Highlights:

- MODIS NDVI based wheat and maize yield forecasting method in Tisza river catchment.
- At least six training years are recommended for RS data based yield prediction.
- Yield can be estimated 6-8 weeks before harvest.
- Forecasting model performs the best in drought periods (at average and low yields).

Wheat and maize yield forecasting for the Tisza river catchment using MODIS NDVI time series and reported crop statistics

Attila Nagy, János Fehér and János Tamás

University of Debrecen, Faculty of Agricultural and Food Sciences and Environmental Management, Institute of Water and Environmental Management, Böszörményi 138, 4032 Debrecen, Hungary

Corresponding author: Attila Nagy; address: Böszörményi 138, 4032 Debrecen, Hungary; e-mail: attilanagy@agr.unideb.hu; phone: +36203360382

ABSTRACT

Stakeholders, policy makers, government planners and agricultural market participants in Central Eastern Europe require accurate and timely information about wheat and maize yield and production. The study site, the lowlands (altitude below 200m) of the Tisza river catchment is by far the most important wheat and corn producing region in the Carpathian basin, and even in Central Eastern Europe. The conventional sampling of on-field data and data processing for crop forecasting requires significant amounts of time before official reports can be released. Several studies have shown that wheat and maize yield can be effectively forecast using satellite remote sensing. In this study, a freely available MODIS NDVI satellite data based wheat and maize yield forecasting methodology was developed and evaluated for estimating yield losses effected by drought.

Wheat and maize yield was derived by regressing reported yield values against time series of 15 different peak-season MODIS-derived NDVI. The lowest RMSE values at the river basin level for both wheat and maize yield forecast versus reported yield were found when using at least six or more years of training data. Wheat forecast for the 2000 to 2015 growing seasons were within 0.819 % and 19.08% of final reported yield values. Maize forecast at county level for the 2000 to 2015 growing seasons were within 0.299 % and 17.14% of final reported yield values. The Nash–Sutcliffe efficiency index (E_1) is positive with E_1 = 0.322 in the case of wheat forecast, and with E_1=0.401 in the case of maize forecast, which means the developed and evaluated forecasting method performs acceptable forecast efficiency. Nevertheless the occurrence of extreme drought or extreme precipitation can alter the forecasting efficiency resulting over or underestimation. Overall statement, which based on MODIS NDVI, possible yield losses can easily be forecasted 6-8 weeks before harvesting and applying simple threshold levels, yield losses can be mapped simply.

Keywords: yield forecast, wheat, maize, MODIS, NDVI

1. Introduction

National and international agricultural agencies, insurance agencies, and international agricultural boards Commodity brokers and governmental agencies are interested in crop yields and acreage under crop production since global trading prices of agricultural commodities depend largely on their seasonal production levels. International humanitarian agencies rely on early and reliable information on crop production to organize emergency response and food aid interventions (Rembolt et al., 2013). In crop production drought is one of the most complex natural hazards because of its slow onset and impact on yield which can be monitored with remote sensing (Zambrano et al, 2016).

Remote sensing techniques are widely used in agriculture and agronomy Atzberger (2013). The agricultural application of satellite RS technology requires a quantitative processing of satellite

RS data with high accuracy and reliability. The reason for this first of all agricultural vegetation develops from sowing to harvest as a function of meteorological driving variables (e.g., temperature, sunlight, and precipitation). The production depends secondly on the physical landscape (e.g., soil type), as well as climatic driving variables and agricultural management practices. All variables are highly variable in space and time. Moreover, as productivity can change within short time periods, due to unfavourable growing conditions such as drought, agricultural monitoring systems need to be timely.

As changes in crop vigour, density, health and productivity affect canopy optical properties, crop development and growth have been monitored by the use of satellite images since the early days of remote sensing; Already in the early 80s, it was shown by Tucker and co-workers that green vegetation can be monitored through its spectral reflectance properties (Tucker, 1979; Tucker et al., 1980) and 79% of the variation in total wheat dry-matter accumulation can be explained by integrating normalized difference vegetation index (NDVI) over the growing season (Tucker et al., 1981). Satellite observations can play a role in providing information about crop type, crop conditions and crop yield from the field level to extended geographic areas like countries or continents.

The success of the remote sensing based biomass monitoring stems from its close relation to the canopy Leaf Area Index (LAI) and fAPAR (fraction of Absorbed Photosynthetically Active Radiation) (Prince, 1991; Baret and Guyot 1991). Due to its almost linear relation with fAPAR, NDVI can be readily used as an indirect measure of primary productivity. The aforementioned relationship between vegetation indices and biomass/fAPAR enables the early estimation of crop yield, since yield of many crops is mainly determined by the photosynthetic activity of agricultural plants in certain periods prior to harvest (Beneditti and Rossini 1993; Baret and Guyot 1989). In Rembold et al. (2013), a comprehensive overview is provided regarding biomass and yield mapping approaches. Most of the experiments and research concentrated on obtaining quantitative relation between satellite (or airborne) RS data and crop yields and used two main types of the possible general strategies (Ferencz et al., 2004). The incorporates satellite RS data into (existing or advanced) agrometeorological or plant-physiological, crop growth models (see e.g. Badhwar and Henderson 1981, Brakke and Kanemasu 1981, Asrar et al. 1984, Wiegand and Richardson 1984, Maas 1992, Delécolle et al. 1992, Reynolds et al. 2000, Senay et al. 2000, Patel et al. 2001,Richter et al., 2011, Voulo et al., 2013). The second type of general strategy is based on direct mathematical relationships between satellite RS data and crop yields. Some direct yield methods use meteorological and agronomical data in operation also; and in a few cases some models use only satellite RS data, with ground-truth reference (crop yield) data necessary only in the calibration phase (e.g. Idso et al. 1977, Aase and Siddoway 1981, Gallo and Daughtry 1981, Tucker et al. 1981, Hatfield 1983, Steven et al. 1983, Rudorff and Batista 1991, Hamar et al. 1996, Maselli et al. 2000, Del Frate and Wang 2001, Yun Shao et al. 2001, Balint et al., 2011, Dempewolf et al. 2014). These models assume basically that the vigour of the crop canopy, observed in the spectral RS data, is directly related to the yield of the given crop.

The objective of this study is to develop and test remote sensing based technology for early season wheat and maize yield forecasting in the lowlands of the Tisza river catchment, Central Eastern Europe with using regression-based modelling combining (Moderate Resolution Imaging Spectroradiometer) MODIS time series data and annual reported crop statistics. The concept was based on our earlier experiences and results (Tamás et al., 2015). The aim is to provide first RS based approximations of wheat and maize yield before the final results using the conventional system become available to help improve timely decision-making. In the validation process, we are not only evaluating the absolute deviations of MODIS normalized difference vegetation index NDVI-derived wheat and maize yield data from reported values, but also the significant difference is being assessed between the predicted and observed yield values within different yield ranges. Thus beside overall forecasting accuracy, those yield range can be identified in which the forecasting model performs the best or extremities (drought or too much precipitation) have significant effect on yield forecasting.

2. Materials and methods

2.1. Study site

The study site is the part of an international catchment, the lowlands (altitude below 200m) of the Tisza river catchment is by far the most important wheat and corn producing region in the Carpathian basin, and even in Central Eastern Europe (Figure 1.). As an example, based on the annual reports of the Hungarian central statistical offices, approximately of 55% of the arable lands covered by wheat and maize. The region suffering from water management problems floods, surplus water and drought phenomena occur regularly. Surplus water and drought often occur in the same year or even in the same vegetation period. For crop production, light or radiation, temperature and water relationships (soil moisture) are the three cardinal climatic factors affecting vegetative development and flowering of crop species. Plain sites of Tisza catchment has a substantial global radiation. The average energy input by radiation onto the surface is 4,430 MJ/m^2/year, which is a vast resource for plant production. This relatively high radiation is due to the long photoperiod, which comprises 2,050 hours/year. In Hungary, the average annual daily temperature is 10-11 C, and for the growing season is 17.5 °C.

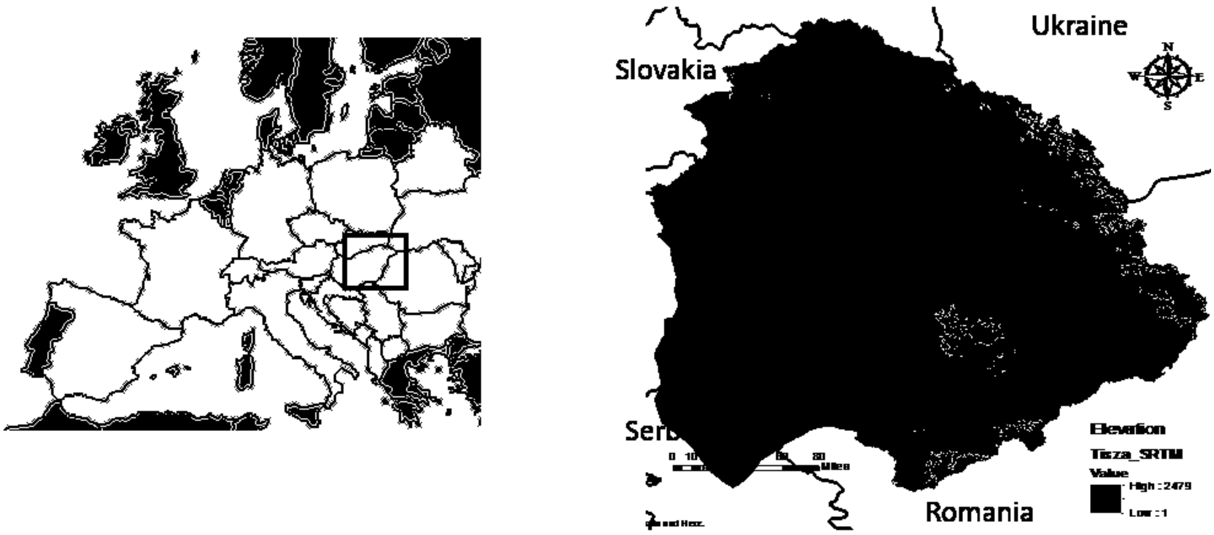


Figure 1. The study site: Tisza river catchment, situated in 5 countries in the Central Eastern Europe

The most variable climate element in the plain site is the precipitation. The average annual precipitation is around 600 mm, but differences between years and the seasonal distribution are extreme. For example, (based on the data of the National Weather Service) looking at figures from Debrecen, middle of the lowland, the minimum and maximum annual precipitations between years 1901 and 2010 were 321 mm and 953 mm, respectively. It is seen that July rainfall may be close to zero or up to 150 mm. This provides an unpredictable water supply for the vegetation and makes crop and fruit production vulnerable. This vulnerability is also explained by the difference between annual precipitation and annual evapotranspiration. It is well known that in mid-season the potential evapotranspiration is high and the precipitation does not meet it, and so there is shortage of soil moisture for crops, furthermore the high clay content can be also a huge problem concerning readily available water content of soils. Climate change models predict that Tisza river basin will experience more serious drought events, and on the other hand more extreme precipitation events in the future. According to statistical data, drought occur in every 2nd, 3rd year in summer period, especially in July and August. Therefore maize is more affected by the drought, than wheat, since wheat is already harvested till the first quarter of July, but maize has its flowering period just in the middle of the most drought risk affected period.

2.2. Crop statistical data

The final official reported yield values were published by the Hungarian Central Statistical Office for the corresponding NUT 3 regions and by Statistical Office of the European Union (EUROSTAT) for Romanian, Slovakian, Serbian NUT 2 regions and collected from 2000 to 2015.

Remarkable yield amounts were detected in 2001, 2004, 2005, 2008 and 2014 (>7 t/ha for maize >4 t/ha for winter wheat); and average in 2006 and 2011 (~6.7 t/ha for maize ~4 t/ha for winter wheat). On the other hand due to drought phenomena severe wheat and maize yield losses were detected in 2000, 2002, 2003, 2007, and 2012 (-3 t/ha loss for maize -1-1.5 t/ha loss for wheat). (Figure 2.).

Figure 2. Average yield changes of maize and wheat, 2000–2015 based on KSH (Hungarian Central Statistical Office) and EUROSTAT

These data are strongly related to the SPI and meteorological data, except for year 2010, when an extreme amount of precipitation (900–1,300 mm/year) was observed on the plain sites of the Tisza river basin, and, due to the surplus drainage water cover on the fields for a long period and plant diseases, the quantity of the yields remained average (Tamás et al., 2015).

2.3. MODIS NDVI data

In the case of low resolution satellite images, thanks to their large swath width, low resolution systems have a much better synoptic view and temporal revisit frequency compared to high resolution sensors (Rembolt et al., 2013). On the other hand the spatial resolution seriously complicates the accuracy of yield detection, the interpretation (and validation) of the signal, as well as the reliability of the derived information products. Although Labus et al. (2002) calculated NDVI from an AVHRR time series for the U.S. state of Montana and found strong correlations between wheat yield and integrated NDVI, as well as late-season NDVI parameters and Reeves et al. (2005) used successfully 1 km Moderate Resolution Imaging Spectroradiometer (MODIS) data to estimate wheat yields in North Dakota and Montana, but an average farm size is smaller in Hungary (which is about 14-15 ha) (Biro et al, 2011) and in Central East European (CEE) region than in the USA. Therefore the monitoring of yield is not appropriate in CEE region with datasets, such as Fraction of Absorbed Photosynthetically Active, Radiation (fAPAR) or AVHRR data, having low spatial resolution (>1 km) (Gobron and Verstraete, 2009), because one pixel exceeds the average crop farm size in CEE region. Meroni et al. (2013) examined the performance of spectral parameters derived from SPOT-VEGETATION data for wheat yield forecasting in Tunisia and, for NDVI, achieved an r-squared value of 0.75 between modelled and observed yield. Although Landsat (or similar sensors such as SPOT) are also the main source of data with sufficient spatial resolution in most agricultural areas, but with a 16-day gap between successive images, and frequent cloud cover in most cropping regions (with the exception of dry, irrigated areas), it can be difficult to obtain more than one or two clear images within a growing season (Lobell, 2013). Sentinel data can be a possible alternative, but in yield prediction the necessary number of training years is at least four years, and the inter-calibration issues among different datasets still must be solved (Yin et al., 2013). On the other hand, Wardlow *et al.* (2007) and Mkhabela et al. (2011) in the USA, and Ferencz et al. (2004) in Hungary, concluded that MODIS time-series at 250 m ground resolution had sufficient temporal and radiometric resolution to discriminate major crop types and crop-related land use practices. Thus MODIS NDVI data with 250 m spatial resolution was chosen in this study for farm and regional scale yield assessment. One should still note, that the 250-m MODIS pixels could contain less than 100% wheat and maize sites and are partially covered by other land cover types, which introduces an inherent uncertainty into the measurements (Dempewolf et al., 2014).

The MODIS has been a key environment remote sensing tool for more than 18 years; it has been used in countless studies of different disciplines all over the world. The MODIS instrument was developed to improve heritage sensors in terms of its spectral, spatial, and temporal resolutions, as well as more stringent calibration requirements. (Xiong et al., 2009). The usefulness of MODIS NDVI for evaluating vegetation response is well known (Huete et al., 2012). In the present case, the vegetation indices (VI) were obtained from the MODIS 'Vegetation Indices 16-Day L3 Global 250 m' short name 'MOD13Q1' product (Didan 2015). A complete 16-year time series (2000–2015) was downloaded through the online Data Pool at the NASA. In this study, we used MODIS data for two purposes, for mapping the presence of wheat and maize and for yield forecasting.

2.4. Data quality issues - smoothing

Several studies pointed out that probably any filtering is better than no filtering (Rembolt et al., 2013; Atzberger and Eilers, 2011; Hird and McDermid, 2009; Meroni et al., 2012). A smoothing process was required to reduce noise in the NDVI time series. Multiple techniques are available in the literature to do this (Hird, J.N. and McDermid, 2009; Julien, Y.and Sobrino 2010; Klisch and Atzberger 2016; Atkinson et al 2012). In a recent comparative study by Atkinson et al., (2012) involving a number of commonly used filters, it was shown that the 'Whittaker smoother' (based on penalized splines) provides robust results for different noise levels and different cropping patterns (e.g., single *vs.* double cropping). Therefore in present case, modified Whittaker smoother was used for MODIS NDVI data smoothing (Figure 3.).

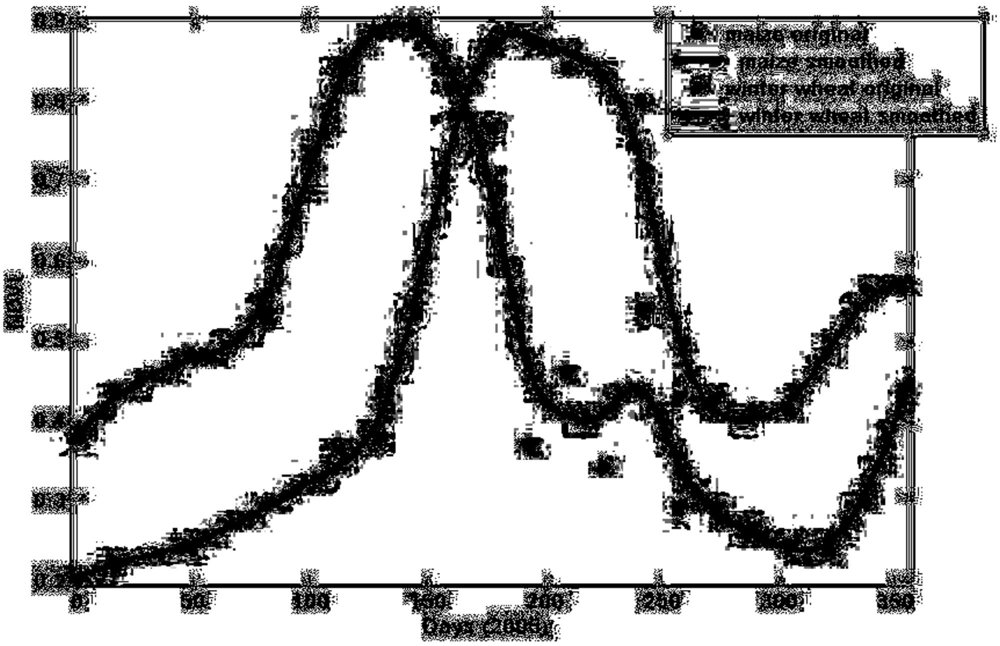


Figure 3. Illustration of the effect of whittaker smoother on the NDVI profile of maize and wheat based on the data from 2005 in HajdúBihar county (part of the examined area)

2.5. Cropland mask

Beside smoothing another obstacle to successful modelling and prediction of crop yields using remotely sensed imagery is the identification of image masks (Kastens et al., 2005). Where the crop area is not known, the NDVI/yield relationship does not provide information on final crop production, which is what many users of crop monitoring information are ultimately interested in (Rembolt et al., 2013). Cropland masking, where all sufficiently cropped pixels are included in the mask regardless of crop type, has been shown to generally improve crop yield forecasting ability (Doraiswamy and Cook, 1995; Lee et al., 2000; Maselli and Rembolt, 2001). Cropland masks usually are derived from existing land use/land cover maps. However, when masking is applied to multiple years of imagery, several difficulties are encountered (Becker-Reshef et al., 2010). A major problem relates to the widespread practice of crop rotation, when a single cropland mask would not be appropriate. For these reasons in general, a direct NDVI/production regression makes only sense under specific conditions, such as a stable crop area over the observed period using cropland mask (Rembolt et al., 2013) or using crop specific masking (i.e., one mask per crop type and year) or yield correlation masking due to changes in crop area as a result of crop rotation (Maselli et al., 2000; Kastens et al., 2005). This would allow one to consider only NDVI information pertaining to the crop of interest.

In this study crop specific masks were produced for wheat and maize and every year. Masking was a robust process. In the data processing we used standardized geographical, landuse and terrain data and information. As a first step the plain area with arable land was clipped out of the NDVI time series data every year. United States Geological Service (USGS) Shuttle Radar Topography Mission (SRTM) model was used to select plain areas, altitude below 200 m (source USGS, http://srtm.usgs.gov/index.php). Thereafter CORINE (COoRdinate INformation on the Environment) Landcover datasets (CLC 2000, CLC 2006 and CLC 2012) were used as a cropland to select arable lands out of plain areas, in order to reduce the possible area for crop specific masking.

For the per pixel characterization of wheat and maize presence in the arable plain land on Tisza river catchment we used the already produced MODIS NDVI cropland site data of July and April for each year for vegetation coverage. Based on the classification of images, vegetation cover Boolean masks were created (two images/year). These images were classified into Boolean masks each indicating vegetation cover and barren site circumstances in the vegetation periods. These masks were used to select vegetation covered and covered places in April and in July to identify wheat, and maize covered sites. With this technique in the case of wheat, all the area covered by alfalfa, maize and industrial crops can be eliminated. Though the final wheat specific masks (based on the crop cover data of) was still contained less than 5% uncertainty mainly due to barely and triticale cover. The uncertainty was defined by the official crop coverage KSH and EUROSTAT statistical data. In the case of maize, though all other cereals, alfalfa, rape were possible to exclude from the investigated area, but there was still a need to overcome the effect of industrial crops, dominantly sunflower cover (95% out of all industrial plants). Taking the advantage of the effect of flowering on NDVI, sunflower masks (1 mask/year) was created using the MODIS data in July, and applied resulting a final maize specific masks for each year. Final wheat and maize masks were then applied on NDVI images. At the end the mean NDVI values of NUT 2 and NUT 3 regions (i.e. counties in Hungary and regions in Romania, Slovakia and Serbia) were extracted as an input for yield regression.

2.6. Yield forecast

The predictive yield models were constructed using simple linear regression analysis of peak-season MODIS-derived NDVI indices against reported crop yields from the years preceding the forecast year. The necessary number of training years was evaluated by calculating forecasts using between two and sixteen training years. The timing of the forecasts within the growing season was evaluated previously in Tamás et al., (2005) study, in which. useful statistical relationships reported using NDVI values at the peak of the growing season (duration approximately four –six weeks before harvest) and final crop yield in correspondence with other studies (Rembolt et al, 2013; Delecolle et al, 1992; Becker-Reshef et al., 2010; Boken et al., 2002, Basnyat et al., 2004.). Therefore, in this study MODIS NDVI data from May and June were used in the case of wheat, and MODIS NDVI data from July and August were used in the case of maize for analysing regression based yield forecasting. An adjustment to the yield forecasts was made by regressing the estimated yield values of the training years against the reported yields and applying the adjustment regression equation to the estimated yield of the forecast year. Due to crop specific masking the result of this study NDVI/yield regression can be an appropriate solution for crop forecasting Rembolt et al. (2013).

The minimum numbers of the years for forecasting and the performance of forecast and the identification of was assessed using the accuracy metrics coefficient of determination (R^2), root means square error (RMSE) and normalized RMSE (NRMSE):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - y_i')^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{1}$$

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

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - y_i')^2}{n}} / (\max(y_i) - \min(y_i)) \tag{3}$$

where y_1 and y_i' are the measured and predicted yield values for sample i, $\overline{y}$ is the mean yield and n is the number of samples used for validation. RMSE provides an absolute measure of prediction errors and NRMSE is useful for comparisons between seasons in case of variable yield ranges (Darvishzadeh et al., 2008). Nash-Sutcliffe efficiency 'E_1' was also calculated. The efficiency E_1 proposed by Nash and Sutcliffe (1970) was defined as one minus the sum of the absolute squared differences between the predicted an observed values normalized by the variance of the observed values during the period under investigation. It is calculated as:

$$E_1 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \quad (4)$$

In the validation process to assess overall forecasting accuracy, we are evaluating the absolute deviations of MODIS normalized difference vegetation index NDVI-derived wheat and maize yield data from reported values. In order to highlight those yield range in which the forecasting model performs the best or extremities (drought or too much precipitation) have significant effect on yield forecasting, significant difference was assessed between the predicted and observed yield values within different yield ranges.

3. Results

In this study wheat and maize yield was derived by regressing reported yield values against time series of 16 different peak-season MODIS-derived NDVI. The use of 250-m MODIS-derived NDVI was analysed and tested for wheat and maize yield production assessment and forecasting for Tisza river catchment area. We assessed the wheat and maize yield forecasting accuracy under a mask derived from MODIS NDVI data, analysed the optimal number of training years for accurate forecast.

The optimal number of training years was determined for wheat yield forecasting by calculating R^2, RMSE and NRMSE for the sixteen peak seasons from 2000 to 2015 using the NDVI index and between two and 16 training years. The values were averaged over the most sensitive (blooming and ripening) period of wheat (the end of May and June). The deterministic coefficients were the highest (R^2>0.7) in using 5-7 training years, with the maximum at five training years with R^2=0.732. On the other hand the NRMSE reaches its minimum values at six training years (NRMSE = 13.9%). The NRMSE did not changed significantly with increasing training years (NRMSEs were between 13.9-14%) (Figure 4.). Since NRMSE performs much better at 6 years than 5 years (16.7%) training data, the minimum data requirements for wheat yield forecasting was identified and we therefore used six training years in the subsequent analysis.

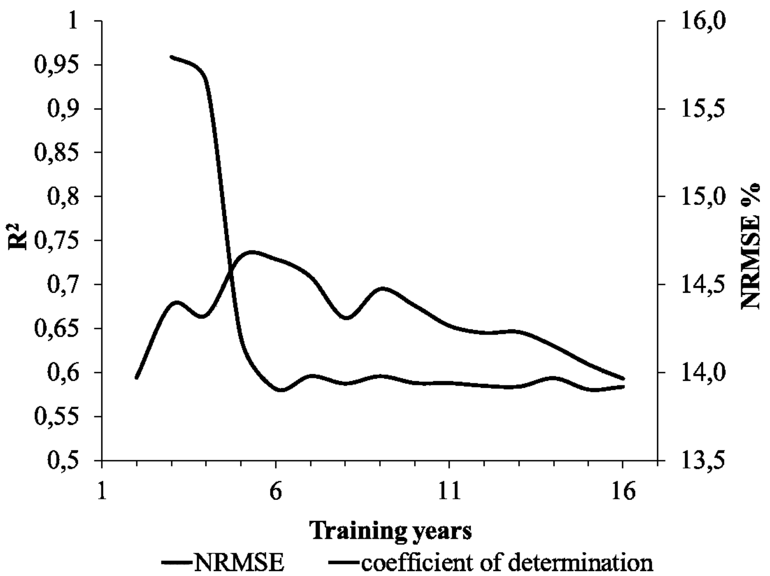


Figure 4. NRMSE and determination coefficient of forecast versus reported wheat yield at the catchment level for an increasing number of training years.

In the case of maize the R^2, RMSE and NRMSE values were averaged over the most sensitive (blooming and ripening) period of maize (July and August) from 2001 to 2015. The deterministic coefficients were the highest (R^2>0.8) in using 5-6 training years, with the maximum at six training years with R^2=0.815. The NRMSE reaches its minimum values at six training years (NRMSE = 15.1%) (Figure 5.). Using twelve training years results in an only slightly lower value (RMSE = 14.9%) compared to six years. Thus in the case of maize six training years were used in further analysis. An increase in NRMSE was measured in the 7th and 8th years, which probably due to higher uncertainty in the relation between NDVI and yield.

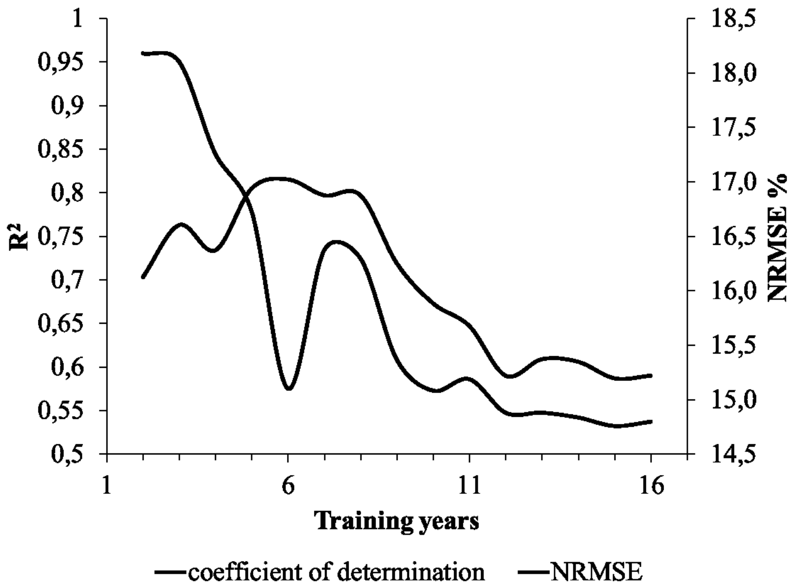


Figure 5. NRMSE and determination coefficient of forecast versus reported maize yield at the catchment level for an increasing number of training years.

The performance of NDVI for wheat and maize yield forecasting was calculated at the county level using six years of training data. The results were compared to official reported yield values every year. At the county level, we calculated the RMSE and the relative deviation (difference in percent) of forecast versus reported yield (Figure 6.). At county level, absolute deviation of NDVI-derived wheat yield from reported values ranged from 0.819 % in the 2004 season to

19.08% in the 2010 season. Absolute deviation of NDVI-derived maize yield from reported values ranged from 0.299% in the 2012 season to 17.14% in the 2014 season.

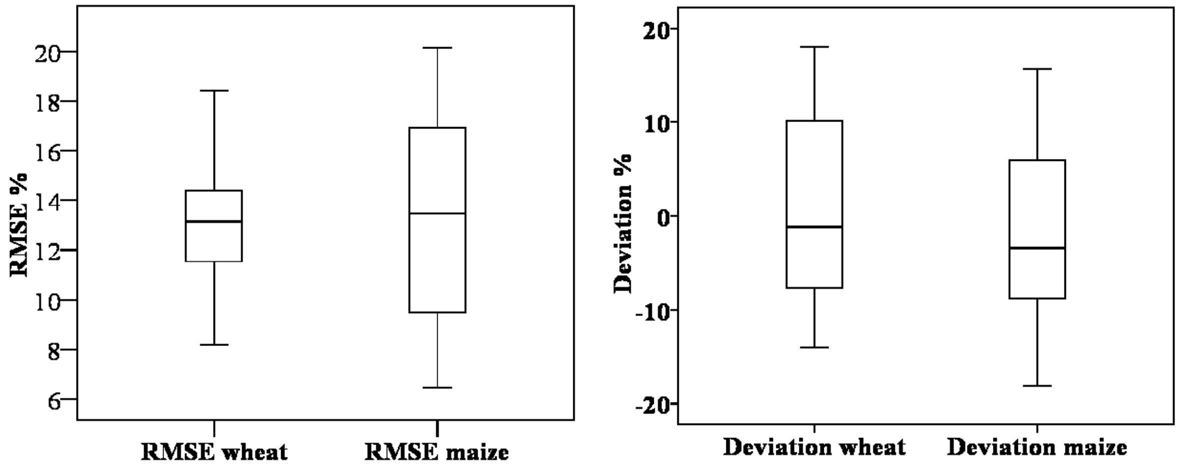


Figure 6. RMSE and deviation of predicted values compared to official reported wheat and maize yield values

The deterministic coefficients for wheat and maize were more than 70% and 80% during the phenlogical peak period using six training years. Although the average absolute deviations between estimated and officially reported county yield data was about 7% for wheat and 8% for maize (Figure 7.). These values were a bit higher than the 5% threshold, which is generally accepted as good (Ferencz et al., 2004). Therefore, yield forecasting results were compared to simply using the three-year or six-year moving averages of the years preceding the forecasting year (Figure 7.). The results show that the forecast yields had, on average, lower deviation from reported values than the moving averages, and thus, the forecast performs better. We also tested the performance of the wheat yield forecast using the Nash–Sutcliffe efficiency index, (E_1), which is a global measure of model efficiency. The Nash–Sutcliffe efficiency index is positive with $E_1 = 0.3$ in the case of wheat forecast, and E_1=0.401 in the case of maize forecast.

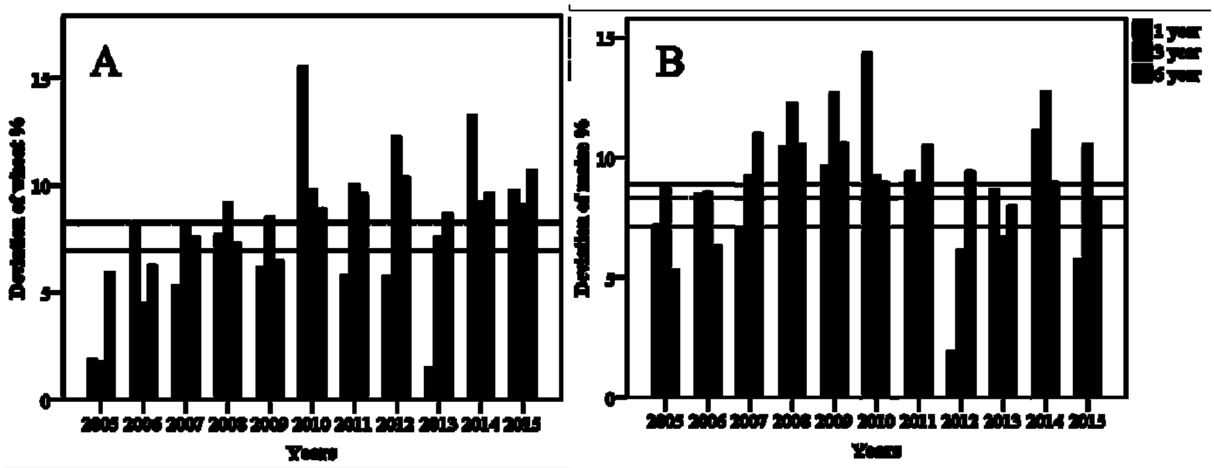
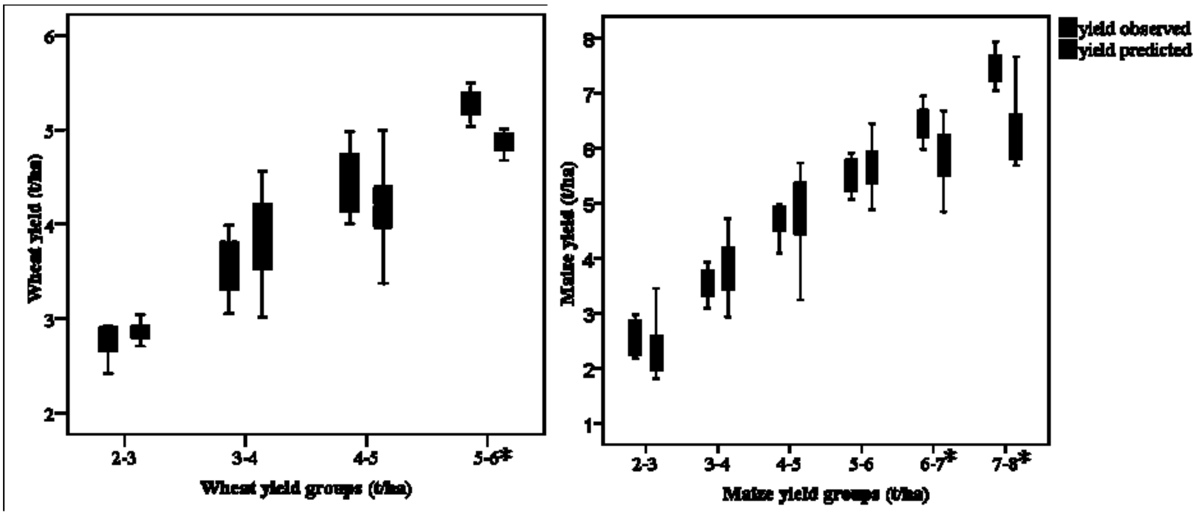


Figure 7. Deviation of forecast from reported wheat (A) and maize (B) yields at catchment level (blue bars) and the overall average (solid horizontal line) for the seasons 2005 to 2015 in comparison to the deviation of the three-year moving average (green bars and dashed line) and the six-year moving average (yellow bars and dotted line) yields.

After assessing the overall yield prediction accuracy, the uncertainties and forecasting precision for different yield ranges was evaluated in order to highlight those yield range in which the forecasting model performs the best. Tukey's B variance analyses were used to assess significant difference between the related observed and predicted yield values within four wheat and six maize yield ranges. As a result, the distribution of the predicted yields was possible to compare to the real, observed data distributions (Figure 8.).



*yield range, in which there was significant difference between predicted and observed yield data ($p<0.05$)

Figure 8. Differences between observed and predicted yield within wheat and maize yield ranges

In the case of wheat higher yield values were significantly underestimated. The difference between predicted and observed yield is 0.56 t/ha (in average). Maize forecast performs similar characteristics, since significant differences is detected in the case of high yield values, the overestimation is between 0.5-0.9 t/ha in the case of yields above 6 t/ha.

4. Discussion

The purpose of this study was to develop a satellite-based system for wheat and maize yield forecasting and to determine the uncertainties of the prediction for different yield amounts for a solution applicable to lowlands of the Tisza river catchment, but also transferable to CEE region.

The results of this study provides specific recommendations for the necessary number of training years. Six years of historical data are the minimum number of training years recommended for forecasting wheat and maize yield. This statement is in accordance with the results of Dempewolf et al. (2014) in the case of wheat forecasting. Fewer years did not seem to provide enough data points for deriving meaningful regression equations

Previous studies have shown the validity of using satellite-derived vegetation indices for wheat and maize yield forecasting. In accordance with the other studies our results achieved good agreement (7%) between wheat yield derived from MODIS-derived NDVI and reported yield: the forecast of yield for the majority of cases was within 10% of final reported values in Pakistan (Dempewolf et al., 2013). Ren et al. used 10-day MODIS NDVI composites to forecast the yield of wheat for a sub-region of Shandong Province in China, and the results were within 5% of official statistics. Sakamoto et al. (2013) was estimated maize yield accurately; yield deviation was below 10%, which is in accordance with our findings. Furthermore, our results performs better than another study using the same 16-day composite in Serbia. In the mentioned study the smallest difference between predicted and actual yield was 1.67% and the largest difference was 44.12% (Govedarica et al., 2016), whilst our result is within 0.299 % and 17.14%.

The satellite-based yield forecasts were much less accurate for the 2010 season than for other seasons in the case of maize. This might be due to unusual weather patterns in 2010, when an extreme amount of precipitation (900–1,300 mm/year), with cooler spring and summer was observed on the plain sites of the Tisza river basin. This circumstances covered the whole vegetation period of both examined crop. Besides, the surplus water cover on the fields was common in spring (Tamás et al., 2015). Thus agricultural works were significantly delayed in spring time, and due to the rainy weather and the hardly accumulating active heat, the normal development of the crops were delayed prolonging the growing season and causing the delay of harvest period. However, due to more favourable conditions in July and August, the wheat and maize caught up subsequently, and the final impact on yield was only small, remained average (Ragán et al. 2014.). This unusual pattern of delayed crop development during the normal seasonal peak time of wheat in spring and early summer in combination with a quick subsequent recovery might explain the lower performance of the forecasting system.

Applying six training years, yield forecasting performs better compared to simpler methods of obtaining yield data, such as using the previous year's value or the three-year or six-year moving average. Nash–Sutcliffe efficiency of higher than zero also indicates that the tested prediction method is a better predictor than the mean value of the observed time series the developed forecasting method is applicable for wheat and maize prediction. Furthermore in the case of wheat our results is better than a study in Pakistan, where the E_1 was only 0.112 (Dempewolf et al., 2014.).

Investigating the forecast in different yield ranges, yield prediction in the case of high yield values have the highest uncertainties, partly due to extreme weather circumstances in 2010 resulting delay of phenological phases resulting smaller NDVI which did not reflect the recovery of the plants in the final stage. As a result in higher yield ranges extremity with cooler weather or too much precipitation has significant effect on yield forecasting. Another possible reason for the uncertainties might be that NDVI is known to saturate at high LAI values (Sellers 1985, Goswami 2015), resulting the decrease in NDVI sensitivity for higher yields. This phenomenon can be explain that the satellite-based yield forecasts were the second less accurate for the 2014 season, whilst there were record yields for maize (7.82 t/ha) and wheat (5.18 t/ha) at the examined site. The forecasting model performs the best from average to the lowest wheat and maize yields, resulting that the prediction can be a very useful tool for detecting yield or yield losses caused by drought phenomena, thus can be a viable option in crop specific drought monitoring as well.

A common problem in crop monitoring and yield forecasting in many countries of the world is the difficulty in extending locally calibrated forecasting methods to other areas or to other scales since most of the studies are linked to the environmental characteristics of specific geographic areas (Rembolt et al., 2013). In this study the results had validated based on yield data from international catchment area, thus valid for the agricultural land in the Tisza river basin, though hadn't validated on yield and NDVI data in wider range of Europe. Based on EUROSTAT data there are few differences in average weather circumstances, in the optimal amount of maize and wheat yields (t/ha) and in the level of agricultural practice in the Carpathian basin and in the CEE region, thus our finding is possible to extend for CEE region. Certainly, there could be small differences in the intensity of crop production, wheat species and especially in maize hybrids between countries, which differences could influence the amount of yield.

The developed model is based on NDVI, (MODIS NDVI). Until recent years, at high revisit frequency, the Earth's land surface could only be covered by coarse/medium resolution sensors, such as MODIS. Nowadays with the Sentinel's 2 and 3 and Proba-V sensors a new era of Earth observation is entered (Rembolt et al. 2013). With new sensors, data availability at coarse/medium resolution increased at high revisit frequency, but still more efforts should be taken in further studies to ensure a suitable sensor inter-calibration, especially because there is not yet enough time series datasets for accurate yield forecasting. Although even with a better sensor inter-calibration, it is not certain that derived products (such as NDVI or fAPAR) are comparable across sensors or even data providers (Meroni et al. 2012).

5. Conclusion

Recent advances in operational space technology have improved our ability to address many issues of early detection of yields. In this way yield forecast support to fill the gap of knowledge between remote sensing data and decision-making, in order to develop yield forecast related decision parameters and application in practice from raw spectral datasets.

The wheat and maize forecasting method estimates the expected yield based on remote sensing data with 250*250 m spatial resolution. Our study was based on multi-spectral remote sensing data (MODIS NDVI) and reported yield data, forecasting method was formulated with calibrating of remote sensing data with the important crops (wheat, maize) which are representative in the Tisza river catchment and in the CEE. The developed wheat and maize yield forecasting provides timely information on crop production, status and yield in a standardized and regular manner at the (sub)regional (county) to the international catchment level. With help from the forecasting method developed based on six training years, the yield can be predicted 6-8 weeks earlier than harvesting. Understanding the applicability and accuracy of yield prediction is also an essential component of forecasting because the ultimate goal is to reduce forecast uncertainties for a particular location and for a specific group of people or agricultural or economic sector. With the forecasting method moderately good estimates are provided as early as possible during the growing season and can be updated periodically through the season until harvest. This information can reduce impacts of possible yield losses if delivered to farmers or decision makers in a timely and appropriate format and if mitigation measures and preparedness plans are in place. Based on the information provided, stakeholders are enabled to take early decisions and identify geographically the areas with large variation in production and productivity which is one of the most vital need for food security and trade. The forecasting needs further development with new sensors with high revisit frequency and good spatial resolution (10-30 m). However, sensor inter-calibration is still an important issue to provide homogeneous and interchangeable data sets with statistically valid precision and accuracy.

Acknowledgements

This research was supported by EFOP-3.6.2-16-2017-00001 Research on the development of complex rural-economy and sustainability with related service network in the Carpathian basin; University of Debrecen Faculty of Agricultural and Food Sciences and Environmental Management Arid Land Research Centre. The basics of this study was provided by the joint Integrated Drought Management Programme of GWP (Global Water Partnership) and WMO (2011-2014) as well as by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.4.A/2-11/1-2012-0001 'National Excellence Program'.

References

Aase, J.K., and Siddoway, F.H., 1981. Assessing winter wheat dry matter production via spectral reflectance measurements. Remote Sensing of Environment, 11, 267–277.

Asrar, G., Fuchs, M., Kanemasu, E.T., and Hatfield, J.L., 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. Agronomy Journal, 76, 300–306.

Atkinson, P.M., Jeganathan, C., Dash, J., Atzberger, C., 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. Remote. Sens. Environ., 123, 400–417.

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

Atzberger, C., Eilers, P.H.C. 2011. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. Int. J. Remote Sens., 32, 3689–3709.

Badhwar, G. D., and Henderson, K.E., 1981. Estimating development of corn from spectral data—an initial model. Agronomy Journal, 73, 748–755.

Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sens. Environ. 35, 161–173.

Baret, F., Guyot, G., Major, D.J., 1989. Crop biomass evaluation using radiometric measurements. Photogrammetria, 43, 241–256.

Basnyat, P., McConkey, B., Lafond, G.P., Moulin, A., Pelcat, Y., 2004. Optimal time for remote sensing to relate to crop grain yield on the Canadian prairies. Can. J. Plant Sci. 84, 97–103.

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 Sens. Environ., 114, 1312–1323.

Benedetti, R., Rossini, P., 1993. On the use of NDVI profiles as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. Remote Sens. Environ. 45, 311–326.

Biro, Sz., Apáti, F., Szőllősi, L., Szűcs, I., 2011. The economic context of irrigation development. (in Hungarian) In: Biro, Sz., Kapronczai, I., Szűcs, I., Váradi, L. 2011. Water use and irrigation development in Hungarian agriculture (in Hungarian). Agricultural Research Institute, Budapest, 45-74.

Boken, V.K., Shaykewich, C.F., 2002. Improving an operational wheat yield model using phenological phase-based Normalized Difference Vegetation Index. Int. J. Remote Sens., 23, 4155–4168.

Brakke, T.W., and Kanemasu, E.T., 1981. Insolation estimation from satellite measurements of reflected radiation. Remote Sensing of Environment, 11, 157–167.

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. Remote Sensing of Environment, 112, 2592–2604.

Del Frate, F., and Wang, L.F., 2001. Sunflower biomass estimation using a scattering model and a neural network algorithm. International Journal of Remote Sensing, 22, 1235–1244.

Delécolle, R., Maas, S.J., Guérif, M., Baret, F., 1992. Remote sensing and crop production models: Present trends. ISPRS J. Photogramm., 47, 145–161.

Dempewolf, J., Adusei, B., Becker-Reshef, I., Barker, B., Potapov, P., Hansen, M., Justice, C., 2013. Wheat production forecasting for Pakistan from satellite data. In Proceedings of 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, VIC, Australia, 21–26 July 2013, 3239–3242.

Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A., Barker, B., 2014. Wheat yield forecasting for Punjab province from vegetation index time series and historic crop data. Remote Sensing, 6, 9653-9675.

Didan, K., 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V006. Technical Report, NASA EOSDIS Land Processes DAAC, 2015. Available online: http://dx.doi.org/10.5067/MODIS/MOD13Q1.006.

Doraiswamy, P.C., Cook, P.W., 1995. Spring wheat yield assessment using NOAA AVHRR data. Can. J. Remote Sens., 21, 41–51.

Ferencz, Cs., Bognár, P., Lichtenberge 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. Int. J. Remote Sensing,. 25(20), 4113–4149.

Gallo, K.P., and Daughtry, C.S.T., 1981. Spectrally derived inputs to crop yield models. Proceedings of the 1981 Symposium on Machine Processing of Remotely Sensed Data (West Lafayette: Purdue University), 52–65.

Goswami S, Gamon J, Vargas S, Tweedie C. 2015. Relationships of NDVI, Biomass, and Leaf Area Index (LAI) for six key plant species in Barrow, Alaska. PeerJ PrePrints 3:e913v1

Govedarica, M., Jovanović, D., Sabo, F., Borisov, M., Vrtunski, M., Alargic, I. 2016. Comparison of MODIS 250 m products for early corn yield predictions: a case study in Vojvodina, Serbia. Open Geosciences, 8(1), 747-759.

Hamar, D., Ferencz, CS., Lichtenberger, J., Tarcsai, GY., and Ferencz-Árkos, I., 1996. Yield estimation for corn and wheat in the Hungarian Great Plain using Landsat MSS data. International Journal of Remote Sensing, 17, 1689–1699.

Hatfield, J.L., 1983, Remote sensing estimators of potential and actual crop yield. Remote Sensing of Environment, 13, 301–311.

Hird, J.N., McDermid, G.J., 2009. Noise reduction of NDVI time series: An empirical comparison of selected techniques. Remote Sens. Environ. 113, 248–258.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote. Sens. Environ., 83, 195–213.

Idso, S.B., Jackson, R.D., and Reginato, R.J., 1977. Remote sensing of crop yields. Science, 196, 19–25.

Julien, Y., Sobrino, J.A. 2010. Comparison of cloud-reconstruction methods for time series of composite NDVI data. Remote. Sens. Environ. 114, 618–625.

Kastens, J.H., Kastens, T.L., Kastens, D.L.A., Price, K.P., Martinko, E.A., Lee, R.-Y., 2005. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. Remote Sens. Environ., 99, 341–356.

Klisch, A., Atzberger, C., 2016. Operational drought monitoring in Kenya using MODIS NDVI time series. Remote Sens., 8, 267.

Labus, M.P., Nielsen, G.A., Lawrence, R.L., Engel, R., Long, D.S., 2002. Wheat yield estimates using multi-temporal NDVI satellite imagery. Int. J. Remote Sens., 23, 4169–4180.

Lee, R., Kastens, D.L., Price, K.P., Martinko, E.A. 2000. Forecasting Corn Yield in Iowa Using Remotely Sensed Data and Vegetation Phenology Information. In Proceedings of the Second International Conference on Geospatial Information in Agriculture and Forestry, Lake Buena Vista, FL, USA, 10–12 January 2000, Volume II, 460–467.

Maas, S. J., 1992. GRAMI: a crop growth model that can use remotely sensed information. ARS-91, USDA Agricultural Research Service, 1–78.

Maselli, F., Rembold, F. 2001. Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries. Photogramm. Eng. Remote Sensing, 67, 593–602.

Maselli, F., Romanelli, S., Bottai, L., and Maracchi, G., 2000. Processing of GAC NDVI data for yield forecasting in the Sahelian region. International Journal of Remote Sensing, 21, 3509–3523.

Meroni, M., Atzberger, C., Vancutsem, C., Gobron, N., Baret, F., Lacaze, R., Eerens, H., Leo, O., 2012. Evaluation of agreement between space remote sensing SPOT-VEGETATION fAPAR time series. IEEE Trans. Geosci. Remote Sens. 51, 1–12.

Meroni, M., Marinho, E., Sghaier, N., Verstrate, M., Leo, O. 2013. Remote sensing based yield estimation in a stochastic framework—Case study of durum wheat in Tunisia. Remote Sens., 5, 539–557.

Mkhabela, M., Bullock, P., Raj, S., Wang, S., Yang, Y., 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric. For. Meteorol., 151, 385–393.

Patel, N.K., Patnaik, C., Dutta, S., Shekh, A.M., and Dane, A.J., 2001, Study of crop growth parameters using Airborne Imaging Spectrometer data. International Journal of Remote Sensing, 22, 2401–2411.

Prince, S.D., 1991. A model of regional primary production for use with coarse resolution satellite data. Int. J. Remote Sens., 12, 1313–1330.

Ragán, P., Bakó, K.I., Sedlák, G. 2014. The impact of environmental changes related different sowing time on the yield of maize. (In Hungarian) Acta Agraria Debreceniensis, 55: 99-104.

Rembold, F., Atzberger, C., Rojas, O., Savin, I., 2013. Using low resolution satellite imagey for yield prediction and yield anomaly detection. Remote Sens., 5: 1704-1733.

Ren, J., Chen, Z., Zhou, Q., Tang, H., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. Int. J. Appl. Earth Obs. Geoinf., 10, 403–413.

Reynolds, C.A., Yitayew, M., Slack, D.C., Hatchinson, C.F., Huete, A., and Petersen, M.S., 2000. Estimating crop yields and production by integrating the FAO Crop Specific Water data and ground-based ancillary data. International Journal of Remote Sensing, 21, 3487–3508.

Richter, K., Atzberger, C., Vuolo, F., D'Urso, G., 2011. Evaluation of sentinel-2 spectral sampling for radiative transfer model based LAI estimation of wheat, sugar beet, and maize. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 4, 458–464.

Rudorff, B.F.T., and Barista, G.T., 1991. Wheat yield estimation at the farm level using TM Landsat and agrometeorological data. International Journal of Remote Sensing, 12, 2477–2484.

Sellers, P.J. 1985. Canopy reflectance, photosynthesis and transpiration. International Journal of Remote Sensing 6(8): 1335–1372.

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J. 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sensing of Environment 131, 215–231.

Senay, G.B., Ward, A.D., Lyon, J.G., Fansey, N.R., Nakes, S.E., and Brown, L.C., 2000. The relation between spectral data and water in a crop production environment. International Journal of Remote Sensing, 21, 1897–1910.

Steven, M.D., Biscoe, P.V., and Jaggard, K.W., 1983. Estimation of sugar beet productivity from reflection in the red and infrared spectral bands. International Journal of Remote Sensing, 4, 325–334.

Tamás, J., Nagy, A., Fehér, J., 2015. Agricultural biomass monitoring on watersheds based on remote sensed data. Water Science and Technology. 72(12): 2212-2220.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ., 8, 127–150.

Tucker, C.J., Holben, B.N., Elgin, J.H., Jr., McMurtrey, J.E., III., 1981. Remote sensing of total dry-matter accumulation in winter wheat. Remote Sens. Environ., 11, 171–189.

Tucker, C.J., Holben, B.N., Elgin, J.H., Jr., McMurtrey, J.E., III., 1980. Relationship of spectral data to grain yield variation. Photogramm. Eng. Remote Sensing, 46, 657–666.

Vuolo, F., Neugebauer, N., Falanga, S., Atzberger, C., D'Urso, G., 2013. Estimation of Leaf Area Index using DEIMOS-1 data: Calibration and transferability of a semi-empirical relationship between two agricultural areas. Remote Sens., 5, 1274–1291.

Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. Remote Sens. Environ., 108, 290–310.

Xiong, X., Chiang, K., Sun, J., Barnes,W., Guenther, B., Salomonson, V., 2009. NASA EOS Terra and Aqua MODIS on-orbit performance. Adv. Space Res., 43, 413–422.

Yin, H., Udelhoven, T., Fensholt, R., Pflugmacher, D., Hostert, P., 2012. How NDVI trends from AVHRR and SPOT VGT time series differ in agricultural areas: An Inner Mongolian case study. Remote Sens., 4(11), 3364-3389.

Yun Shao, Xiangtao Fan, Hao Lin, Jianhua Xiao, Ross, S., Brisco, B., Brown, R., and Staples, G., 2001. Rice monitoring and production estimation using multitemporal RADARSAT. Remote Sensing of Environment, 76, 310–325.

Zambrano, F., Lillo-Saavedra, M., Verbist, K., Lagos, O., 2016. Sixteen Years of Agricultural Drought Assessment of the BioBío Region in Chile Using a 250 m Resolution Vegetation Condition Index (VCI). Remote Sens., 8, 530.