Phenology based NDVI time-series compensation for yield estimation analysis

Conference Paper · August 2017

DOI: 10.1109/Agro-Geoinformatics.2017.8047038

2 authors:

Ayda Aktaş
Istanbul Technical University

Burak Berk Üstündağ
Istanbul Technical University

Phenology Based NDVI Time-series Compensation for Yield Estimation Analysis

Ayda F. Aktas
Informatics Institute
Istanbul Technical University
Istanbul, Turkey
akkartala@itu.edu.tr

B. Berk Üstündağ
Agricultural & Environmental Informatics Research and Application Center
Istanbul Technical University
Istanbul, Turkey
berk@berk.tc

***Abstract*— Normalized difference vegetation index (NDVI) has been correlated with various vegetation parameters using different preprocessing methods, corrections and sampling time based on the aim of the study. In yield estimation studies, maximum NDVI value of the season and the same day of the year NDVI value, which are based on chronological sampling time, are used within different techniques from statistical analysis to machine learning. However, analysis of biological systems based on their chronological timing results in an error increase at data extraction phase due to the non-linearity among phenological stages, representing plant development and its variability. In this study, a phenology based optimum NDVI sampling time is determined and proposed as a replacement of chronologically sampled NDVI time for yield estimation analysis. It may not be possible to have or acquire satellite images for the desired NDVI date due to the temporal resolution of existing remote sensing satellites and meteorological limitations. Therefore, a compensation process based on Adaptive Savitzky-Golay filter and using the existing images is proposed to constitute a new NDVI value for the desired day of the season.**

The study area is situated in the Southeastern Anatolia region of Turkey within the Fertile Crescent where wheat was first cultivated 10000 years ago. The region has the highest durum wheat production, supplying %46 of the whole production in Turkey.

8-day interval, Landsat-7 and Landsat-8 NDVI time-series are analyzed for seasonal vegetation development with TIMESAT software for the 2014-2016 period. Ground-based ancillary data was obtained within the Turkish Agricultural and Environmental Informatics Research and Application Center (TARBIL) project.

Trend analysis of NDVI time-series was performed using Adaptive Savitzky-Golay filter, form of a moving average, adapting to the upper envelope of the data points. Two different sampling methods representing chronological and phenological approaches in addition to the max NDVI value are used to determine the optimum NDVI day. Phenological sampling is carried out as 10-day intervals starting from the emergence phase indicating the start of the season whereas 15 April, representing the long-term annual mean peak NDVI date of the study area was used for chronological sampling.

Adaptive Savitzky-Golay filtering and different sampling combinations were used to perform correlation analysis with annual yield data. Best sampling method along with the optimum NDVI sampling day of the season was determined based on the correlation analysis. It is observed that the combinations with phenological sampling corresponding to the first node stage according to Food and Agriculture Organization (FAO) guidelines have the highest correlation.

Regression analysis between agrometeorological data with and without compensated NDVI and yield variables showed that the usage of compensated NDVI had higher correlation for wheat yield estimation. The results showed that, in comparison with the conventional approaches, the usage of phenology based compensated NDVI, enhanced the yield estimation percentage. Along with the possibility of producing ancillary data from remote sensing images, this approach will minimize the need for ground-based observations that are time and money consuming.

***Keywords*— NDVI; time-series; Plant Phenology; Data Sampling; Data Compensation; TIMESAT; Yield Estimation**

I. INTRODUCTION

Remote sensing derived vegetation indices, especially NDVI, have been used to correlate with various vegetation parameters, however different sampling frequencies, preprocessing methods and corrections need to be performed based on the target parameter. Even though it is the most frequently used vegetation index it has its caveats and limitations. In addition, it may not be possible to obtain NDVI values for the targeted day due to weather conditions and revisit times of remote sensing satellites. In this study, an optimum NDVI day for agricultural applications was determined based on phenological sampling and the new compensated NDVI along with agrometeorological data was tested within direct mathematical relations for crop yield estimation.

Yield is one of the most important parameters as it plays a crucial role in setting up production strategies and ensuring sustainability for future generations with increasing number of population. Conventionally, NDVI is used as an input for statistical correlations with yield data or for existing crop physiological models for yield estimation analysis [1]. Max NDVI value of the season and the same day of the year NDVI value, which are based on chronological sampling, are used for yield estimation analysis consisting various methods such as statistical correlations, neural networks and machine learning [2], [3]. Chronological sampling is established on the idea that the plant will be in the same condition and phase at the same time of every planting season. Contrary, a phenological sampling method which represents the plant development and its variability due to agrometeorological factors, is proposed in this study.

Turkey is among the top ten wheat producers in the world with around 20 million metric tons of annual production [4]. Wheat has the highest total area of production with 66.9 % among other cereals [5]. The study area is situated in Southeastern Anatolia Region that has one of the highest wheat production rates throughout Turkey whilst geographically located in the fertile crescent. The biodiversity in Turkey and 10000 years of wheat cultivation history reveals itself with many cultivated and wild wheat species in the region [4]. Different varieties have different response to climatic conditions and differences in phenological stages, effecting the complexity of agricultural analysis. In addition to this, regional conditions such as small plot farming practices and lack of ground based, up to date, high accuracy crop pattern data effect the overall accuracy of remote sensing integrated agricultural analysis.

Remote sensing based time-series analysis have some challenges related to the data policy indicating mostly the availability of data, sensor type and calibration, study area location and processing phase of data [6]. Beside their coarse resolution, MODIS and AVHRR images are the most used satellite images for time-series analysis targeting agricultural variables with their long-term data support and near-global coverage [7]. As USGS opened Landsat repositories to public in 2008, analysis of higher spatial and temporal resolution time-series became possible. In addition, Collection-1 Level-1 data products produced from surface reflectance data with radiometric and geometric corrections performed and cross-calibrated for different Landsat sensors, suitable for time series analysis, were released in 2016.

Clouds, water vapor and atmospheric contaminants are well known phenomena that decrease the NDVI values. It's also reported that in wet season NDVI tends to have lower values whereas having higher values during the dry season [8]. On the other hand, due to transmission errors, sensor calibration, signal digitization and sensor drift NDVI tends to have localized increases. In addition, when the background soil reflection is dominant rather than the vegetation NDVI does not reflect the vegetation characteristics [9]. Also, when the vegetation gets dense due to the relationship between NDVI and near infrared reflection, it saturates and reaches a plateau [10]. Considering all these limitations of NDVI, even though the data have been preprocessed, there is still some noise present in the time series data [11]. In order to reduce the effect of these limitations, various smoothing techniques have been proposed.

Adaptive Savitzky-Golay filter, one of the most used smoothing filter with NDVI time series in the agricultural analysis, is used to smoothen the data [12]. It is based on a simplified least-squares-fit convolution and referred to as a moving average that preserves the area and the mean position of the peak [13].

Chronological and phenological sampling approaches were performed on the sampled NDVI data as described in section III. The results of correlation and regression analysis with an addition of phenological stage based segmented agrometeorological ground data in order to improve the statistical relationship of NDVI with yield is given in section IV.

II. Study Area and Data Used

A. *Study Area*

Twenty rain-fed winter wheat observation parcels, located in the Southeastern Anatolia Region, were chosen as the study area (Fig.1). Winter wheat, barley, cotton, corn and lentil are the main products of this region. There are two seasons within a year whereas crop rotations and a season of fallow are common practices in the area. The region has its own characteristics with small plot farming and fragmented crop pattern.

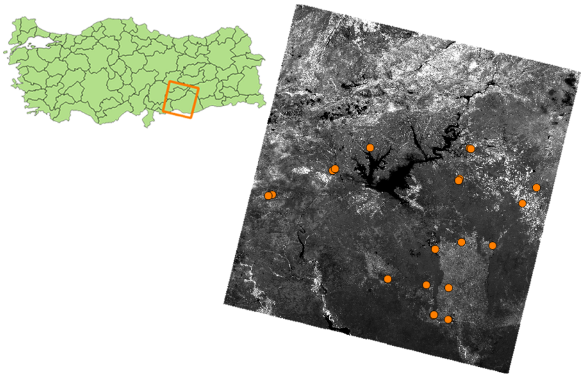

Fig. 1. Study Area

B. *Satellite Images and Ancillary Data*

Collection-1 Level-1 Landsat 7 and Landsat 8 NDVI images, computed from surface reflectance values, were used for time-series analysis. In addition, associated CFmask images were used for the quality assessment. 8-day interval Landsat time-series data is composed of 184 images acquired between 18.05.2013 and 21.05.2017. There were 12 missing images included to the time-series dataset that were processed as no-data and assigned a corresponding pixel quality value.

Selected parcels were part of TARBILs observation network where every TARBIL monitoring station is equipped with 35 sensors capable of atmospheric, soil and phenological measurements. Ancillary ground based observational data and parcel vector info for the selected 20 wheat planted parcels between 2014 and 2016 were provided within TARBIL project. Agrometeorological parameters were calculated and segmented with respect to phenological stage transition terms stated in FAO guidelines (Fig. 2.) [14]. In addition to this, county based yield data was provided from Turkish Statistics Institute (TÜİK).

Fig. 2. Phenological stages of wheat

III. Methodology

Each observation parcel was represented by a specified pixel, based on the digital images of the parcels taken by cameras mounted on TARBIL stations. These pixels NDVI and quality values were extracted and formed as time-series in an ASCII file.

A. Trend Analysis and Savitzky-Golay Filter

Pixel based NDVI time-series analysis of each parcel is performed using the TIMESAT 3.2 software package [13]. Adaptive Savitzky-Golay filter, a form of moving averaging given in its simplest form in equation 1 and 2, was used to smoothen the NDVI growth trend curves (Fig. 3.). c_j corresponds to the Savitzky-Golay coefficient, where n is the half width of the smoothing window. As a moving average window, it preserves the area and mean position of the seasonal peak but not the line width. Savitzky-Golay filter is aimed to assign coefficients that preserve the line width by approximating the underlying data value. This is achieved by using a value from a least-squares fit to a polynomial instead of a constant given in equation 2. For each NDVI value a quadratic polynomial is fitted to the data in the moving window and the y_i is replaced with the value of the polynomial. The window size was set to 4 in order to be able to capture the sudden rise in NDVI values.

$$y' = \sum_{j=-n}^{n} c_j y_{i+j} \tag{1}$$

$$c_j = {}^{1}\!/_{(2n+1)} \tag{2}$$

Pixel quality values were used to define weights for each pixel in the time-series data for filtering process.

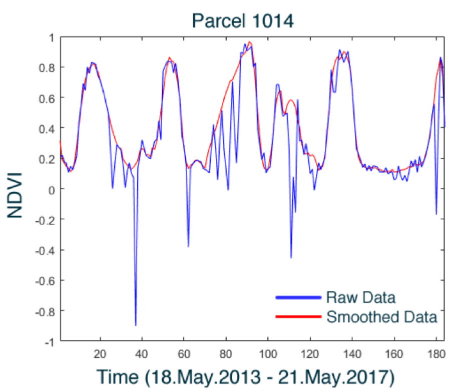

Fig. 3. Adaptive Savitzky-Goray NDVI graph of a wheat parcel

T.R. Ministry of Food, Agriculture and Livestock and I.T.U TARBIL Agro-informatics Research Center.

Growing seasons and phenological parameters such as the start and end of season, length of season, maximum value and seasonal amplitude were extracted based on the smoothened dataset (Fig. 4.).

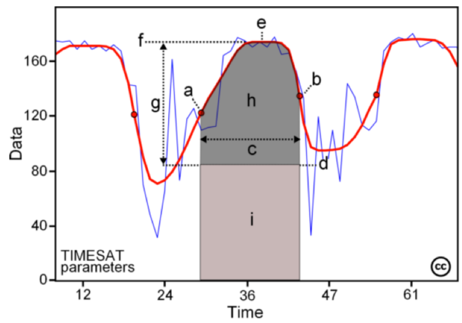

Fig. 4. Some of the seasonality parameters generated in TIMESAT: (a) beginning of season, (b) end of season, (c) length of season, (d) base value, (e) time of middle of season, (f) maximum value, (g) amplitude, (h) small integrated value, (h+i) large integrated value [13].

B. Sampling and Correlation Analysis

In order to determine the optimum NDVI time interval for agricultural analysis, two different sampling methods representing chronological and phenological approach is used in addition to the maximum NDVI.

Smoothened NDVI data for each parcel was sampled with 10-day interval beginning from the start of emergence phase till harvest. This proposed phenological sampling approach is providing the consideration of biological timing of wheat. In contrast with this approach, based on the consideration of chronological timing, NDVI data was sampled regarding to the peak date, representing maximum NDVI and 15th April, representing the long term mean peak date determined for the area.

Among different agricultural analysis, sampled NDVI data was tested for yield estimation, whereas it can be also used for other analysis such as time to harvest. Correlation analysis between sampled data and county based annual yield were performed to examine the performance of phenologically and chronologically sampled NDVI data.

C. Regression Analysis

Growing Day Degree (GDD), Vapor Pressure Deficit (VPD), Photo Thermal Unit (PTU), Minimum Temperature (T_{min}), Rainfall (RF) and Potential Evapotranspiration (PET) are some of the agrometeorological parameters known to have a correlation with plant growth characteristics and thus yield efficiency [15], [16]. These parameters were calculated with using measurements acquired by TARBIL stations and segmented based on phenological stages. The phenological segmentation of agrometeorological parameters provides the mapping of crops biological timing [16].

Ground based agrometeorological data is used to improve the statistical relationship of the remotely acquired NDVI data. The selection and grouping of agrometeorological data is based on the maximum correlation and minimum covariance with yield data. Regression analysis of 15th April NDVI, maximum

NDVI, 7th set of phenologically sampled NDVI with various selected agrometeorological parameters were performed.

IV. Results

Based on the extracted phenological parameters in the trend analysis process, mean peak dates of planting seasons were derived as 23.04.2014, 08.05.2015 and 08.04.2016. It's observed that the maximum NDVI value did not correspond to the same time or even same phenological stage every year. It varies based on environmental conditions and crop structure. Beginning with the drought period in 2013, the annual average rainfall in the region was below the base rainfall (1981-2010) for the following years [17].

As a result of sampling, it is observed that different sampling methods not only have different NDVI values but also correspond to different phenological stages (Table I). It is also observed that %25 of the max NDVI values were overlapping with 15th April NDVI values while the rest have slightly different values with an average of 0.05.

TABLE I. Sampling Methods and Corresponding Phenological Stages

Phenological Stage	Phenological Sampling	NDVI Max.	Chronological Sampling (15 April)
3	5 %	-	-
4	85 %	5 %	5 %
5	10%	25 %	65 %
6	-	50 %	30 %
7	-	20 %	-

In the phenology based sampled dataset, there were 20 sampling sets with a 10-day interval from emergence to harvesting. It is observed that the 7th sampling set, corresponding to the end of first node, beginning of heading stages, has the highest correlation between annual yield data with 0.66 (Fig. 5.). When the NDVI have higher values the correlation between NDVI and yield decreases after the 7th sampling set. Chronologically sampled datasets of maximum NDVI and 15th April NDVI have a correlation value of 0.16 and 0.25 respectively.

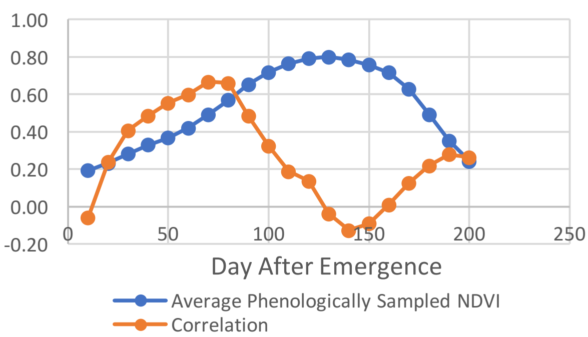

Fig. 5. Avarage phenologically sampled NDVI with respect to day after emergence with 10-day interval and its variation of correlation between wheat yield

The correlation and covariance analysis of phenological stage based segmented agrometeorological parameters were performed for parameter selection and phenological grouping (Fig.6.).

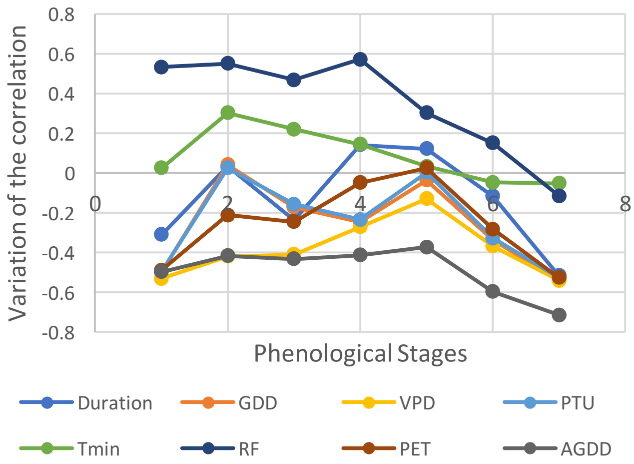

Fig. 6. Correlation of yield versus phenologically segmented agrometeorological parameters

Grouped and selected agrometeorological parameters were used to improve the correlation of chronological and phenological NDVI datasets. Regression analysis between variations of these parameters along with the 15th April NDVI, maximum NDVI, 7th set of phenologically sampled NDVI and annual yield were performed (Table II).

TABLE II. Coefficient of Determination (R^2) Comparison of Chronologically Sampled NDVI, Phenologically Sampled NDVI and Maximum NDVI

	15th April NDVI	$NDVI_{max}$	$NDVI_7$
f (GDD_1)	0.26	0.26	0.74
f (RF_{23})	0.47	0.42	0.68
f (GDD_1, RF_{23})	0.55	0.53	0.85
f (GDD_{1234})	0.20	0.17	0.58
f (GDD_{12})	0.19	0.19	0.68
f (GDD_{12}, RF_{23})	0.51	0.49	0.81
f (PET_1)	0.25	0.26	0.62
f (PET_1, RF_{23})	0.59	0.59	0.82
f (PET_1, RF_{1234})	0.82	0.85	0.84
f (GDD_{12}, $\sum RF$)	0.69	0.69	0.78
f (GDD_1, $\sum RF$)	0.70	0.69	0.8
f (GDD_1, RF_{1234})	0.79	0.79	0.83
f ($\sum GDD$)	0.51	0.51	0.7
f ($\sum GDD$, RF_{23})	0.66	0.66	0.8

It is observed that the usage of different adequate combinations improves the NDVIs relationship with yield. The highest acceptable coefficient determination (R^2) of 0.85 is achieved by the usage of GDD in the first phenelogical stage and grouping of rainfalls for the second and third phenological stages with the given formula in (3).

$$Y = -41.97 + 358.68\ NDVI_{comp} - 0.32\ GDD_1 + 0.73\ RF_{23} \tag{3}$$

There are other combinations that have high R^2 values for the 7^{th} set of phenologically sampled NDVI but they have higher p values than GDD_1 and RF_{23}. The PET_1 and RF_{234} combination seem to have high R^2 values of 0.82 and 0.85 for 15^{th} April NDVI and maximum NDVI, however they have an unsignificant p values that are higher than 0.05. It is shown that the phenology based sampled NDVI corresponding to the end of first node and beginning of heading around 70 days after emergence is the most adequate NDVI for agricultural purposes, in this case for yield estimation.

V. Conclusion

Different mean peak dates for each year is one of the indications that wheat did not show the same chronological growing schedule for every year. The change in the rainfall regime and the trend analysis showing a decrease for further years in addition of increase in temperature effects the wheat growth resulting a shifting of its phenological stages [18]. In addition to the timing of maximum peak date, the correlation analysis of the sampled data showed that after the end of first node and beginning of heading the correlation between NDVI and yield data begin to decrease even though the NDVI value gets higher. These analyses also indicate that, when the NDVI value is at maximum or around maximum value the variation of correlation between yield is around minimum values. Based on the analysis, it is shown that the highest correlated NDVI is the phenologocally sampled NDVI calculated for 70 days after emergence which is proposed sampling time for agricultural analysis.

Ground based agrometeorological data was used to improve the statistical relation of NDVI with yield. It's observed that the usage of different agrometeorological parameters and their variations with the highest correlated phenogically sampled NDVI dataset improved the R^2 value from 0.66 to 0.85. Among all the variations for all the parameters used for analysis, it is shown that the phenologically sampled NDVI have the highest correlation with yield and the NDVI data for further agricultural analysis should be compensated to this NDVI date.

The usage of chronologically sampled parameters in agricultural analysis effects the overall accuracy of the analysis. In this study, the stated improvement of correlation with yield is not only achieved by using phenologically sampled NDVI but also using segmented and grouped agrometeorological parameters based on phenological stages.

Acknowledgment

Authors would like to thank to Istanbul Technical University - Agricultural and Environmental Informatics and Applied Research Center (TARBIL) for providing the field data and Republic of Turkey Ministry of Food and Agriculture and Livestock for support of the project.

References

[1] S. K. Ambast, A. K. Keshari, and A. K. Gosain, "Satellite remote sensing to support management of irrigation systems: Concepts and approaches," Irrig. Drain., vol. 51, no. 1, pp. 25–39, 2002.

[2] I. Becker-Reshef, E. Vermote, M. Lindeman, and C. Justice, "A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data," Remote Sens. Environ., vol. 114, no. 6, pp. 1312–1323, 2010.

[3] H. Jianxi, "Winter Wheat Yield Forecasting Based on Time Series of MODIS NDVI," vol. 47, no. 2, pp. 1–7, 2016.

[4] İ. Özberk, S. Atay, F. Altay, E. Cabi, H. Özkan, and A. Atlı, "Türkiye'nin Buğday Atlası," İstanbul, 2016.

[5] M. Kan et al., "Wheat Landraces in Farmers' Fields in Turkey: National Survey, Collection, and Conservation, 2009-2014," FAO, Ankara, 2015.

[6] C. Kuenzer, S. Dech, and W. Wagner, Remote Sensing Time Series, vol. 22. Springer International Publishing Switzerland, 2015.

[7] C. Atzberger, "Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs," Remote Sens., vol. 5, no. 2, pp. 949–981, 2013.

[8] N. Pettorelli, The Normalized Difference Vegetation Index, First Edit. Oxford University Press, 2013.

[9] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," Remote Sens. Environ., vol. 25, no. 3, pp. 295–309, 1988.

[10] Q. Wang, S. Adiku, and J. Tenhunen, "On the relationship of NDVI with leaf area index in a deciduous forest site," vol. 94, pp. 244–255, 2005.

[11] Y. Curnel and R. Oger, "Agrophenology Indicators From Remote Sensing: State of the Art," in ISPRS Archives, 2006, vol. XXXVI-8/W4, pp. 31–38.

[12] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures.," Anal. Chem., vol. 36, no. 8, pp. 1627–1639, 1964.

[13] L. Eklundh and P. Jönsson, "TIMESAT 3.2 with parallel processing Software Manual," pp. 1–88, 2015.

[14] E. Acevedo, P. Silva, and H. Silva, "Wheat growth and physiology," in Bread Wheat, B. C. Curtis, S. Rajaram, and H. Gómez Macpherson, Eds. Rome: FAO, 2002.

[15] T. Amrawat, N. S. Solanki, S. K. Sharma, D. K. Jajoria, and M. L. Dotaniya, "Phenology growth and yield of wheat in relation to agrometeorological indices under different sowing dates," vol. 8, no. 49, pp. 6366–6374, 2013.

[16] B. Üstündağ, "An adaptive Mealy machine model for monitoring crop status," J. Integr. Agric., vol. 16, no. 2, pp. 252–265, 2017.

[17] A. Uğurlu, B. Aydın, S. Kervankıran, H. Şahin, H. Özcan, and Ö. Öz, "2016 Yılı Yağış Değerlendirmesi," Ankara, 2016.

[18] M. Bahadır, "Temperature and precipitation trend analysis in southeast anatolia project (gap) area," J. Int. Soc. Res., vol. 4, no. 16, 2011.