https://ntrs.nasa.gov/search.jsp?R=20190001674 2020-05-09T10:14:43+00:00Z

ENHANCING REMOTE SENSING BASED YIELD FORECASTING: APPLICATION TO WINTER WHEAT IN UNITED STATES

B. Franch[1,2], *E. Vermote*[2], *S. Skakun*[1,2], *J.-C. Roger*[1,2], I. Becker-Reshef[1], *C. Justice*[1]

[1]Department of Geographical Sciences, University of Maryland, College Park MD 20742, USA
[2]NASA Goddard Space Flight Center Code 619, 8800 Greenbelt Road, Greenbelt, MD 20771, USA

ABSTRACT

Accurate and timely crop yield forecasts are critical for making informed agricultural policies and investments, as well as increasing market efficiency and stability. In Becker-Reshef et al. (2010) and Franch et al. (2015) we developed an empirical generalized model for forecasting winter wheat yield. In this study we present a new model based on the extrapolation of the pure wheat signal (100% of wheat within the pixel) from MODIS data at 1km resolution and using the Difference Vegetation Index (DVI). The model has been applied to monitor the national and state level yield of winter wheat in the United States from 2001 to 2016.

Index Terms— Agriculture, wheat, yield, MODIS

1. INTRODUCTION

Observations from the EOS/ MODIS sensors have several of the key qualities needed for global crop yield monitoring such as global, daily coverage at coarse spatial resolution (250m) and a suite of validated products. [1] developed an empirical but generalized crop yield model based on MODIS CMG (0.05 deg spatial resolution) data and applied it to Kansas and Ukraine. The model is based on the relationship between the Normalized Difference Vegetation Index (NDVI) value at the peak and the final yield value, corrected by the purity or the percentage of crop within the area studied. Later, [2] improved the timeliness of the model by including the Growing Degree Days (GDD) information. The model was applied satisfactorily (errors lower than 10%) at national level over the US, Ukarine and China. It also showed a good performance when applied to the AVHRR LTDR [3]. However, it is well known that the NDVI saturates when monitoring dense vegetation. In fact, recently, [4] when applying the [1] yield model to Landsat, showed that the NDVI saturates for yield values higher than 4MT/ha. In this work we improve the yield model both at national and regional scale by improving its accuracy and response to extreme events. Compared with [1] method, this model is based on two major improvements:

1. we develop a method to derive the wheat surface reflectance, that is, at AU level we estimate the signal as if the pixels were covered 100% by wheat.
2. we apply the model to 1km resolution MODIS data instead of CMG spatial resolution. In this way, we are able to improve the yield estimation at subnational level.
3. we consider a vegetation index (DVI) that is better correlated with the yield than the NDVI.

2. STUDY AREA & DATA DESCRIPTION

The study was performed for the US and Ukraine from 2001 until 2016. The US is one of the main producers and exporters of wheat globally. Wheat is produced in almost every state in the United States and winter wheat varieties dominate US production, representing between 70 and 80% of the total. The main class is Hard Red Winter Wheat, which is grown primarily in the Great Plains, with Kansas being the largest producing state.

Ukraine, is another critical player in the global wheat market. Wheat is grown all across the country, although the central and southern regions are the key growing areas (Forest-Steppe and Steppe zones). About 95% of Ukraine wheat production is winter wheat, planted in the fall and harvested during July and early August of the following year. Generally, wheat is not irrigated in this country. Ukraine produces mostly the Hard Red Winter Wheat.

In each country we estimate the wheat yield for each administrative unit. Thus, we work at different spatial scales depending on the data availability. In the case of the US, we work at county level and in the case of Ukraine the administrative units are oblasts.

For the US, we use the official archive of county-level statistics on yield, area harvested, and production that is available from the USDA National Agricultural Statistics Service (NASS) Quick Stats database. For Ukraine, oblast-level crop statistics were obtained from the State Statistical Committee of Ukraine (SSC) for winter wheat area harvested and yield.

In this work the crop type masks are critical in order to isolate the wheat signal. For the US, we used crop type masks from the Cropland Data Layer (CDL) produced by NASS. We studied the main land surfaces surrounding the

wheat fields in the US and used a total of eight different crop type masks: winter wheat, spring wheat, corn, soybean, alfalfa, potato, grassland and forest. For Ukraine, we built winter wheat crop masks using an approach that we developed for MODIS [5] that allows automatic mapping of winter crops using *a priori* knowledge on crop calendar and without using reference (ground truth) data.

3. METHODOLOGY

3.1. Extrapolation of the wheat signal

For each AU and for a given date, the total DVI signal from each pixel, i, can be written as the sum of the DVI signal from the wheat (DVI_{wheat}) multiplied by the percentage of wheat within the pixel or wheat purity (Wpct) and the DVI from other surfaces within the pixel (DVI_{others}) multiplied by the remaining percentage (1- Wpct) (1).

$$DVI_i = DVI_{i,wheat} \cdot Wpct_i + DVI_{i,others} \cdot (1 - Wpct_i) \quad (1)$$

Assuming that the DVI_{wheat} and the DVI_{others} remain constant through an AU for a given date, the DVI of pixel i can be written as:

$$DVI_i = (DVI_{wheat} - DVI_{others}) \cdot Wpct_i + DVI_{others} \quad (2)$$

Which has the structure of a linear regression function where the term $(DVI_{wheat} - DVI_{others})$ is the slope and DVI_{others} is the intercept and we can easily derive the DVI signal from the wheat. In the case of the US, where we consider a total of 8 different classes (plus the "others" contribution) equation 1 is written as:

$$DVI_i = DVI_{i,wheat} \cdot Wpct_i + DVI_{i,corn} \cdot Cpct_i + \cdots + DVI_{i,others} \cdot (1 - Wpct_i - Cpct_i - \cdots) \quad (3)$$

that also has structure of a linear regression function.

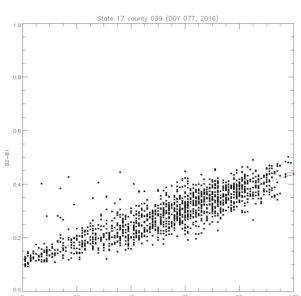


Figure 1. Linear regression of the DVI_i versus the wheat purity ($Mpct_i$) through the Harper county for DOY 121 of 2008

Figure 1 shows an example of the linear regression between the DVI_i and the wheat purity, $Wpct_i$, (black dots) the Day of the Year (DOY) 121 of 2008 in the Harper county (Kansas).

3.2. Calibration of the yield model

The U.S.

We calibrated the yield model based on the years when the CDL covered all the country, that is from 2008 to 2016. We just consider in the study those states with pixel wheat purities higher than 40%. We divided the calibration of the model based on the different two main wheat classes within the states considered: SWW and HRW.

Figure 2 left shows the HRW calibration over the states of Kansas, Colorado, Oklahoma, Montana, Nebraska, South Dakota and Texas. Figure 2 right shows the SWW calibration over the states of Washington, Oregon, Idaho, Illinois and Missouri.

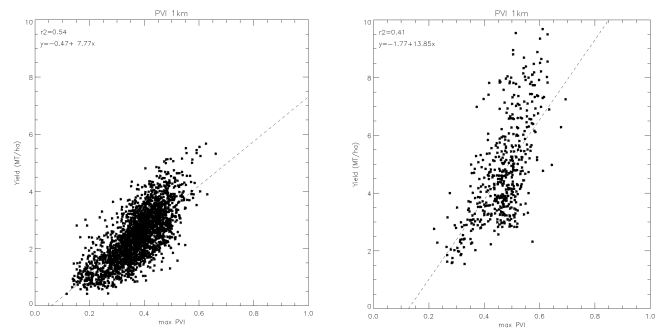


Figure 2. HRW (left) and SWW (right) calibration in the US using county level statistics from 2008 to 2016.

Given that the Central Great Plains is prone to severe droughts we explored the use of the EF in the model. Figure 3 left shows the validation at county level of the linear regression of the yield versus the DVI and Figure 3 right shows the county level validation of the same linear regression but including the EF. Kansas and Colorado show an improvement of the yield estimations when including the EF information while the other states don't show any improvement. Thus, we just consider the EF on those two states.

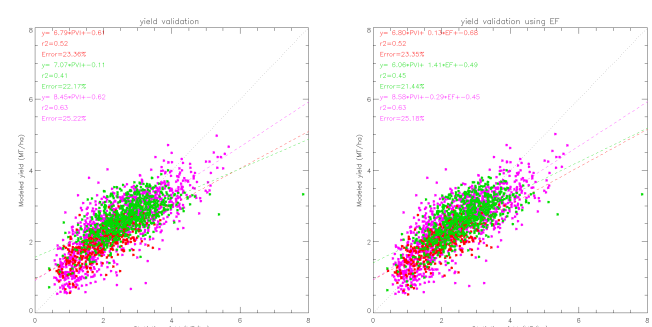


Figure 3. County level validation using just DVI (left) or using DVI and EF (right).

Ukraine

In contrast to the US, there is limited information in Ukraine about the main planted wheat classes. However, given that the yield model developed in this study is based on the dependency of the DVI at the peak with the yield, we noticed that this dependency in Ukraine was different for the two main climatic regions. Thus, we divided the calibration into two major classes: the North and the South (Figure 4).

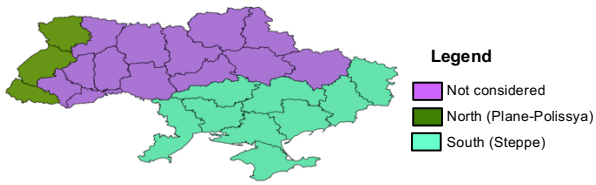


Figure 4. Ukaine main climatic regions considered to calibrate the model.

Figure 5 shows the calibration of the model for the north (a,b) and the south (c). The statistics showed that the north drastically increased the yields from 2 to 4MT/ha (2001-2013) to 4 to 6 MT/ha (2014-2016). Possible causes of such increase were the chage of wheat variety or the improve of fertilizers. Therefore, we considered 2014 to 2016 as a different class of wheat for calibration purposes (Figure 5b)

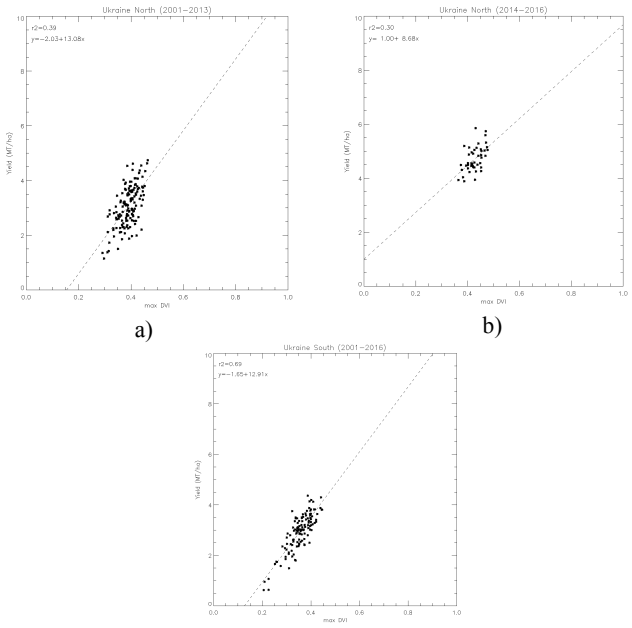


Figure 5. Calibration of the model in Ukraine.

4. RESULTS

The U.S.

Figure 6 shows the state level error (top) and correlation (bottom) of the modeled yield when compared to the official statistics from 2001 to 2016. Generally, the model shows a good performance in the most important wheat producing states (Kansas, Oklahoma, Colorado) with low errors and high correlation. However, the figure shows worse results in Montana and Texas mainly caused by the accuracy of the wheat mask during 2001-2006.

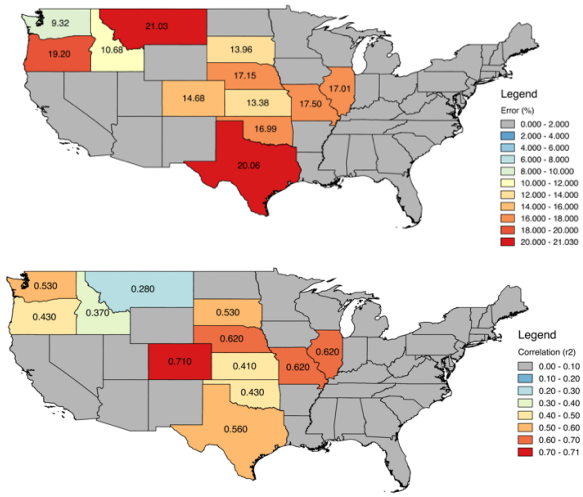


Figure 6. Error (top) and correlation coefficient (bottom) at state level in the U.S.

Figure 7 shows the national validation of the model in the U.S. The statistics show a good performance of the model with low error (4.68%) and good correlation coefficient (0.68) despite the low variability of the yield during the time series considered.

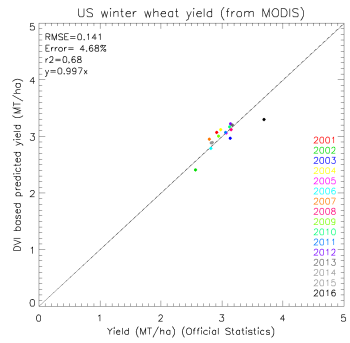
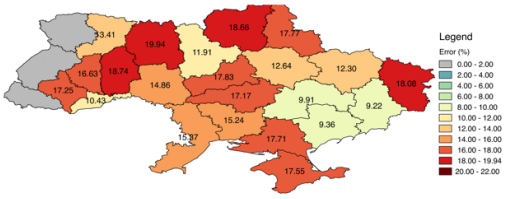


Figure 7. U.S. national validation.

Ukraine

Figure 8 shows the oblast level error (top) and correlation (bottom) of the yield from 2001 to 2016. The southern oblasts, that are also the major wheat producing oblasts, show a better performance (lower errors and higher correlation) than the northern oblasts.



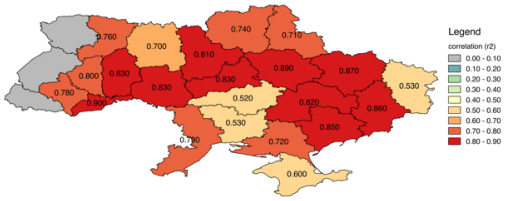
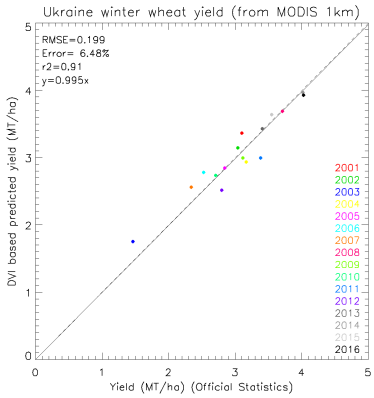


Figure 8. Error (top) and correlation coefficient (bottom) at oblast level in Ukraine.

Figure 9 shows the national validation of the model in Ukraine. The statistics show a low error of 6.48% and a really good correlation of 0.91, responding successfully to extreme events (high and low yields).



5. CONCLUSIONS

This study presents a new remote sensing based yield model based on the DVI at the peak and the unmixing of the wheat signal. The model is efficient in responding to extreme conditions (high and low yields) with good correlation and good accuracy (error < 7%) at national level both in the US and Ukraine. The model provides also good results (error < 15%) at subnational level in the major wheat producing oblasts/states.

6. REFERENCES

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

[2] Franch B, Vermote EF, Becker-Reshef I, et al. (2015) "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 Sensing of Environment* 161: 131–148.

[3] Franch, B., Vermote, E., Roger, J.C., Murphy, E., Becker-Reshef, I., Justice, C., Claverie, M., Nagol, J., Csiszar, I., Meyer, D., Baret, F., Masuoka, E., Wolfe, R. and Devadiga, S., (2017) "A 30+ year AVHRR Land Surface Reflectance Climate Data Record and its application to wheat yield monitoring." *Remote Sensing*, 9, 296; doi:10.3390/rs9030296

[4] 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.

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