variation of error value according to MODIS (a) and Landsat (b) time series data.

Yield estimation with ML in producer fields

The performances of MLR, MLP, RF and SVM models were compared for yield estimations for producer fields. NDVIp, NDVIm and NDVIc data were used as model input data. Model estimation results were made four weeks before the harvest date. The SVM model showed the highest performance among the models using MODIS data. With SVM, the best estimation was in 2017 (RMSE:0.74 t/ha, MAPE: 20.3%), while 2016 (RMSE:0.78 t/ha, MAPE: 21.4%) and 2018 (RMSE:0.80 t/ha, MAPE: 16.2%) have followed. With the MLP model, the best estimation was achieved in 2017 (RMSE: 0.80 t/ha, MAPE: 21.8%), while the worst estimation was for 2018 (RMSE: 0.90 t/ha, MAPE: 19%). The traditional MLR model is comparable in 2017 (RMSE:0.82 t/ha, MAPE: 25.3%), 2016 (RMSE:0.85 t/ha, MAPE: 21.8%) and 2018 (RMSE:0.88 t/ha, MAPE: 18.2%). For MLR, similar error values were obtained for each year. The RF model results have higher error (RMSE > 0.95 t/ha) values than other models (Table 2).

A remarkable improvement was seen in the model results using Landsat 8 data. Among the models, the SVM model showed the highest performance. While the best estimation with SVM was made in 2017 (RMSE: 0.51 t/ha, MAPE: 13.9 %), they were followed by 2016 (RMSE: 0.53 t/ha, MAPE: 13.1 %) and 2018 (RMSE: 0.60 t/ha, MAPE: 13.1 %) respectively. The best estimate after SVM was achieved with MLP. The best estimation for the MLP model is in 2017 (RMSE:0.55 t/ha, MAPE:15.5 %). Meanwhile, 2016 (RMSE:0.58 t/ha, MAPE:15.1 %) and 2018 (RMSE:0.63t/ha, MAPE:12.7 %) error values slightly increased. MLR model best estimates are respectively 2017 (RMSE:0.58 t/ha, MAPE:14.9 %), 2016 (RMSE:0.60 t/ha, MAPE: 15.3 %) and 2018 (RMSE:0.65 t/ha, MAPE: 13.5%) occurred over the years. Slightly higher error values were obtained by the RF model compared to other models in 2016 (RMSE:0.69 t/ha, MAPE:16.5 %), 2017 (RMSE:0.66 t/ha, MAPE:18.1 %) and 2018 (RMSE:0.71 t/ha, MAPE:15.2%) (Table 2).

Table 2 Forecast model performances (RMSE_t/ha, MAPE_%) by years for MODIS and Landsat 8 in producer fields

		MLR		MLP		RF		SVM	
	Year	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
MODIS	2018	0.88	18.2	0.90	19.0	0.97	21.3	0.80	16.2
	2017	0.82	25.3	0.80	21.8	1.00	28.3	0.74	20.3
	2016	0.85	21.8	0.83	22.1	0.95	26.2	0.78	21.4
	Average	**0.85**	**21.8**	**0.84**	**21.0**	**0.97**	**25.3**	**0.77**	**19.3**
Landsat 8	2018	0.65	13.5	0.63	12.7	0.71	15.2	0.60	13.1
	2017	0.58	14.9	0.55	15.5	0.66	18.1	0.51	13.9
	2016	0.60	15.3	0.58	15.1	0.69	16.5	0.53	13.1
	Average	**0.61**	**14.6**	**0.59**	**14.4**	**0.69**	**16.6**	**0.55**	**13.4**

In the modeling made using MODIS data, the most successful prediction was made with the SVM model. Model estimation error with SVM was found in the range of RMSE: 0.74-0.80 t/ha (avg.:0.77 t/ha). Other models followed with MLP (RMSE: 0.80-0.90 t/ha, avg.:0.84 t/ha), MLR (RMSE: 0.82-0.88 t/ha, avg.:0.85 t/ha) and RF (RMSE: 0.95-1.0 t/ha, avg.:0.97 t/ha). In the models made for Landsat 8 data, the best prediction performance was obtained with SVM (RMSE: 0.51-0.60 t/ha, avg.:0.55 t/ha). The SVM model was followed by the MLP (RMSE: 0.55-0.63 t/ha, avg.:0.59 t/ha), MLR (RMSE: 0.58-0.65 t/ha, avg.:0.61 t/ha) and RF (RMSE: 0.66-0.71 t/ha, avg.:0.69 t/ha) models.

When the model performances of MODIS and Landsat 8 satellites in producer fields were compared for 2016, 2017 and 2018, it was determined that SVM technique made the best estimation compared to MLP, RF and MLR models. In some studies comparing linear and ML models, it was stated that ML model performances were better (Jeong, 2016; Chen and Jing, 2017; Cai et al,, 2019; Guo et al., 2021). It is observed that different models stand out in accuracy in studies conducted in different geographical locations using ML techniques. According to Gomez et al. (2021), among the ML techniques applied in the study conducted in Mexico, it was determined that the RF model produced results with the highest accuracy (RMSE: 0.78-0.83 tons/ha). Wang et al. (2020) compared linear and nonlinear models for yield estimation for winter wheat in the USA. Among the ML techniques, Adaptive Boosting (R^2: 0.86, RMSE:0.51 t/ha) showed the best performance. ML models were able to predict better than linear models. In Kamir et al. (2020) study, wheat yield in Australia was estimated using ML models. Among the ML models, the SVM (R^2:0.86, RMSE:0.51 t/ha) model showed the best prediction accuracy.

In the SVM model, where the best estimation is made, the RMSE values obtained for MODIS and Landsat 8 are 0.80 t/ha and 0.60 t/ha for 2018, 0.74 t/ha and 0.51 t/ha for 2017 and 0.78 t/ha and 0.53 t/ha for 2016 respectively. The error value obtained with MODIS was approximately 1.4 times higher than the Landsat 8 data in producer fields (Figure 8).

Similar results are observed for the MLP, RF and MLR models. In the study based on crop development periods, a good prediction accuracy was obtained with the SVM model. Obtaining results at similar levels in previous studies supported the argument that the SVM model is one of the appropriate methods for estimating crop yield (Joshi et al., 2021; Ju et al. 2021; Abebe et al., 2022; Li et al., 2022).

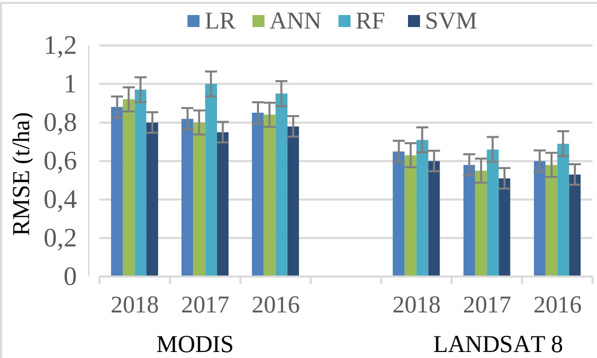

Fig. 8 Model performances by years in producer lands

Discussion and Conclusion

Yield estimation was implemented for TIGEM and producer fields using NDVI data obtained from MODIS and Landsat satellite images using LR, MLP, SVM and RF modeling techniques. It is aimed that the applied model best predicts the yield value and the error rate is at the lowest level. Model input data, yield values, and extra climatic conditions depending on the phenological period affect the prediction model accuracy. The spatial and temporal resolution of satellite images, which are the main data source, overlap with the data source, which will increase the performance of the applied model. In this context, model performance behaviors were investigated with data obtained from different field sizes.

When the empirical relationship between the satellite-derived vegetation index and historical yield values is modeled with statistical regression-based approaches, the model is localized and cannot be easily generalized to other different areas. The main disadvantage of empirical models estimating yield from spectral data is the difficulty of extending locally calibrated estimation methods to other areas or other scales, as it is linked to the environmental characteristics of the geographic area (Becker-Reshef et al., 2010; Vannoppen and Gobin, 2021). Effective performance can be achieved in areas with similar climate and ecological structure. As crop development varies over time, correlating crop phenology with vegetation indices can provide a significant benefit for crop yield prediction models. Thus, considering crop phenology in yield estimation can increase the accuracy of the estimation.

During the crop development period, the seasonal NDVI value and the cumulative NDVI value obtained at maximum crop coverage were highly and linearly correlated. With the acquisition of new NDVI data every month, forecast performances were improved, reaching the highest level when all data were collected. Therefore, a more accurate estimation of its yield is possible approximately 4-6 weeks before harvest.

The differences between the years in the climate data, especially the irregularity and variability in the precipitation, also affect the estimation results. Ecological differences and the tolerance of the crop variety increase the variability in yield during the periods when the crop is under water stress. Ecological differences cause temporal shifts in the phenological stages of the crop in this case it affects the model performance. The effect of the climatic conditions occurring during the critical development periods on the harvest yield of the crop is also quite large. The fact that the models created can explain this change in vegetation will cause a decrease in the error rate. The model that can best explain the positive and negative aspects of the season will show the best performance. Estimation results can be improved by using effective climate parameters for the studied area as well as vegetation indices obtained from satellite images.

Among the factors affecting model performances are mixing of field sizes and product reflectance values, historical yield values and accuracy. Our most important advantage in this study was that we have wheat planted field borders. Thus, the interference of satellite image reflection values with other products is prevented to a certain extent. This situation creates negativity in reflectance values due to field interactions in low resolution images, especially in regions with small field area. In the creation of NDVI time series obtained periodically, MODIS satellite image provides an advantage in overcoming problems such as cloudiness compared to Landsat satellite image due to its high temporal resolution, while it has a disadvantage due to low spatial resolution in producer fields.

Another important issue is the accuracy of the yield values. If the efficiency values are not healthy, this will increase the error rate and decrease the model performance. In this study, we used measured field yields to reduce this error as much as possible. It is important that the spatial resolution of the satellite image is high so that the data obtained on the basis of location in the producer fields can represent the field. In medium and low resolution images such as MODIS, the prediction accuracy results are reduced due to the mixing of product reflection values. In the results we obtained in the producer fields, we saw that the MODIS (250 m) data made an estimation with approximately 1.4 times more error than the Landsat 8 (30 m) data.

In large fields, the models created with MODIS and Landsat have similar results. Although the image resolutions are different, due to the field sizes of more than 100 ha, it has provided sufficient performance for medium resolution images. The use of 30 m resolution data obtained from the Landsat satellite image in the producer's fields has had a positive effect by increasing the performance of the model. Yield estimation models can be developed by using sensors with higher repetition frequency and better spatial resolution (10 m) for more accurate estimation in small field structures such as producer fields.

The best estimation was made with the SVM model among the model performances in the producer's fields. This has been confirmed over the years and for different spatial resolutions. ML models have shown that it can play an important role in crop yield prediction. With the results obtained, it will be possible to predict the wheat yield early, cost-effectively and more accurately and to reveal the yield variability in terms of area.

Acknowledgements

This study was supported by TUBITAK (Scientific and Technological Research Council of Turkey) 3001 program with project number 1150850.

References

Abebe, G., Tadesse, T., Gessesse, B. (2022). Combined Use of Landsat 8 and Sentinel 2A Imagery for Improved Sugarcane Yield Estimation in Wonji-Shoa, Ethiopia. *Journal of the Indian Society of Remote Sensing*, 50(1):143–157.

Atzberger, C. (2013). Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. *Remote Sens.* 5, 949–981.

Aydoğdu, M., Yıldız, H., Ünal, E., Külen, S. (2020). Evaluating Hyperspectral Vegetation Indices for Estimating Nitrogen Concentration of Winter Wheat in Different Growth Stages, *International Journal of Environment and Geoinformatics*, 7(3), 325-334. doi.10.30897/ijegeo.673038

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, 1312–1323.

Boken, V. K., Shaykewich, C. F. (2002). Improving an operational wheat yield model for the Canadian Prairies using phenological-stage-based normalized difference vegetation index, *International Journal of Remote Sensing*, 23 (20):4157-4170.

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. https://doi.org/10.1023/A:1010933404324

Cai, Y., Guan, K., Lobell, D., Potgieter, A. B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., & Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorology,* 274, 144–159.

Chen, P., Jing, Q. (2017). A comparison of two adaptive multivariate analysis methods (PLSR and MLP) for winter wheat yield forecasting using Landsat-8 OLI images. ScienceDirect, 59, 987–995.

Chlingaryan, A., Sukkarieh, S., Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agric.* 151, 61–69.

Cooper, M., Voss-Fels, K. P., Messina, C., Tang, T., Hammer, G. L. (2021). Tackling G×E×M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. Theoretical and Applied Genetics (2021)134:1625–1644.

Dempewolf, J., Adusei, B., Becker, I., Hansen, M., Potapov, P., Khan, A., Barker, B. (2014). Wheat yield forecasting for Punjab Province from vegetation index time series and historic crop statistics. *Remote Sens.* 6 (10):9653–9675.

Fischer, R. A., Byerlee, D., Edmeades, G. O. (2014). Crop Yields and Global Food Security: Will Yield Increase Continue to Feed the World; *Australian Centre for International Agricultural Research:* Canberra, Australia, 158, 634.

Gomez, D., Salvador, P., Sanz, J., Casanova, J. L. (2021). Modelling wheat yield with antecedent information, satellite and climate data using machine learning methods in Mexico. *Agricultural and Forest Meteorology*.doi.org/10.1016/j.agrformet.2020.108317

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning.

Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS Tech. Rep.,* 8, 14, 5–16.

Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., Bryant, C. R., Senthilnath, J. (2021). Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecol. Ind.,* 120, 106935.

Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., Zhang, J. (2020). Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sens.* 2020, 12, 236. doi:10.3390/rs12020236

Hearst, M. A., Dumais, S .T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intell. Syst. Appl.,* 13, 18–28.

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). doi:10.1371/journal.pone.0070816

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 Sens. Environ,* 83, 195–213.

Jelínek, Z., Kumhálová, J., Chyba, J., Wohlmuthová, M., Madaras, M., Kumhála, F. (2020). Landsat and Sentinel-2 images as a tool for the effective estimation of winter and spring cultivar growth and yield prediction in the Czech Republic. *Int. Agrophys.*, 34, 391-406 doi:10.31545/intagr/126593

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., Kim, S. (2016). Random forests for global and regional crop yield predictions. *PLoS One* 11, e0156571, doi.org/10.1371/journal.pone.0156571

Ji, B., Sun, Y., Yang, S., Wan, J. (2007). Artificial neural networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science,* 145, 249–261.

Ji, Z., Pan, Y., Zhu, X., Zhang, D., Wang, J. (2022). A generalized model to predict large-scale crop yields integrating satellite-based vegetation index time series and phenology metrics. *Ecological Indicators,* 137, 108759.

Johnson, D. M. (2016). A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products. *Intern J Appl Earth Obs Geoinform* 52: 65-81.

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., Bédard, F. (2016). *Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods*, 218-219 74-84.

Jönsson, P., Eklundh, L. (2004). TIMESAT-A program for analyzing time-series of satellite sensor data. *Computers and Geoscience*, 30, 833−845.

Joshi, V. R., Kazula, M. J., Coulter, J. A., Naeve, S. L., Garcia, A. G. (2021). In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. International Journal of Biometeorology, 65:489–502, doi.org/10.1007/s00484-020-02039-z

Ju, S., Lim, H., Won Ma, J., Kim, S., Lee, K., Zhao, S., Heo, J. (2021). Optimal county-level crop yield prediction using MODIS-based variables and weather data: A comparative study on machine learning models. *Agricultural and Forest Meteorology* 307, 108530.

Kamir, E., Waldner, F., Hochman, Z. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J. Photogramm. Remote Sens.* 160, 124–135.

Kaul, M., Hill, R. L., Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems,* 85, 1–18.

Klompenburg, T. V., Kassahun, A., Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture,* 177, 105709.

Kouadio, L., Duveiller, G., Djaby, B., El Jarroudi, M., Defourny, P., Tychon, B. (2012). Estimating regional wheat yield from the shape of decreasing curves of green area index temporal profiles retrieved from MODIS data, *International Journal of Applied Earth Observation and Geoinformation,* 18, 111–118.

Li, Z., Chen, Z., Cheng, Q., Duan, F., Sui, R., Huang, X., Xu, H. (2022). UAV-Based Hyperspectral and Ensemble Machine Learning for Predicting Yield in Winter Wheat. *Agronomy* 2022, 12, 202. doi.org/10.3390/agronomy12010202

Lopresti, M. F., Di Bella, C. M., Degioanni, A. (2015). Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province. Argentina, Information *Processing In Agriculture,* 73–84.

Lyle, G., Lewis, M., Ostendorf, B. (2013). Testing the Temporal Ability of Landsat Imagery and Precision Agriculture Technology to Provide High Resolution Historical Estimates of Wheat Yield at the Farm Scale. *Remote Sens.*, *5*, 1549-1567.

Mashaba, Z., Chirima, G., Botai, J. O., Combrinck, L., Munghemezulu, C., Dube, E. (2017). Forecasting winter wheat yields using MODIS NDVI data for the Central Free State region. South African Journal of Science, 113 (11/12). dx.doi.org/10.17159/ sajs.2017/20160201

Mehta, D. , Shukla, S. H., Kalubarme, M. H. (2021). Winter Crop Growth Monitoring using Multi-Temporal NDVI Profiles in Kapadvanj Taluka, Gujarat State. *International Journal of Environment and Geoinformatics,* 8(1), 33-38, doi.10.30897/ijegeo.773860

MGM (2021). *General Directorate of Meteorology.*

Mirasi, A., Mahmoudi, A., Navid, H., Kamran, K. V., Asoodar, M. A. (2019). Evaluation of sum-NDVI values to estimate wheat grain yields using multi-temporal Landsat OLI data. *Geocarto International,* doi.org/10.1080/10106049.2019.1641561

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.

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, 652.

Panchal, J., Shukla, S. H., Kalubarme, M. (2021). Analysis of Optimum Growth Stages for Winter Crop Separability using Multi-Temporal NDVI Profiles in Vijapur Taluka, Gujarat State. *International Journal of Environment and Geoinformatics,* 8(2), 135-143, doi.10.30897/ijegeo. 803303.

Panek, E., Gozdowski, D. (2021). Relationship between MODIS Derived NDVI and Yield of Cereals for Selected European Countries. *Agronomy*, 11, 340.

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems.* Volume 187.

Rahman, H. (2022). Monitoring Directional Dynamics of Growing Wheat Crop Canopy Using Ground based Time Series Remote Sensing Radiative Measurements, *International Journal of Environment and Geoinformatics,* 9(1), 25-39. doi.10.30897/ijegeo.877226

Rasmussen, M. S. (1997). Operational yield forecast using AVHRR NDVI data: reduction of environmental and interannual variability. *International Journal of Remote Sensing,* 18(5), 1059 -1077.

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., Chatterjee, S. (2019). Climate change has likely already affected global food production. *PLoS ONE* 14(5): e0217148.

Ren, J., Chen, Z., Zhou, Q., Tang, H. (2008). Regional yield estimation for winter wheat with 5 MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation,* 10, 403–413.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. In Third ERTS Symposium, Washington DC, USA, NASA SP 351 I 309-317.

Satir, O., Berberoglu, S., (2016). Crop yield prediction under soil salinity using satellite derived vegetation indices. Field Crops Res. 192, 134–143.

Savitzky, A., Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry,* 36, 1627– 1639.

Sayago, S., Bocco, M. (2018). Crop yield estimation using satellite images: comparison of linear and non-linear models. *AGRISCIENTIA*, 2018, VOL. 35: 1-9.

Shiferaw, B., Smale, M., Braun, H. J., Duveiller, E., Reynolds, M., Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security,* 5, 291–317.

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 Geosci,* 3, 163–186.

Teasdale, R. J., Cavigelli, A. M. (2017). *Meteorological fluctuations define long-term crop yield patterns in conventional and organic production systems. Scientific Reports* .7: 688.

TUİK. (2021). Turkish Statistical Institute.

Vannoppen, A., Gobin, A. (2021). Estimating Farm Wheat Yields from NDVI and Meteorological Data. *Agronomy,* 11(5):946. doi.org/10.3390/agronomy. 11050946

Wang, Y., Zhang, Z., Feng, L., Du, Q., Runge, T. (2020). Combining Multi-Source Data and Machine Learning Approaches to PredictWinter Wheat Yield in the Conterminous United States. *Remote Sens.,* 12, 1232; https://doi.org/10.3390/rs12081232.