Winter wheat yield assessment from Landsat 8 and Sentinel-2 data

*Sergii Skakun[1,2], Eric Vermote[2], Belen Franch[1,3], Jean-Claude Roger[1,2]

1. University of Maryland College Park, 2. NASA Goddard Space Flight Center, 3. University of Valencia

Combination of data acquired by Landsat 8 and Sentinel-2 remote sensing satellites can provide high temporal resolution globally (3-5 days) (Li & Roy, 2017), which is critical for various applications requiring dense data time series. Previously, such (or better) high temporal resolution at global scale was available mainly for remote sensing sensors, which acquire daily data over Earth' s surface, but at coarser spatial resolution (> 250 m) (Justice et al., 1998, 2013). 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 (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).

Consider a crop yield assessment/forecasting application (Becker-Reshef et al., 2010a). 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 (Johnson 2016, Meroni et al., 2016). 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 (Becker-Reshef et al. 2010b, Franch et al., 2015; Kogan et al., 2013). 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, which enable high likelihood of obtaining cloud-free data over the Earth' s surface (Whitcraft et al., 2015a, 2015b). This is also evidenced by multiple successful applications of applying coarse spatial resolution satellite data to crop yield assessment and forecasting (Becker-Reshef et al. 2010b, Franch et al., 2015; Johnson 2016; Kogan et al., 2013; Kolotii et al., 2015; López-Lozano et al., 2015; Mkhabela et al., 2011).

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. Therefore, satellite data should be normalized for the further processing. Though a combination of Landsat 8 and Sentinel-2 offers high frequency of observations, discrepancies in available cloud-free data, however, will still exist. With more and more applications transitioning to a higher spatial resolution, it is important to explore and analyze the effect of those discrepancies on the quality of the resulting products. In the case of crop yield assessment, one of the important questions is as follows: How do discrepancies in dates of satellite observations influence the performance of empirical crop yield models at regional scale? The present study aims to address this question by documenting results of the study on building empirical models for winter wheat yield assessment with Landsat 8 and Sentinel-2 data in Ukraine. Through simulation modelling and analysis of real datasets, 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.