THE USE OF LANDSAT 8 AND SENTINEL-2 DATA AND METEROLOGICAL OBSERVATIONS FOR WINTER WHEAT YIELD ASSESSMENT

S. Skakun[1,2], *B. Franch*[1,2], *E. Vermote*[2], *J.-C. Roger*[1,2], *N. Kussul*[3], *J. Masek*[4]

[1]Department of Geographical Sciences, University of Maryland, College Park MD 20742, USA
[2]NASA Goddard Space Flight Center Code 619, Greenbelt, MD 20771, USA
[3]Space Research Institute NAS Ukraine & SSA Ukraine, Kyiv 03680, Ukraine
[4]NASA Goddard Space Flight Center Code 618, Greenbelt, MD 20771, USA

ABSTRACT

This study focuses on winter wheat yield assessment from NASA's Harmonized Landsat Sentinel-2 (HLS) product and meteorological observations through phenological fitting. Vegetation indices (VIs), namely difference vegetation index (DVI), normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI2), extracted from satellite optical data, are fitted per pixel against accumulated growing degree days (AGDD) using a quadratic function. Accumulated VIs are correlated against winter wheat yields. Results show a better performance from DVI compared to NDVI and EVI2.

Index Terms— Wheat, yield, Landsat 8, Sentinel-2, Ukraine

1. INTRODUCTION

Combination of data acquired by Landsat 8 and Sentinel-2 remote sensing satellites can provide high temporal resolution (3-5 days) [1], which is critical for various applications requiring dense data time series. Previously, such (or better) high temporal resolution was available mainly for remote sensing sensors, which acquire daily data over Earth's surface, but at coarser spatial resolution (>250 m) [2]. The latter, for example, includes space-borne remote sensing sensors, such as MODIS, VIIRS, AVHRR, SPOT-VEGETATION. Taking into account an increased frequency of observations at moderate spatial resolution (<30 m), the assumption is that methods and models developed for generating products for coarse spatial resolution sensors can be ported to moderate spatial resolution sensors (for example, Landsat 8/OLI, Sentinel-2/MSI). However, the practice shows that such transition is not always straightforward due to larger data gaps because of clouds and uneven coverage, sensor characteristics and increased spatial resolution (at least at the order of 10, when going from 250 m to 30 m).

In this work, we focus on a crop yield assessment/forecasting application. The hypothesis is that satellite-based features, such as vegetation indices (VIs) or biophysical parameters derived at a single date or accumulated over some time period, can be correlated to crop yields [3]. Since the reference data on crop yields are mainly available at regional scale, the corresponding empirical models are built by averaging satellite-based features over those regions and correlating these derived variables to crop yields [4],[5],[6],[7]. It is assumed that there is a homogeneity within the region in terms of crops grown and agricultural practices applied and, therefore, the averaging should be performed for satellite data acquired at the same (or approximately the same) stages of crop growth, meaning that the data are normalized. This is usually the case for coarse spatial resolution remote sensing sensors, whose high temporal resolution enable high likelihood of obtaining cloud-free data over the Earth's surface [8]. This is also evidenced by multiple successful applications of applying coarse spatial resolution satellite data to crop yield assessment and forecasting [3]-[7]. However, this is not the case for moderate spatial resolution satellite data (<30 m). Irregular spatial coverage, when the area in question is covered by several "stripes" sensed at different times, and high revisit cycles lead to discrepancies in dates of cloud-free observations. The present study aims to address this issue by building empirical models for winter wheat yield assessment with Landsat 8 and Sentinel-2 data in Ukraine. We show that satellite data normalization is critical in building robust crop yield models. Not performing satellite data normalization may lead to poorer performance of the empirical crop yield model, which would be attributed not to the lack of correlation with satellite-derived variables, but rather to observation irregularities.

2. STUDY AREA & DATA DESCRIPTION

The study was performed for Kirohohradska oblast in Ukraine for 2016-2018. Oblast in Ukraine is a high-level administrative division of the country, and each oblast is further divided into districts. There are 24 oblasts in Ukraine and Autonomous Republic of Crimea. Kirovhradska oblast is located in the central part of Ukraine and composed of 21

districts with geographical area ranging from 65 to 165 thousand ha and cropland area ranging from 27 to 112 thousand ha. It is a highly intensive agricultural region with heterogeneous land cover land use (LCLU) classes [9], which include forest, grassland, barren soil, water, winter crops (wheat and rapeseed) and summer crops (mainly corn, sunflower, soybean and sugar beet).

Winter wheat is one of the major crops in Kirovhradska oblast accounting for 20% of production of all crops in the region. Winter wheat is mainly rain-fed in the region and usually planted in September-October. After dormancy during the winter, it re-emerges early spring reaching maturity by the end of June. Harvest of winter crops is typically undertaken in July.

Reference data on crop yield and harvested area at district level were collected from the Department of Agro-Industry Development of Kirovohrad State Administration (http://apk.kr-admin.gov.ua). The data were made available online as the harvest progressed and were based on farm surveys of all large agricultural enterprises (that account of more than 90% of all winter crops production in the region) and samples of household farms the same way as official statistics is collected [10]. The final estimates for winter crop yields and areas were available at the end of November and were used as reference in this study. Uncertainty of reference data shall not exceed 10%.

The main source of remote sensing data is NASA's HLS product [11]. HLS, which stands for Harmonized Landsat Sentinel-2, provides a Level-2 product of Nadir BRDF (Bidirectional Reflectance Distribution Function)-Adjusted surface Reflectance (NBAR) at 30 m. HLS combines, in a single data set, observations of the land surface from the Landsat 8's Operational Land Imager (OLI) and Sentinel-2's Multi-Spectral Instrument (MSI). MSI data are also spectrally adjusted to OLI spectral bands, so to generate a single harmonized dataset. We used HLS version 1.4. A detailed description of HLS product generation is given in [11].

The study area is covered by eight HLS tiles (each tile is approximately 110 km × 110 km), namely 35UQQ, 35UQP, 36UUV, 36UUU, 36UVV, 36UVU, 36UWV, and 36UWU. Overall, 3565 HLS scenes (L30 and S30) were downloaded, which covered the whole 2016-2018.

Meteorological data included a 2 m air temperature, which is used to calculate accumulated growing degree days (GDDs). The temperature data were extracted from NASA's Modern-Era Retrospective analysis for Research and Applications (MERRA2) product [12].

3. METHODOLOGY

Winter wheat yield mapping and assessment at regional scale consists of the two major steps: (i) winter crop mapping, so yields are estimated for relevant fields; (ii) yield assessment at 30 m spatial resolution. Fig. 1 illustrates all processing steps along with the input datasets.
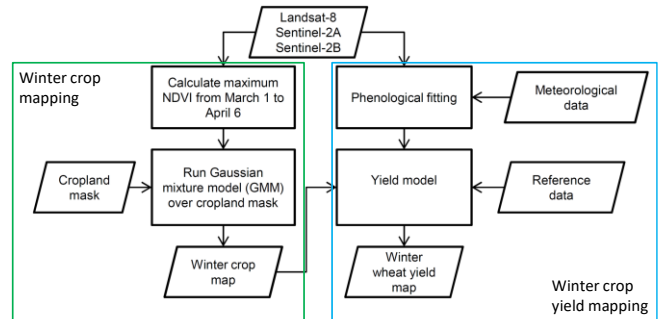

Fig. 1. Main algorithmic steps for winter wheat yield assessment using Landsat 8 and Sentinel-2 data.

For winter crop mapping, we adopted a previously developed approach for MODIS [13],[14] that allows automatic mapping of winter crops using *a priori* knowledge on crop calendar and without using reference (ground truth) data. The method is based on per-pixel estimation of the peak NDVI (hereafter referred as the metric) during early spring (or early fall depending on the Earth hemisphere), when winter crops have developed biomass, while other crops (spring and summer) have no biomass in that time period. The calculated metric will have high NDVI values for winter crops and low NDVI values for other crops. Then, the metric is fitted using a Gaussian mixture model (GMM) to automatically discriminate different crop types (winter versus others). Parameters of the GMM model are estimated using an expectation-maximization (EM) algorithm that is run for all pixels identified as cropland. In our study, we used a cropland layer derived from the land cover map generated for Ukraine at 30 m spatial resolution [15],[16]. The component with the largest mean, i.e. NDVI value, in the obtained GMM model is considered to belong to the winter crop class. Finally, the derived GMM model is applied to all cropland pixels, and a posteriori probability of the pixel belonging to the winter crop class is estimated in the final resulting map. Pixels, with the probability larger than 0.5, are considered as winter crops.

Vegetation indices, derived from HLS, were used for correlating with winter wheat yields, and included NDVI, DVI, and EVI2. Before deriving satellite based features, one has to ensure that satellite data are normalized. Though a combination of Landsat 8 and Sentinel-2 offers high frequency of observations, discrepancies in available cloud-free data, however, will still exist (Fig. 2).

Therefore, when building a crop yield model by averaging RS-derived features from different pixels, one has to ensure that the values from those pixels are taken from the same stage of crop growth. This normalization can be done through phenological fitting between VI and accumulated GDD (AGDD). To do such a fitting, it is advisable to have as much quality samples as possible to better capture the crop growth dynamics. The combined use of Landsat 8 and Sentinel-2 provides on average 2.7 and 1.6 times more cloud-free observations compared to Landsat 8 only and Sentinel-2 only, respectively. Fig. 3 shows an average number of cloud-

free observations for winter crop pixels during its crop calendar (March-July).
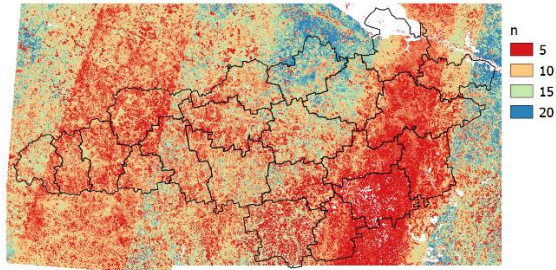

Fig. 2. Distribution of the number of cloud free observations from Landsat 8 and Sentinel-2 from March through the end of June 2018.
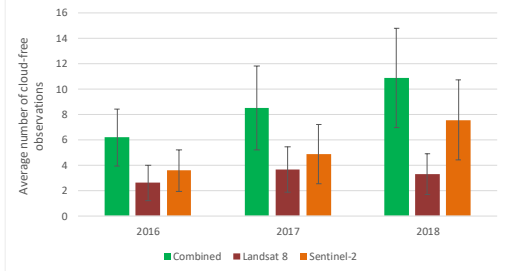

Fig. 3. Average number of cloud-free observations for winter crop pixels depending on satellite data usage. The number of pixels was taken from March until the end of June.

We used a quadratic model between VIs and accumulated GDD (AGDD) [17],[18]:

$$VI = a_0 + a_1 * AGDD + a_2 * AGDD^2. \qquad (1)$$

4. RESULTS

Fig. 4 shows results of validation for 2016-2018 combined, when comparing the areas of winter crops from official statistics and satellite-derived.
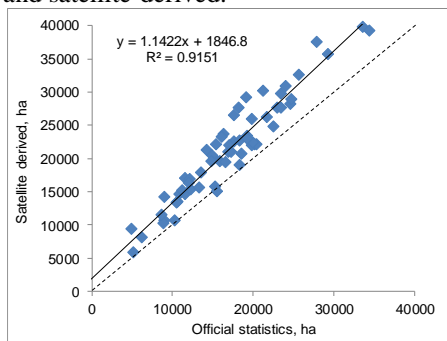

Fig. 4. Validation of winter crop maps for 2016-2018 for Kirovohrad oblast (Ukraine) at district level.

Fig. 5 shows an example of a quadratic relationship between DVI and AGDD.

The quality of fitting was assessed using two metrics, namely coefficient of determination (R^2) and root mean square error (RMSE) (Fig. 6). In general, all VIs (NDVI, DVI, EVI2) exhibited strong relationships between VIs and AGDD with averaged R^2>0.87.
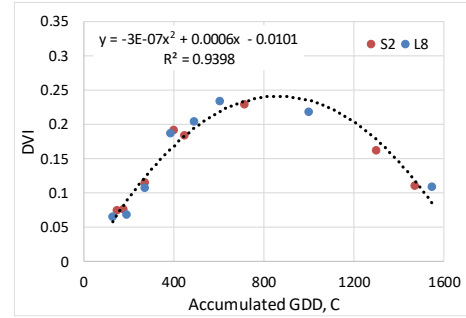

Fig. 5. Quadratic relationship between DVI and AGDD for 2017 for a winter crop pixel.
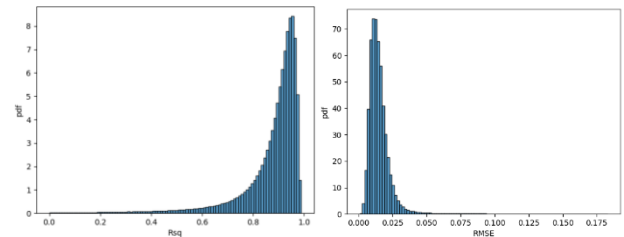

Fig. 6. Distribution of the coefficient of determination (R^2) and RMSE, when fitting DVI with accumulated GDD for 2018 for winter crop pixels. Number of winter crop pixels is 7,445,822. Average R^2=0.87±0.13, and average RMSE=0.015±0.011. Only pixels with >8 cloud-free observations during April-June period were considered.

Relationships between winter wheat yields and accumulated VIs (DVI, EVI2, NDVI) for a district level for 2016-2018 combined are shown in Fig. 7. Accumulation of VIs was performed from green up phase (with DVI_{min}=0.05, $NDVI_{min}$=0.15 and $EVI2_{min}$=0.15 that were determined empirically) until VI peak.

Better results were obtained for DVI, which showed the highest correlation and coefficient of determination for the relationships between accumulated DVI and crop yields.

5. CONCLUSIONS

This study showed the importance of satellite data normalization for crop yield assessment task. Combination of Landsat 8 and Sentinel-2 offers high temporal frequency of data, however discrepancies in acquisitions should be accounted, so satellite-derived features, e.g. VIs, from different pixels can be combined together. Normalization was performed through phenological fitting, whereas VIs were fitted to accumulated GDDs through quadratic function. All indices considered, namely DVI, NDVI and EVI2, showed a strong relationships yielding a coefficient of determination R^2>0.87. In terms of correlating with crop yields accumulated DVI showed a better performance (R^2=0.68) compared to NDVI (R^2=0.31) and EVI2 (R^2=0.55). Further works will be focused on incorporating all spectral bands into the yield models.
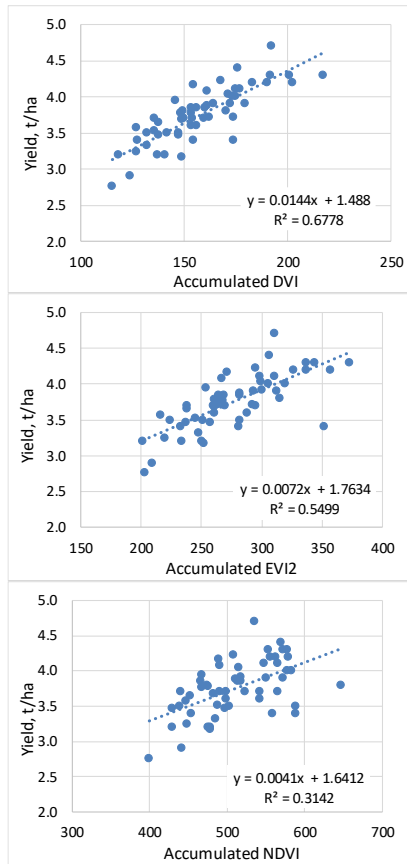
Fig. 7. Relationships between winter wheat yields and accumulated VIs from Landsat 8 and Sentinel-2, namely DVI, EVI2, and NDVI, at district level in Kirovohradska oblast for 2016-2018 combined.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] J. Li, and D. P. Roy, "A global analysis of sentinel-2A, sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring," *Remote Sens.*, vol. 9, no. 9, art. no. 902, 2017.

[2] C. O. Justice, *et al.*, "Land and cryosphere products from Suomi NPP VIIRS: Overview and status," *J. Geophys. Res.: Atm.*, vol. 118, no. 17, pp. 9753-9765, 2013.

[3] D. M. Johnson, "A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 52, pp. 65–81, 2016.

[4] I. Becker-Reshef, *et al.*, "A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data," *Remote Sens. Env.*, vol. 114, no. 6, pp. 1312–1323, 2010.

[5] B. Franch, *et al.*, "Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information," *Remote Sens. Env.*, vol. 161, pp. 131–148, 2015.

[6] F. Kogan, *et al.*, "Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 23, pp. 192–203, 2013.

[7] B. Franch, *et al.*, "Remote sensing based yield monitoring: Application to winter wheat in United States and Ukraine," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 76, pp. 112-127, 2019.

[8] A. K. Whitcraft, *et al.*, "Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations," *Remote Sens. Env.*, vol. 156, pp. 438-447, 2015.

[9] N. Kussul, *et al.*, "Crop inventory at regional scale in Ukraine: developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery," *Europ. J. Remote Sens.*, vol. 51, no. 1, pp. 627-636, 2018.

[10] F. J. Gallego, *et al.*, "Efficiency assessment of using satellite data for crop area estimation in Ukraine," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 29, pp. 22-30, 2014.

[11] M. Claverie, *et al.*, "The Harmonized Landsat and Sentinel-2 surface reflectance data set," *Remote Sens. Env.*, vol. 219, pp. 145-161, 2018.

[12] A. Molod, *et al.,* "Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2," *Geosci. Model Develop.*, vol. 8, no. 5, pp. 1339-1356, 2015.

[13] S. Skakun, *et al.,* "Early season large-area winter crop mapping using MODIS NDVI data and growing degree days information," *Remote Sens. Env.*, vol. 195, pp. 244–258, 2017.

[14] S. Skakun, *et al.*, "Combined Use of Landsat-8 and Sentinel-2A Images for Winter Crop Mapping and Winter Wheat Yield Assessment at Regional Scale," *AIMS Geosciences*, vol. 3, no. 2, pp. 163–186, 2017.

[15] M. Lavreniuk, *et al.*, "Regional retrospective high resolution land cover for Ukraine: Methodology and results," In *IGARSS2015*, 2015, New York: IEEE, pp. 3965–3968.

[16] N. Kussul, *et al.*, "Regional scale crop mapping using multi-temporal satellite imagery," *Int. Arch. Photogram., Remote Sens. Spat. Inform. Sci. (ISPRS Arch.)*, vol. 40, no. 7, pp. 45–52, 2015.

[17] W. G. Alemu, and G. M. Henebry, "Land surface phenologies and seasonalities using cool earthlight in mid-latitude croplands," *Environ. Res. Let.*, vol. 8, no. 4, art. no. 045002, 2013.

[18] K. M. De Beurs, and G. M. Henebry, "Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan," *Remote Sens. Env.*, vol. 89, no. 4, pp. 497-509, 2004.