The ARYA crop yield forecasting algorithm: Application to the main wheat exporting countries

B. Franch [a,b,*], E. Vermote [c], S. Skakun [b,c], A. Santamaria-Artigas [b,c], N. Kalecinski [b,c], J.-C. Roger [b,c], I. Becker-Reshef [b], B. Barker [b], C. Justice [b], J.A. Sobrino [a]

[a] Global Change Unit, Image Processing Laboratory, Universitat de Valencia, Paterna, Valencia 46980, Spain
[b] Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA
[c] NASA Goddard Space Flight Center Code 619, 8800 Greenbelt Road, Greenbelt, MD 20771, USA

ABSTRACT

Wheat is the most important commodity traded in the international food market. Thus, accurate and timely information on wheat production can help mitigate food price fluctuations. Within the existing operational regional and global scale agricultural monitoring systems that provide information on global crop yield and area forecasts, there are still fundamental gaps: #1. Lack of quantitative Earth Observation (EO) derived crop information, #2. Lack of global but detailed (national or subnational level) and timely crop production forecasts and #3. Lack of information on forecast uncertainties. In this study we present the Agriculture Remotely-sensed Yield Algorithm (ARYA) an EO-based method, advancing the state of EO-data application and usage (addressing gap #1) to forecast wheat yield. The algorithm is based on the evolution of the Difference Vegetation Index (DVI) using MODIS data at 1 km resolution and the Growing Degree Days (GDD) from reanalysis data. Additionally, we explore how Land Surface Temperature (LST) can be included into the model and whether this parameter adds any value to the model performance when combined with the optical information. ARYA is implemented at the national and subnational level to forecast winter wheat yield in the main wheat exporting countries of US, Russia, Ukraine, France, Germany, Australia and Argentina from 2001 to 2019 (covering over 70% of wheat exports globally) in a timely manner by providing daily forecasts (addressing gap #2). The results show that ARYA provides yield estimations with RMSE's within 0.3 ± 0.1 t/ha at national level and 0.6 ± 0,1 t/ha at subnational level after Day Of the Year (DOY) 140 (mid May) in the Northern Hemisphere and DOY 280 (beginning of October) in the Southern Hemisphere. This means that ARYA can provide crop yield estimates of wheat yield with 5–15% error at national and 7–20% error at subnational level starting from 2 to 2.5 months prior to harvest.

1. Introduction

Despite significant improvements in technology that the agriculture sector has experienced in the last century, food production are still highly dependent on climate (Rosenzweig et al., 2001). In the context of a changing climate average yields are predicted to decrease from 30% up to 82% depending on the warming scenario (Schlenker & Roberts, 2009). When extreme weather events impact major food producing countries, these can often result in production shortfalls and dramatic food price spikes. Global food price crises are generally unpredictable in their occurrence and often in their duration. Droughts, floods or extreme temperatures are the main extreme weather events that cause crop

failure and reduce global food supply with a consequent increase in prices (Tadesse et al., 2014).

Enhanced, quantitative and timely information can improve agricultural analysis and assessment and support improved decision making from international to local level. In this context, better informed decisions can contribute to increased food security, greater sustainability, and greater resilience in the agriculture sector. In 2011, the Group of Twenty (G20) Agricultural Ministers developed the G20 Action Plan on Food Price Volatility and Agriculture that launched the Agricultural Market Information System (AMIS; URL1), and the Group on Earth Observations Global Agricultural Monitoring initiative (GEOGLAM; URL2). AMIS is an international platform whose main goal is enhancing

food market transparency and policy response for food security. GEO-GLAM's main objective is to strengthen the use of Earth Observations (EO) internationally to produce and disseminate relevant and timely information on agricultural production at national, regional, and global scales (Becker-Reshef et al., 2019, 2018). GEOGLAM and AMIS requirements are lately evolving from qualitative to quantitative metrics based on EO of within season production forecasts in a timely manner and monitoring their evolution through time.

Crop production estimation is based on two components: crop area and crop yield estimation. For most purposes related to food market and food security the principal need is information on production. Area estimation is needed in areas with strong crop area inter-annual variability and relatively stable yields. On the contrary, the priority will be on yield for countries were crop area is stable or predictable and yield change is stronger (Gallego, 2008). The main purpose of leveraging EO information is reducing the uncertainty of these estimations.

The main goal of global and regional scale agricultural monitoring systems is providing up-to-date information on food production to decision makers in support of global and national food security. A number of systems now exist that fill this role, but just three provide freely available production forecasts at the global scale:

- The Foreign Agricultural Service (FAS) of the United States Department of Agriculture (USDA) provides information from multiple sources including EO which is used by agricultural economists and researchers to predict global crop production for all major commodities, for all countries in a monthly report.
- The Global Information and Early Warning System (GIEWS) from the Food and Agriculture Organization (FAO). The system provides quarterly global reports of food crop production and market prices at continental scale, and more specific regional reports based on intelligence from FAO's regional and country offices (Rojas, 2015).
- CropWatch, which is led by the Institute of Remote Sensing and Digital Earth at the Chinese Academy of Sciences, evaluates national and global crop production. Since 2013, CropWatch has been releasing quarterly and annual bulletins internationally covering most of the prominent food-producing countries in the world from global to sub-national (for the nine largest countries) level. These bulletins include predictions of crop conditions and production for the main commodities (wheat, maize, rice, and soybean) (Wu et al., 2014).

In Europe, the Monitoring Agricultural Resources Unit (MARS) of the European Joint Research Centre (JRC) provides information on the status of crops in the EU and neighboring countries and forecasts crop yields in monthly bulletins (Baruth et al., 2008; López-Lozano et al., 2015; Rembold et al., 2019).

Despite the improvements in access to high resolution satellite imagery over the last decade and the use of numerous remote-sensing based products by the different systems, there are still fundamental gaps (Fritz et al., 2019). Next, we identify three gaps addressed in part by this study:

#1. Lack of quantitative EO-derived crop information. Current agriculture monitoring systems generally use EO to assist qualitatively assessments of crop conditions and identifying anomalies which are then used by analysts to infer yield, area and production reductions. However, these systems do not provide the quantitative area and production forecasts, which are needed. Since the early days of EO satellites, agricultural monitoring was considered a major targeted application of these technologies (Macdonald, 1984). However, it is widely recognized that EO could play a much stronger and transformational role in informing better policies, programs, and more effective implementation to benefit food security and agricultural practices (Fritz et al., 2019).

#2. Lack of global but detailed (national or subnational level) and timely crop production forecasts. Current global systems provide at most regular monthly reports of national level yield or production forecasts. Monthly reports are a very good source of information for the commodity markets and decision-making around the world (from production, marketing, domestic and global trade, among others). However, higher frequency global reports that contain production forecast regionally detailed information would potentially minimize the gap between one report and the update (currently one month period), reducing the "surprise effect" that some reporting has on the market. Minimizing the "surprise" effect, especially in times when an important weather event is occurring to a crop in an important production region, or in a tight situation of stock-to-use ration, would potentially alleviate excess price volatility. For instance, severe weather events in major food producing countries were a major factor in recent (2007/8; 2011/12) food-price spikes, pushing the number of food insecure people to over 1 billion and leading to civil unrest, economic strife and geopolitical tension (Anderson et al., 2014; Janetos et al., 2017). The most recent catastrophic droughts in Southern (2015–16) and Eastern (2017) Africa led to major food shortages (URL3), requiring a timely international response. These events underscore the importance of timely and accurate production forecasts in stabilizing markets, mitigating food supply crises, and mobilizing humanitarian assistance.

#3. Very limited information on forecast uncertainties. Current forecasting systems do not report uncertainties of their outputs, and it is not clear the targeted uncertainty for yield estimations.

There are a growing number of studies based on exploiting remote sensing data for agriculture monitoring. In this context, Weiss et al. (2020) provides an extensive overview of the recent research developments of how remote sensing can be used for agricultural applications. Focusing on crop yield estimation and forecasting, remote sensing data does not directly provide these parameters. Thus, most methods rely on establishing empirical relationships between yield statistics (based on surveys or field measurements) and remote sensing parameters that can be retrieved from satellite data. The EOS/ Moderate Resolution Imaging Spectroradiometer (MODIS) sensors provide EO data globaly with a daily (twice a day) coverage at coarse spatial resolution (250 m) and has a suite of validated products. All these qualities, and especially a good temporal coverage, are key for agriculture monitoring applications. Among many empirical yield models developed based on MODIS data (e.g. Doraiswamy et al., 2005; Mateo-Sanchis et al., 2019; Sakamoto et al., 2013), Becker-Reshef et al. (2010) developed a crop yield model based on its relationship with the Normalized Difference Vegetation Index (NDVI) seasonal peak from MODIS CMG (0.05 deg spatial resolution). The model was successfully applied to Kansas and Ukraine. Later, Franch et al. (2015) improved the timeliness of the model by including information on the NDVI evolution with the Growing Degree Days (GDD). The model was applied to forecast wheat yields in the US, Ukraine and China at national level with errors lower than 10% up to 2 months before harvest. The method also showed a good performance when applied to the AVHRR LTDR data set (Franch et al., 2017). At coarse resolution the signal retrieved by each pixel is a mixture of the targeted crop and the surrounding surfaces. The previous works cited solve this issue by averaging the NDVI signal from the pixels that reach the maximum percentage of the crop within each Administrative Unit (AU), which limits the spatial representativity of the whole AU. Additionally, it is well known that the NDVI saturates when monitoring dense vegetation (Baret & Guyot, 1991; Buschmann & Nagel, 1993; Gitelson et al., 2003; Sellers, 1985). Therefore, Franch et al. (2019) presented a model based on the Difference Vegetation Index (DVI) from MODIS data at 1 km resolution and yearly wheat crop type maps (Skakun et al., 2017) to un-mix the DVI signal of all pixels within an AU to a 100% wheat signal. This average wheat signal of each AU is used to build a linear regression model using DVI seasonal amplitude

and length and the average of the Evaporative Fraction (EF) 30 days after the DVI peak as inputs. The model was applied from 2001 to 2017 to estimate the winter wheat yield in the US and Ukraine both at national and subnational level. At the subnational level the model showed a Root Mean Square Error (RMSE) lower than 0.6 t/ha (15–18%). At national level the model showed a RMSE of 0.11 t/ha (3.7%) for the US and 0.27 t/ha (8.4%) for Ukraine. However, this is not a forecast model and it must be applied at the end of the season in order to retrieve the three regressors. Next, we briefly describe the main findings of some works on wheat yield forecasting and their reported metrics that are focused on the countries analyzed in this manuscript. In Ukraine, Kogan et al. (2013) used NDVI data from MODIS to forecast winter wheat yield at *oblast* (sub-national) level with RMSE ranging from 0.5 to 0.8 t/ha from 2 to 3 months prior to harvest. In Argentina, (Lopresti et al., 2015) developed an early method of wheat yield estimation in Northern Buenos Aires province based on MODIS NDVI data during 2008–2011 that allowed predicting wheat yields 30 days before harvest with a determination coefficient of 0.75 at department level. In Europe, Pagani et al. (2017) improved the Crop Growth Modeling System (CGMS) of the European Commission JRC, within MARS, to forecast maize, wheat, and barley yield in Europe. Their results when applied from 1995 to 2013 to forecast wheat yield from flowering to maturity stages ranged between a relative RMSE of 3.86–4.62% in France and 4.26–5.66% in Germany. Recently, Kamir et al. (2020) used machine-learning regression methods based on climate records and satellite image time series from 2009 to 2015 to estimated wheat yields in the Australian wheat belt. Though this method was applied at pixel level, the performance at national level explained showed a determination coefficient of 0.73 and a RMSE of 0.59 t/ha. Note that these performance metrics correspond to end-of-season estimates.

Finally, looking for additional variables that can be retrieved from EO, in this work we explore the use of the satellite-derived Land Surface Temperature (LST) to complement the optical data and to correct the effects of extreme events on the crop yields. The LST is generally used in the literature to derive downstream products to monitor water deficit (Kogan, 1997; Anderson et al., 2011; Anderson et al., 2016). There are few studies using the LST directly in crop yield forecasting models. Originally, Idso et al. (1979) successfully combined the stress-degree-day (SDD) defined as the difference between the canopy temperature and the air temperature approximately 1 m above the crop, with the GDD to predict spring wheat yields in 11 field plots in Arizona. More recently, Johnson (2014) built a regression tree-based modeling technique combining the NDVI and daytime LST to forecast maize and soybean yield at the county level in the United States. The study found that the LST was negatively correlated with both crops.

In this work, we present the Agriculture Remotely sensed Yield Algorithm (ARYA) forecasting algorithm and we contribute to address the aforementioned gaps by

- Amplifying the use of EO data in end user decision making, focused on agriculture.
- Advancing the state of the art of yield modelling at the national and subnational levels
- Consolidating the yield model into a forecasting algorithm to provide timely national and subnational yield forecasts over the major exporting countries.
- Tracking and reporting the accuracy of the yield estimates.

The main science questions that we will answer in this study are: can we quantitatively assess the crop yield in the main wheat exporting countries and how early in the growing season can we provide a good (with an associated error) yield forecast in each country?

The ARYA forecasting algorithm is based on moderate resolution data acquired by MODIS rescaled to 1 km spatial resolution, using the crop signal unmixing method presented in (Franch et al., 2019) and exploring the use of the LST in the algorithm. The forecasting of the optical parameters is based on the evolution of the optical signal with the accumulated GDD information.

2. Study area and data description

2.1. Study area

In this study we analyze the main wheat exporting countries: the US, France, Germany, Ukraine, Russia, Australia and Argentina, which together account for c. 70% of global wheat exports (Fig. 1 top). Wheat is the most important cereal crop traded on international markets and winter wheat constitutes approximately 80% of global wheat production. Fig. 1 (bottom) shows the evolution of wheat yield at the national level in each country analyzed and during the period 2001–2019. During this period, all countries showed some variability and a positive trend, with Russia, Argentina and especially Ukraine showing the largest increases in yields. Next, we describe the main characteristics of wheat production in each country.

Wheat yield levels in the US have remained relatively stable during the past 40 or more years. Five major classes are grown in the US: Hard Red Winter (HRW), Soft Red Winter (SRW), Hard Red Spring (HRS), White and Durum. Most US wheat is grown in the Great Plains from Texas to North Dakota. The large majority of winter wheat is rainfed and approximately seven percent is irrigated.

In Ukraine wheat is grown mainly on the central, southern and eastern regions. Most of the wheat planted in Ukraine is winter wheat. Generally, wheat is not irrigated in this country. Ukraine produces mostly the HRW class. The amount of winterkill varies widely from year to year, from 2% in 1990 to 65% in 2003, when a persistent ice-crust smothered the crop. Ukraine is characterized by highly variable wheat and coarse grains productivity. On average, every three years, wheat production changes by 20%. Fig. 1 shows Ukraine as the country with the largest yield variation amongst the major wheat exporting countries with a positive trend of 0.09 t/ha per year.

Russia was the largest wheat exporter in 2020 according to data from U.S. Department of Agriculture's Foreign Agricultural Service (FAS USDA). The varieties of winter wheat are mostly cultivated in the Caucasus, in the Central Black Earth region, and in the Volga region while spring wheat (whose yield is generally half that of winter wheat) is mostly planted in the Eastern region (Western Siberia). Major constraints to production in Russia include droughts, diseases, lodging and winterkill of plants. Most of the wheat cultivation is rainfed. Thus, droughts also occur in some years and can result in serious crop losses. This occurred in 2010, when a large fraction of the wheat area in Russia experienced extraordinarily high temperatures through the summer and leading to a drop of about 40% in production from previous years' levels.

The major wheat-growing countries in Europe in order of size of production are France, Germany, the United Kingdom, Italy, Spain and Portugal. In these countries, the most common wheat class is SRW of generally low quality due to its low protein content and poor gluten content. However, these countries have the highest yields in the world.

Wheat production in Australia accounts for 55% of the total national cropland. Most wheat is grown in the arcuate belt of land curving across the eastern and southern regions where winter rainfall is sufficient to produce a crop. Spring wheat is grown as a winter crop sown in the autumn (May to June) and harvested in early summer (November to December). Wheat yields in Australia are low and highly variable primarily due to extreme fluctuations in annual rainfall.

Over 85% of the grain production in Argentina takes place in the Pampas region, which covers the central provinces of Santa Fe, Cordoba, Buenos Aires, Entre Rios and La Pampa. Argentina has a favorable climate for rainfed crop production, and despite the potential for irrigation expansion, current irrigated area is negligible. In Argentina, environmental factors affecting wheat during the growth cycle are early and late heat, early and midterm drought, frosts at flowering and rains at harvest.
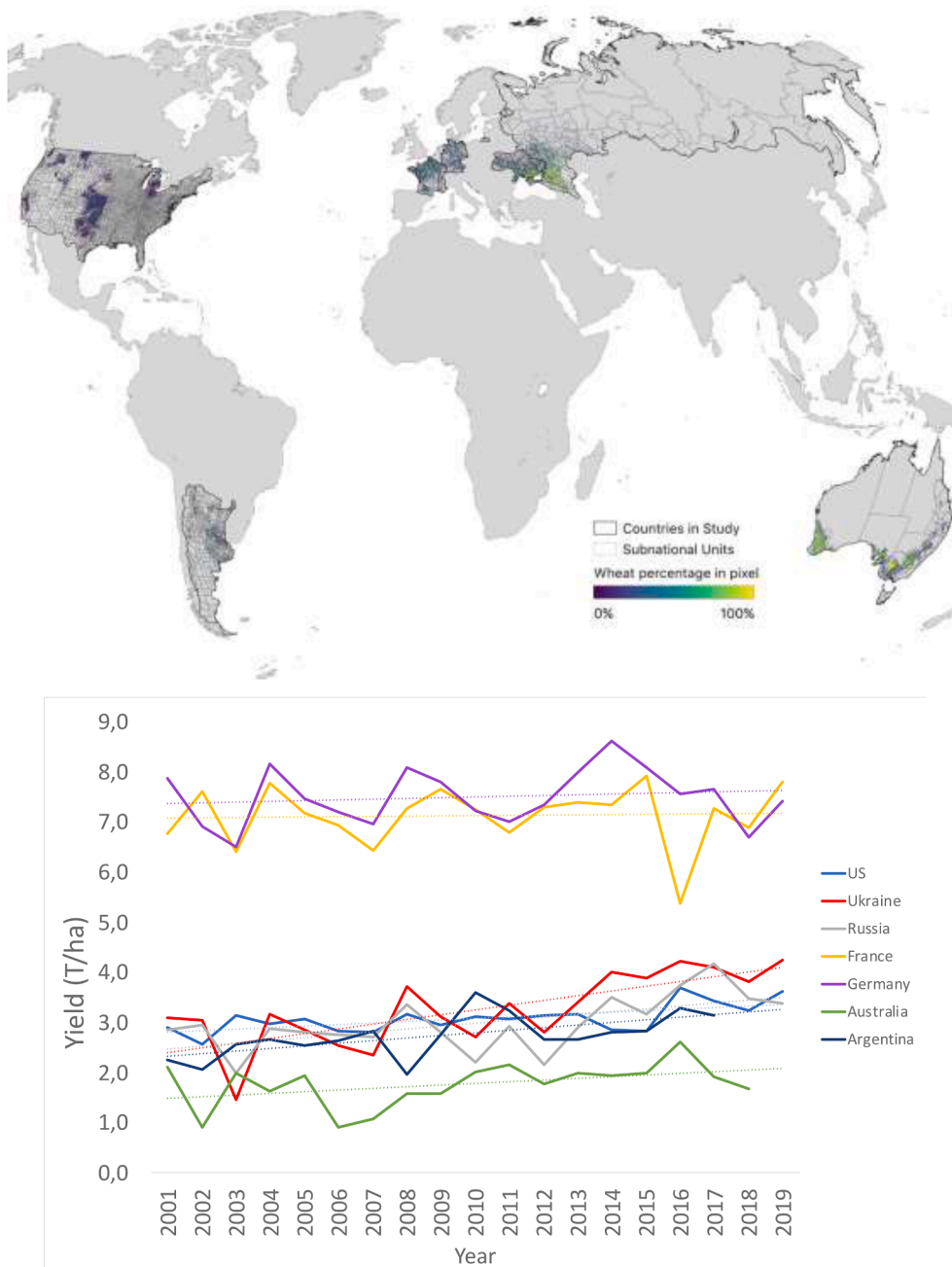
Fig. 1. Top: countries and subnational units divisions as well as percentage wheat across the countries considered. Bottom: trends of the national yield in the countries considered over the period 2001–2019.

2.2. Crop statistics

We gathered the official national and subnational level statistics on yield, area harvested, and production that is available for each country from the different national agriculture agencies shown in Table 1.

2.3. Crop type masks

Crop type masks are needed to unmix the wheat signal. In the US we used the Cropland Data Layer (CDL) product produced by the National Agricultural Statistics Service (NASS) (D M Johnson & Mueller, 2010). We studied the main land surfaces surrounding the wheat fields in the US and considered eight different crop types: winter wheat, spring wheat, soybean, corn, potato, alfalfa, grassland and forest.

In the other countries, the crop type masks were generated using the

approach presented in (Skakun et al., 2017) that allows automatic mapping of winter crops based on MODIS data using a priori knowledge of crop calendars and without using ground truth data. These maps were produced yearly for each country. When compared to the CDL for Kansas and ground measurements for Ukraine, crop masks produced with this method showed accuracies > 90% when generated 1.5–2 months before the harvest (Skakun et al., 2017).

2.4. EO data

In this study, we use MODIS daily surface reflectance Collection 6 data (M{O,Y}D09GQ) at 250 m resolution and gridded in the sinusoidal projection. Besides, the product M{OY}D09GA was also downloaded to extract the observation/illumination geometry of each image. Additionally, the 250 m surface reflectance data were re-sampled to 1 km

Table 1

Official wheat statistics considered and their source of information. *These years in the Southern Hemisphere correspond to the planting year. Acronyms used in the table: Russia United Interdepartmental Information and Statistical System (EMISS), National Agricultural Statistics Service (NASS), State Statistical Committee of Ukraine (SSC), Sistema Integrado de Información Agropecuaria de Argentina (SIIA).

Country	Data source	Administrative unit (AU)	Number of AU	Years considered in the study	Average national yield $\pm$ standard deviation (t/ha)	Average date of flowering (DVI peak)
Russia	EMISS	Oblast	18	2001–2019	3.1 ± 0.3	152 ± 9
United States	USDA/NASS	County	249	2001–2019	3.0 ± 0.5	130 ± 16
France	Ministry of Agriculture and Food	Province	14	2001–2019	7.1 ± 0.6	140 ± 13
Australia	Australian Bureau of Statistics	Districts	13	2001–2018*	1.7 ± 0.6	267 ± 7
Ukraine	SSC	Oblast	17	2001–2019	3.3 ± 0.7	149 ± 5
Argentina	SIIA	District	101	2001–2017*	2.4 ± 0.9	280 ± 10
Germany	Eurostat	Province	13	2001–2019	7.5 ± 0.5	147 ± 6

spatial resolution by aggregating 4x4 pixels to avoid the inaccuracies of incorrect registration of the 250 m product (Breon et al., 2015). The solar and view geometry effects on the surface reflectance were corrected based on the VJB method (Vermote et al., 2008; Franch et al., 2014). Following this method, the normalized surface reflectance (ρ^N), that is the Bidirectional Reflectance Distribution Function (BRDF) corrected surface reflectance, is written as

$$\rho^N(45,0,0) = \rho(\theta_s,\theta_v,\phi)\frac{1 + V\cdot F_1(45,0,0) + R\cdot F_2(45,0,0)}{1 + V\cdot F_1(\theta_s,\theta_v,\phi) + R\cdot F_2(\theta_s,\theta_v,\phi)} \quad (1)$$

where ρ is the surface reflectance, θ_s is the sun zenith angle, θ_v is the view zenith angle, ϕ is the relative azimuth angle, F_1 is the volume scattering kernel, based on the Ross-Thick function derived by (Roujean et al., 1992) but corrected for the Hot-Spot process proposed by (Maignan et al., 2004), F_2 is the geometric kernel, based on the Li-sparse model (Li & Strahler, 1986) but considering the reciprocal form given by (Lucht, 1998), V represents the volume parameter since it is linked to the Volume kernel and R represents the roughness parameter since it is linked to the geometric kernel. These parameters (V and R) represent the shape of the BRDF.

The LST was extracted from the official MODIS LST Collection 6 official product (M{O,Y}D11A1) that provides daily LST and surface emissivity at 1 km spatial resolution. Fig. 2 (top) shows an example of the LST evolution in Harper County Kansas, USA. While the cloud mask layer is useful to detect cloudy pixels at the time of the satellite overpass, the effects of clouds on LST, decreasing its values, can last well after the cloud has passed. Therefore, to minimize this effect we dilated the cloud mask using a buffer of 5 pixels around each cloudy pixel (Fig. 2 bottom).

Finally, near-surface air temperature (T2M) was extracted from the MERRA2 reanalysis inst1_2d_asm_Nx product. The inst1_2d_asm_Nx product provides hourly T2M values at a spatial resolution of 0.5° latitude $\times$ 0.625° longitude. For this study, we computed both the daily average T2M value, which was used to compute the accumulated GDD); and the value at the time of the satellite overpass, used to compute the difference between the T2M and LST as described in the following section. In both cases, the T2M data was spatially interpolated to 1 km using a bilinear method. Near-surface air temperature from MERRA-2 has been shown to have good agreement with observations from weather stations, with R2 $\sim$ 0.98 and RMSD $\sim$ 1.7 (Santamaria-Artigas et al., 2019).

3. Methodology

Fig. 3 shows the workflow of the data preprocessing.

3.1. Un-mixing of the wheat signal

According to Franch et al. (2019), for each AU and date, the DVI signal retrieved by pixel, *i*, is:
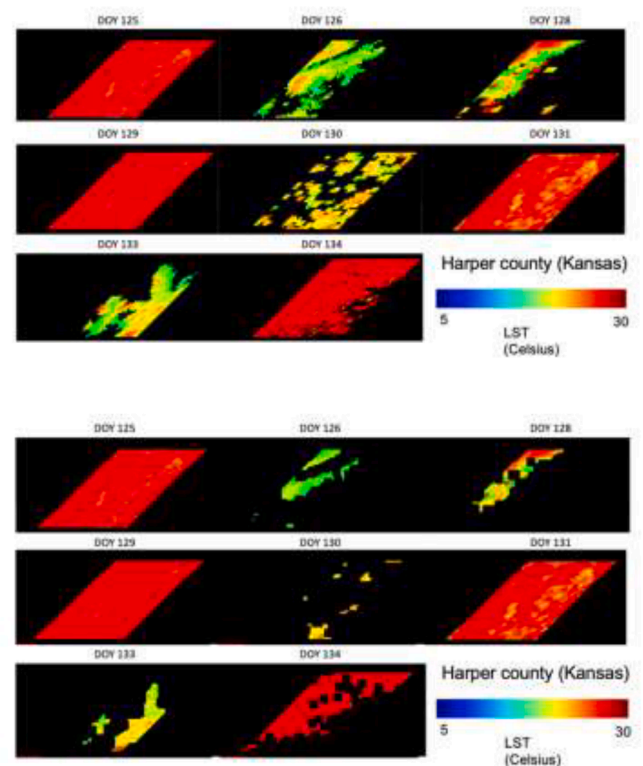

Fig. 2. LST images of the Harper County (KS, USA) in 2003 using the original cloud mask (top) and dilating the cloud mask 5 pixels (bottom).

$$DVI_i = DVI_{i,wheat}\cdot Wpct_i + DVI_{i,others}\cdot(1 - Wpct_i) \quad (2)$$

where $DVI_{i,wheat}$ is the DVI signal from the wheat in the pixel, Wpct is the percentage of wheat within the pixel and DVI_{others} is the DVI from other surfaces observed by the same pixel.

In the case of the US:

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

Therefore, based on the DVI of all pixels through the county (DVI_i) and the crop type maps, we can use Eqs. (2) or (3) to unmix the wheat signal.

3.2. Wheat DVI forecast

To forecast the wheat signal, first, we derive the Gaussian fitting parameters (Eq. (4)) of the DVI evolution with the accumulated GDD for
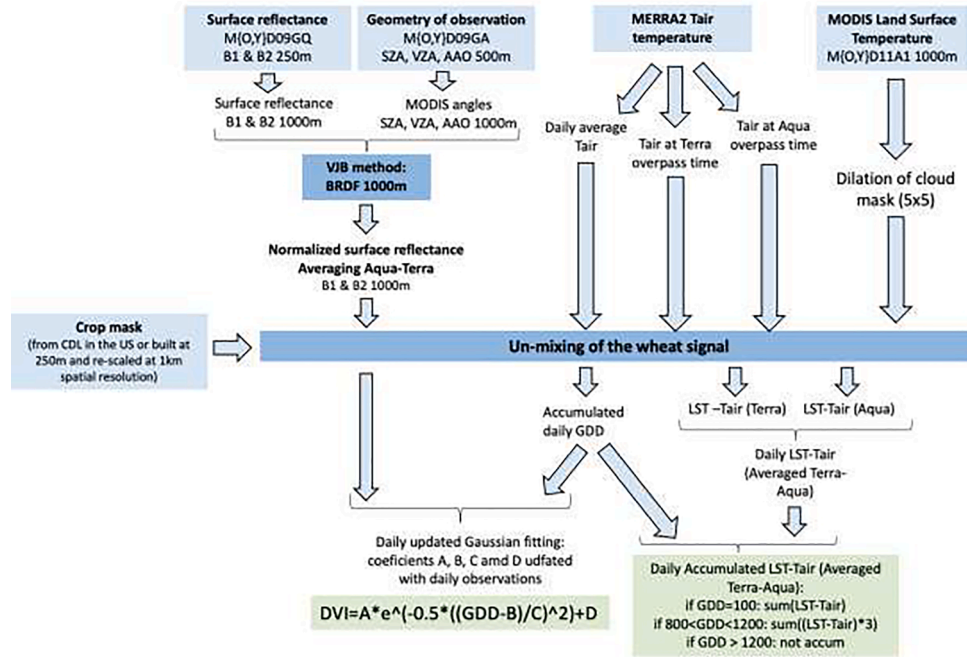
Fig. 3. General workflow of the data preprocessing.

each growing season.

$$DVI(day) = A \cdot e^{-0.5 \cdot ((GDD_{accum}(day) - B)/c)^2} + D \qquad (4)$$

where A,B,C and D are the fitting parameters and GDD$_{accum}$(day) is estimated by accumulating the daily GDDs during the growing season starting from a biofix date:

$$GDD_{accum}(day) = \sum_{d=biofixdate}^{day} GDD_d \qquad (5)$$

and GDD$_d$ is the average daily maximum (T_{max}) and minimum temperatures (T_{min}) minus a base temperature (T_{base}) that in the case of wheat is equal to 0 °C. Note that the biofix date is established as January 1st in

the Northern Hemisphere analogously to Franch et al. (2015) while for the Southern Hemisphere (Argentina and Australia) the DOY 180 (June 29th) is selected as the average start of season for both countries according to Crop Monitor calendars [URL3].

$$GDD_d = \frac{T_{max} + T_{min}}{2} - T_{base} \qquad (6)$$

where if $[(T_{max} + T_{min})/2 < T_{base}]$, then $[(T_{max} + T_{min})/2] = T_{base}$ (McMaster & Wilhelm, 1997).

Fig. 4 (left) shows an example of the DVI wheat signal Gaussian fitting for Harper County in 2004.

Fig. 4 (right) shows the evaluation of the Gaussian fitting through the US and for all years analyzed in the dataset (2001–2019). The plot shows the predicted DVI (based on the GDDaccum and the gaussian fitting
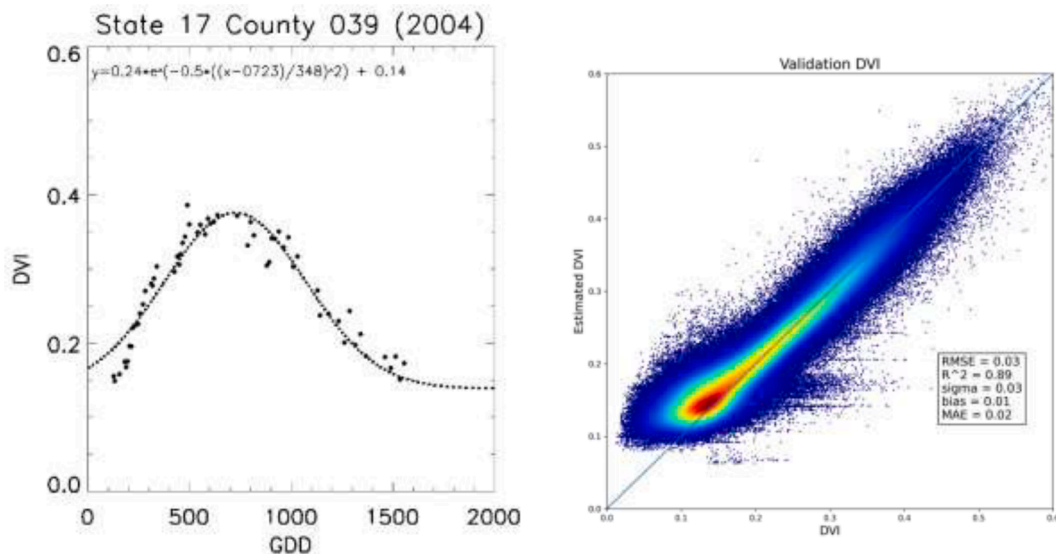

Fig. 4. DVI wheat signal Gaussian fitting on Harper County in 2004 (left) and evaluation of the US DVI estimation based on the Gaussian fitting parameters from 2001 to 2019 across all counties (right).

paramenters) versus the real DVI from DOY 1 to 200. Note that this selection may include some periods (during dormancy and after harvest) where the DVI is correspond to bare soil conditions and the Gaussian

Tair) from Terra and Aqua, which minimizes the absence of data caused by clouds. Nevertheless, to avoid any remaining data gaps, we interpolate the $TDif_{AU,d}$ daily based on the GDD_{accum} evolution.

$$TDif_{AU,d} = LST_{AU,d} - Tair_{AU,d} = mean\left(LST_{AU,Terra} - Tair_{AU,Terratime}, LST_{AU,Aqua} - Tair_{AU,Aquatime}\right) \tag{7}$$

fitting provides constant DVI values. The metrics show a good a good agreement with a RMSE of 0.03 and a coefficient of determination of 0.89, showing the low values of DVI (when there is no crop or it has not emerged) the largest errors.

Second, each date of the forecast is fitted using a Gaussian function considering the DVI wheat signal up to the date considered and the average bare soil conditions after the growing season. Fig. 5 shows an example of the Gaussian fitting on Day Of the Year (DOY) 110 (a) and DOY 130 (b) of 2004 in the Harper County.

Note that we use some data after the end of the season (around $GDD_{accum} = 1500$C-days) as a basis for the Gaussian fitting during the forecasting process. This data is estimated by averaging the bare soil signal after the growing season. To do so, we use the timeseries database in each AU and the average GDD_{accum} value from which historically the bare soil signal is stable (DVI lower than 0.2 and standard deviation lower than 0.04).

3.3. Including the LST into the model

In order to account for any stress conditions related to water deficiency or extreme temperatures, we analyzed the impact of including the difference between the LST and the air temperature in the model. When this difference is positive, it means that the surface is warmer than the surrounding air which can be related to any water stress conditions. In contrast, if the difference is negative, it means that the surface is colder than the surrounding air and may be consequence of a frost event. Based on this principle we accumulated the difference between LST and the air temperature analogously to the GDD_{accum}. To do so, first we applied the unmixing methodology to derive both the daily LST and the air temperature both at Terra and Aqua overpass times to estimate the average temperatures of the wheat pixels across each AU applying Eq. (3). Second, we averaged the daily difference from Terra and the Aqua overpass times (7). Note that we consider all cloud free pixels within each AU to unmix the LST and then we average the temperature difference (LST-

Finally, we accumulated this TDif daily from $GDD_{accum} = 100$C-days until $GDD_{accum} = 1200$C-days. However, the daily averaged difference was strengthened by the weight (*3) during and after the peak DVI happens, that is from $GDD_{accum} = 800$C-days d to $GDD_{accum} = 1200$C-days. This range of GDDs is equivalent to 30 day period after the peak considered in (Franch et al., 2015) to account for water stress conditions that may reduce the final yield and are not captured by the optical data. Additionally, the selected period (800–1200C) includes the grain filling stage when any drought condition is critical to determine the final yields. Note that we apply the same approach (i.e. the same factor) to all regions analyzed in this work.

$$TDif_{AU,t} = \sum_{d(GDD_{accum}=100)}^{t} LST_{AU,d} - Tair_{AU,d} \tag{8}$$

Fig. 6 (left) shows the accumulated difference evolution for each season. The plot shows that the year with the lowest accumulation is 2007 and such year also shows the impact of the late frost in early April which drives the accumulated difference to negative values on DOY 91. On the contrary, 2018 and 2014 show the largest temperature difference accumulation. However, while 2018 shows the largest accumulated values through the season, 2014 showed normal accumulated differences up to April, when the drought conditions started to be more severe (USDA-NASS, 2019). Besides, Fig. 6 (right) shows the yield versus the accumulated difference on DOY 150 for Harper County where each dot represents a different growing season (from 2001 to 2019). This plot shows a quadratic relationship between the accumulated temperature and the final yield. Low accumulated TDif, which might be related to frost events show lower yields. This is the case of 2007 when a late frost hit Kansas. Similar is the case for high accumulated TDif, which can be related to drought conditions and show the lowest yields. Finally, the highest yields are obtained for mid- accumulated temperatures' differences (around 500 K).
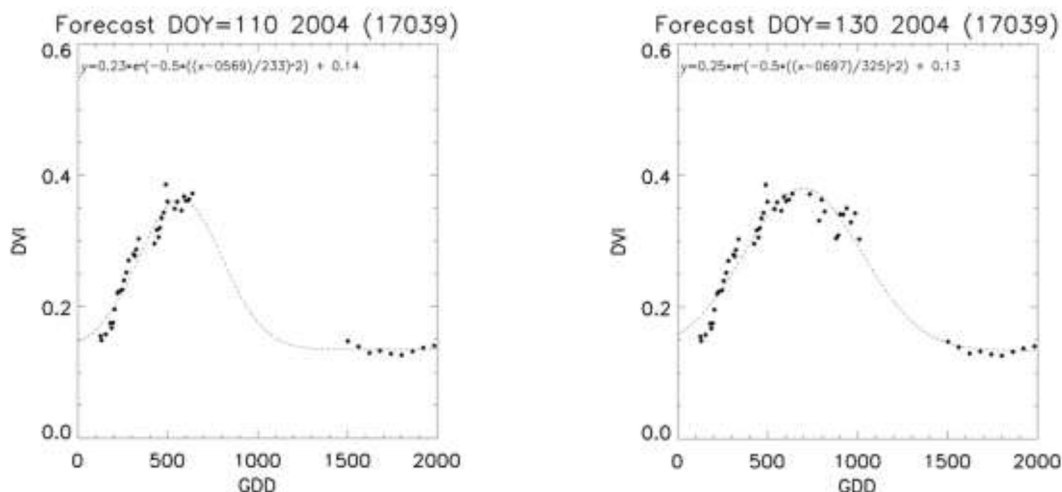

Fig. 5. DVI wheat signal forecast on DOY 110 (left) and DOY 130 (right) vs the GDD_{accum} on Harper County in 2004.
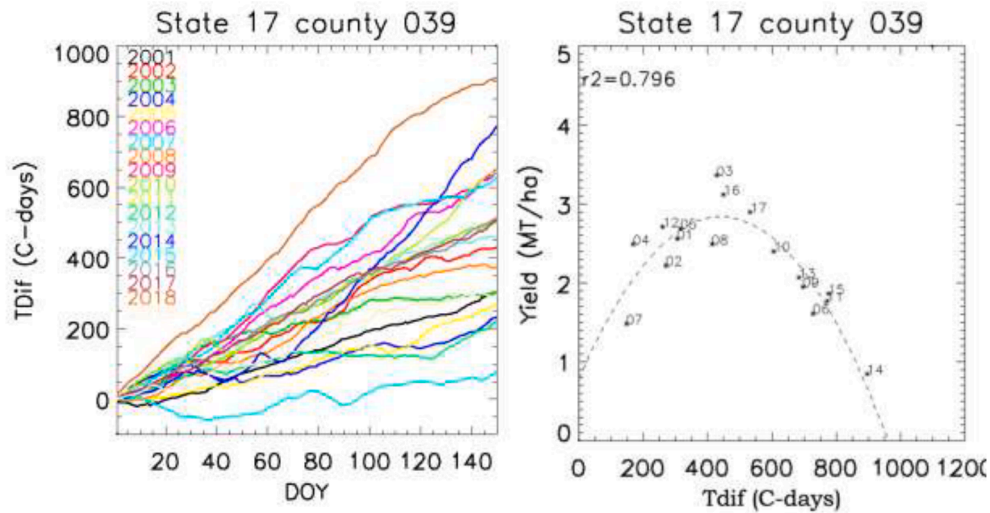
Fig. 6. Accumulated difference of LST and air temperature for each DOY (left) and on DOY 150 versus the final yield (right) in Harper County during all growing seasons from 2001 until 2019.

3.4. Calibration of the yield model

We calibrated the yield model based on all years' official statistics at the subnational level. In this study, we just consider in the study those AU with pixel wheat purities higher than 40% as suggested by (Franch et al., 2019). In each AU we tested two approaches. First, we evaluated the model performance based solely on the solar spectral range information:

$$yield_{AU,t} = c1_{AU,t} \cdot A_{AU,t} + c2_{AU,t} \cdot C_{AU,t} \tag{9}$$

where $c1_{AU,t}$ and $c2_{AU,t}$ are the calibration coefficients and $A_{AU,t}$, $B_{AU,t}$ and $C_{AU,t}$ are the Gaussian fitting parameters described in the previous section for a given AU and day t.

Second, we test the model performance when including the thermal information as explained in the previous section:

$$yield_{AU,t} = c1_{AU,t} \cdot A_{AU,t} + c2_{AU,t} \cdot C_{AU,t} + c3_{AU,t} \cdot TDif_{AU,t} + c4_{AU,t} \cdot TDif_{AU,t}^2 \tag{10}$$

where $c1_{AU,t}$, $c2_{AU,t}$, $c3_{AU,t}$ and $c4_{AU,t}$, are the calibration coefficients and $TDif_{AU,t}$ is the difference between LST and air temperature of the AU and accumulated until day t. However, given the relatively large number of regressors compared to the data available in each AU (19 years maximum depending on data availability for the particular AU), we run an iterative process to reduce the number of regressors. To do so, we computed the relative importance of each regressor based on the permutation feature importance method (Altmann et al., 2010). The permutation feature importance is calculated as follows. First, the model is fitted and a baseline metric is calculated (the determination coefficient in this case). Next, one of the regressors from the same data set is randomly permuted and the metric is evaluated again. The permutation importance is defined as the difference between the metric obtained after the permutation and the baseline metric. These steps are repeated independently for all the regressors in the dataset to obtain the importance of all the features. A high value means that the feature is important for the model. Based on this statistic we iteratively removed the regressors with relative importance (evaluated against the determination coefficient) lower than 20%. Note that two of the regressors considered (TDif and TDif^2) are directly correlated. Thus, the relative importance of these regressors might be underestimated by the permutation importance method.

3.5. Cross-validation and performance metrics

Each country is validated independently both at the national and the subnational level from 2001 to 2019. The national yield forecast (ForNatYield$_{DOY,year}$) is estimated following this equation:

$$ForNatYield_{DOY,y} = \frac{\sum_{SU} ForYield_{AU,DOY,y} \cdot Area_{AU,y}}{\sum_{SU} Area_{AU,y}} \tag{11}$$

where ForYield$_{AU,DOY,y}$ is the ARYA yield forecasted in each AU in a given DOY and year y and Area$_{AU,y}$ is the harvested area in each AU and year y.

Note that at the End Of the Season (EOS), when aggregating the subnational yield estimations to national level the total yield forecast is in some cases biased against the national statistics in some countries. This is a consequence of limiting the applicability of the model to AU with purities higher than 40%. This ensures that the wheat signal is strong enough (compared with the signal from surrounding surfaces) to apply the un-mixing algorithm. Table 2 shows the bias generated when comparing the subnational yield forecasts at the EOS aggregated at national level to the official national statistics considering the whole time series in each country. In order to correct the bias, the results shown in in the following section are re-calibrated dividing the yield results of each forecast by the constant re-calibration coefficients (RCC) displayed in Table 2.

$$RCC = \frac{\sum_{y=2001}^{2019} Nationalyield_y \cdot ForNatYield_{EOS,y}}{\sum_{y=2001}^{2019} Nationalyield_y^2} \tag{12}$$

Additionally, the variability in climatic zones can result in different timing of crop development. This means that in cooler parts of a country, wheat will emerge after dormancy later than in warmer areas. ARYA can

Table 2
Re-calibration coefficients (RCC) at national level.

Country	RCC
Russia	1.066
United States	0.941
France	1.000
Australia	0.984
Ukraine	0.960
Argentina	1.000
Germany	1.000

just be applied after the emergence of the crop, when there is already vegetation signal, so the Gaussian fitting can be implemented. This is one of the challenges in crop forecasting over large areas using remote sensing data. Thus, during the early forecasts there is an additional bias at national level caused by the limited number of AU where the algorithm can be applied. To minimize this effect, early national yield forecasts consider the average yield of the timeseries (2001–2019) in those AU where ARYA can be applied at the EOS but not during the early forecasts.

The performance metrics are based on a leave-one-out cross-validation (LOOCV) procedure, which is based on iteratively retaining a year for validation, while using the other years for model's calibration. The main performance metrics analyzed are the determination coefficient (r^2), the RMSE and the Relative RMSE (RRMSE). Particularly, the evolution of the RMSE for each DOY of the forecast is estimated following this equation:

$$RMSE_{DOY} = \sqrt{\frac{\sum_{y=2001}^{2019} \left(ForNatYield_{DOY,y} - StatsYield_y\right)^2}{19}} \quad (13)$$

Finally, to evaluate how the performance of the model is translated into the production, the forecast production is estimated by multiplying the ARYA yield forecasts by the harvested area from official statistics.

$$Total\ Forecast\ Production = \sum_{SU} Yield\ Forecasted \cdot Harvested\ Area \quad (14)$$

The uncertainty of the production forecasts, given that we consider the official harvested area with unknown error, is estimated by multiplying the absolute error (RMSE) of the yield forecasts by the harvested area as defined by the theory of error propagation.

4. Results

4.1. Calibration of ARYA.

Fig. 7 shows the relative importance of each regressor and the number of regressors considered in each AU for all the countries analyzed. Generally, ARYA considers 2 or 3 regressors. Note that regressors with relative importance lower than 20% are not considered into the model. The regressor with the largest relative importance is the peak value followed by the width. For the US, the thermal regressors show more importance in the central counties where generally TDif2 shows higher importance than TDif. This is also the case in Russia and Ukraine, where the thermal regressors are more important in southern oblasts. However, TDif is slightly more important than TDif2 in France and Germany though there is a limited number of provinces where the thermal regresors show an importance. In Australia, where the thermal information plays a more important role (as we detail in the next
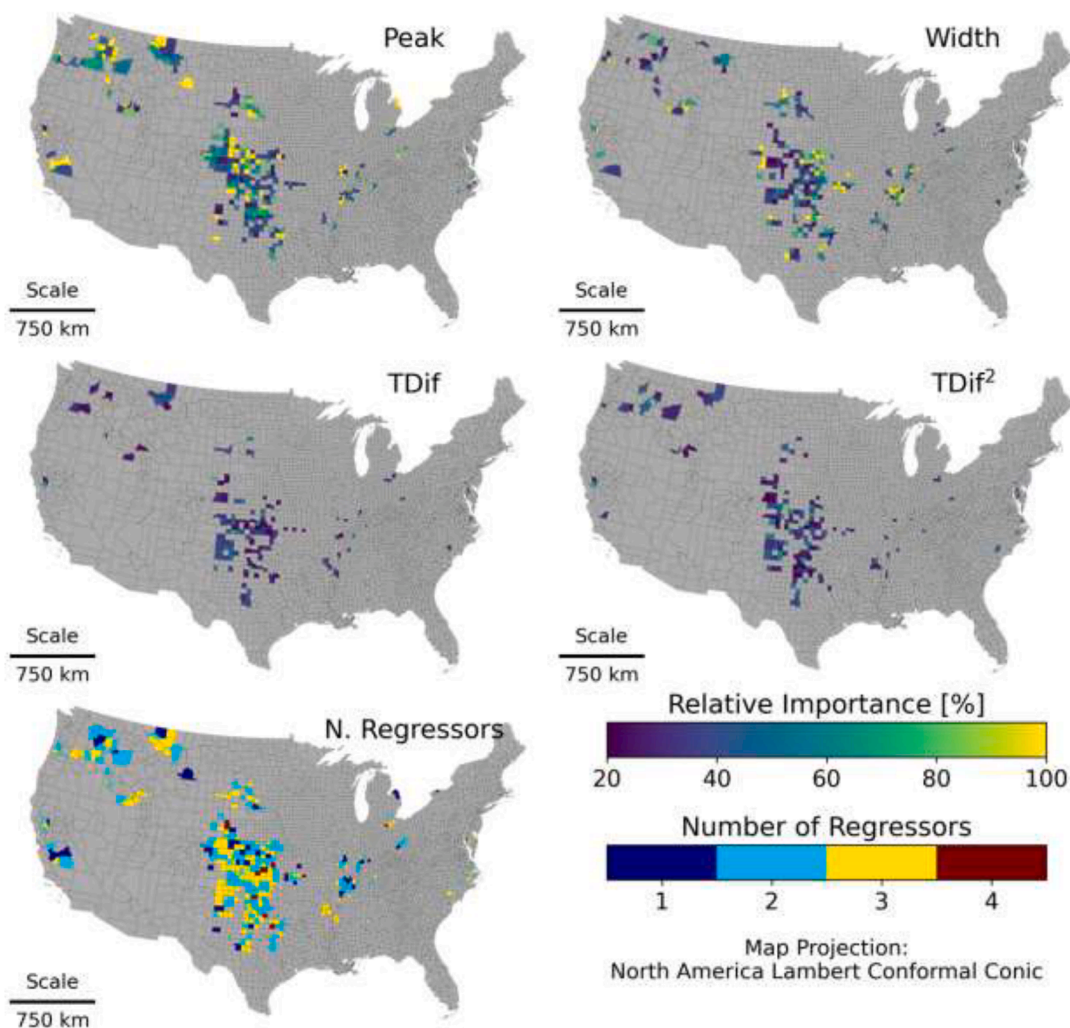

Fig. 7. Relative importance of each regressor and number of regressors considered in each AU in the US, Russia and Ukraine, France and Germany, Australia and Argentina.
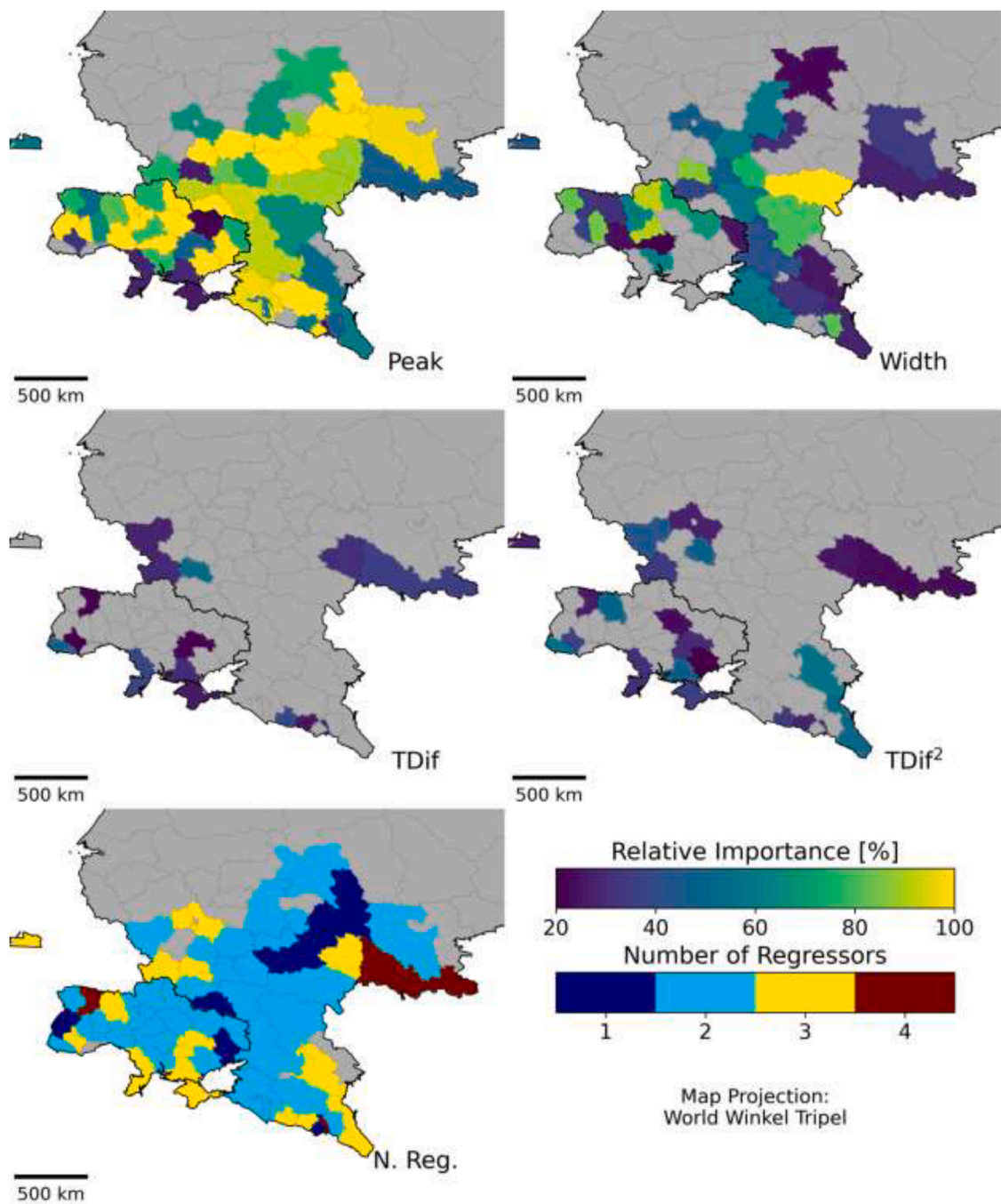
Fig. 7. (*continued*).

section), TDif is relatively more important in the west of Victoria and South East of South Australia while TDif2 is more important across New South Wales. Finally, in Argentina both TDif and TDif2 have very similar importance across the country and there are a few districts in the south of Buenos Aires or along the Cordoba province where the thermal parameters are the only regressors considered in the model.

4.2. Performance metrics

Fig. 8 shows the subnational (red) and national (black level error evolution depending on the date of the forecast of the modeled yield when compared to the official statistics from 2001 to 2019 based on the cross validation approach. The solid line shows the results when using just optical data (Eq. (8)) while the dashed line represents the error

evolution when integrating both the optical and the thermal information (Eq. (9)). Additionally, in the background the bars represent the number of AU where ARYA is applied. This number increases as the season advances.

All countries show some spikes at the beginning of the season which are generated first by the large uncertainty at the SOS, when the peak has not yet happened and second by the limited number of AU where ARYA can be applied at the beginning of the forecast.

In every country the inclusion of thermal data in the model provides better statistics, reducing specially the national level error up to 0.2 t/ha in Australia, and around 0.1 t/ha in Ukraine, Germany and Argentina. However, the thermal inclusion generates slightly higher errors at subnational level in Russia, however this is not translated into the national level performance.
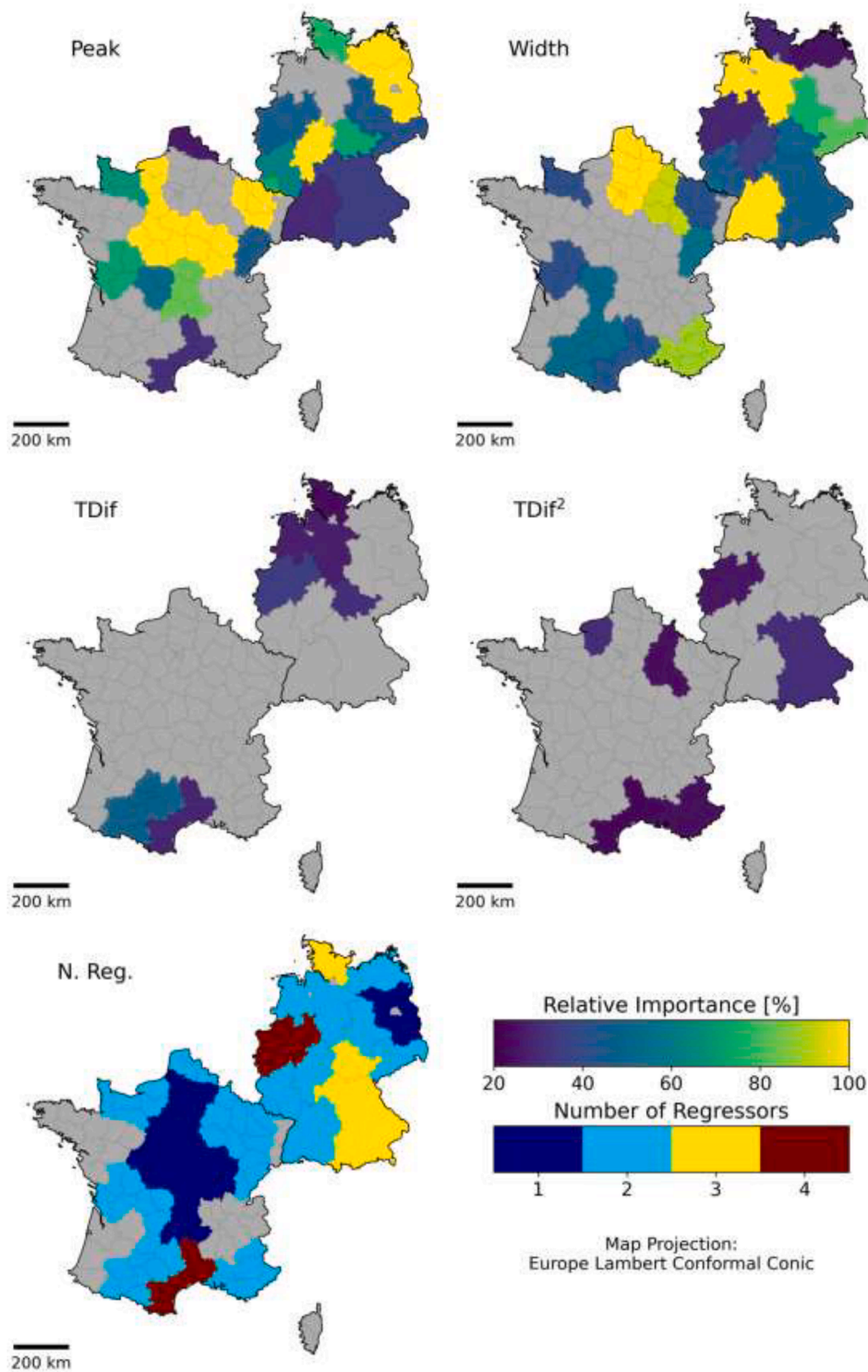
Fig. 7. (*continued*).

To better analyze the results, Table 3 summarizes the national level RMSE and RRMSE at particular DOYs.

The minimum error at the EOS in absolute units is in the US followed closely by Australia, while in relative units is Germany followed by the US. On the contrary, the largest error in absolute units is in France and Ukraine and in relative units in Ukraine and Australia. Focusing on the error evolution, the errors remain stable after DOY 180 (end of june) in the northern hemisphere and after DOY 310 (beginning of November) in the southern hemisphere. Before that, the error increases slightly

(within a 2% variation) in most countries until the average date of the peak that is between DOY 130 and DOY 150 in the northern hemisphere and DOY 280 in the southern hemisphere. Besides, the countries where the error increases more rapidly are Australia and Ukraine. Overall, the error remains low (within 0.2–0.4 t/ha at national and 0.5–0.7 t/ha at subnational level) after DOY 140 (middle of May) in the Northern Hemisphere and DOY 280 (beginning of October) in the Southern Hemisphere. Note that the harvest is completed by the beginning of August in the Northern Hemisphere and end of December in the
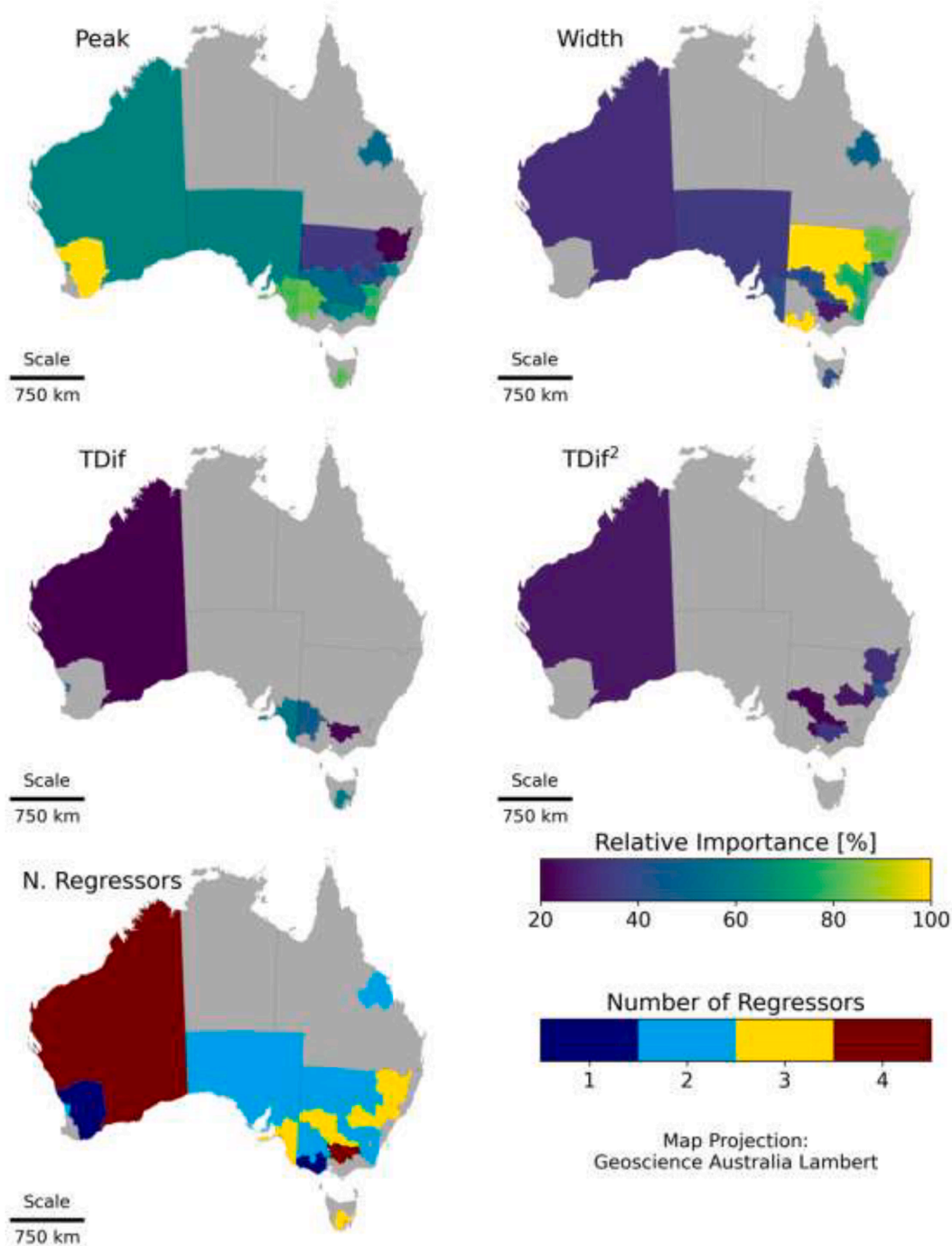
Fig. 7. (*continued*).

Southern Hemisphere, which means that we can provide a good estimate 2 to 2.5 months prior to harvest.

Next, Fig. 9 shows the national and subnational level cross validation at the last day of the forecast, which corresponds to the EOS. This date was selected based on GEOGLAM Crop Monitor crop calendar harvest dates for each country [URL3].

To better compare their performance, Fig. 10 shows the error at the EOS for each country. The US and Australia show the lowest errors at both the subnational (~0.5 t/ha) and national (~0.2 t/ha) levels, while Ukraine and France show the largest errors for both the subnational (~0.7 t/ha) and national (~0.5 t/ha) levels.

Finally, Fig. 11 (left) shows the national level production estimation for all the countries considered. The performance statistics show a good agreement with a RMSE of 3.1 MT (10%) and a determination coefficient of 0.94. Additionally, Fig. 11 (right) shows the aggregated global production based on the yearly total production of each country included in the study added up at global level. Despite 2007, when a late frost in Kansas caused a major winter wheat loss and leads to overestimation of the final production, all the other years' production estimations errors include the official production from statistics and show a good agreement with a RMSE of 12 MT/ha (5.9%) and a determination coefficient of 0.74.
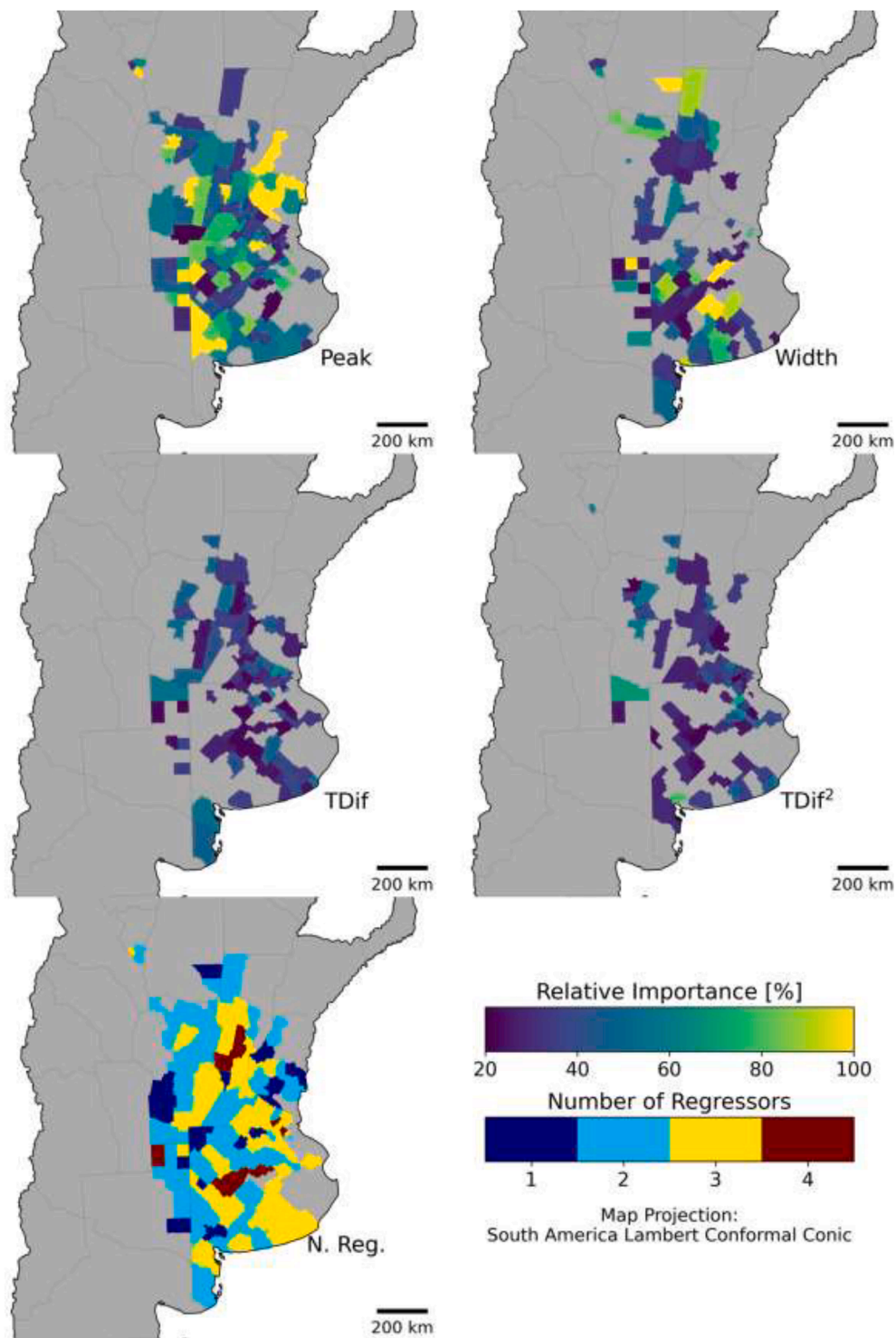
Fig. 7. (*continued*).

5. Discussion

In this work we present the ARYA forecasting algorithm. This method is based on EO and reanalysis data, thus contributing to address gap #1 cited in the introduction section. With this method we amplify the use of EO data in agriculture monitoring. Additionally, ARYA is evaluated in seven major wheat exporting countries over 19 years which provides a good understanding of the forecast uncertainties, contributing to address gap #3. Particularly, the wheat yield can be forecasted with RMSE within 0.2–0.4 t/ha (5–15%) at national and 0.5–0.7 t/ha

(7–20%) at subnational level after DOY 140 (middle of May) in the Northern Hemisphere and DOY 280 (beginning of October) in the Southern Hemisphere. Finally, the forecasts can be implemented based on the daily acquisitions of MODIS and MERRA2 data estimations. However, MERRA2 data has one month of latency. Therefore, in the operational context other products with similar performance metrics for agriculture applications (Santamaria-Artigas et al., 2019) such as NCEP2 (Kanamitsu et al., 2002), with three to five days latency, or ERA5 (Hersbach et al., 2020), that can provide preliminary data within 5 days of real time, could be an alternative. Therefore, weekly forecasts can be
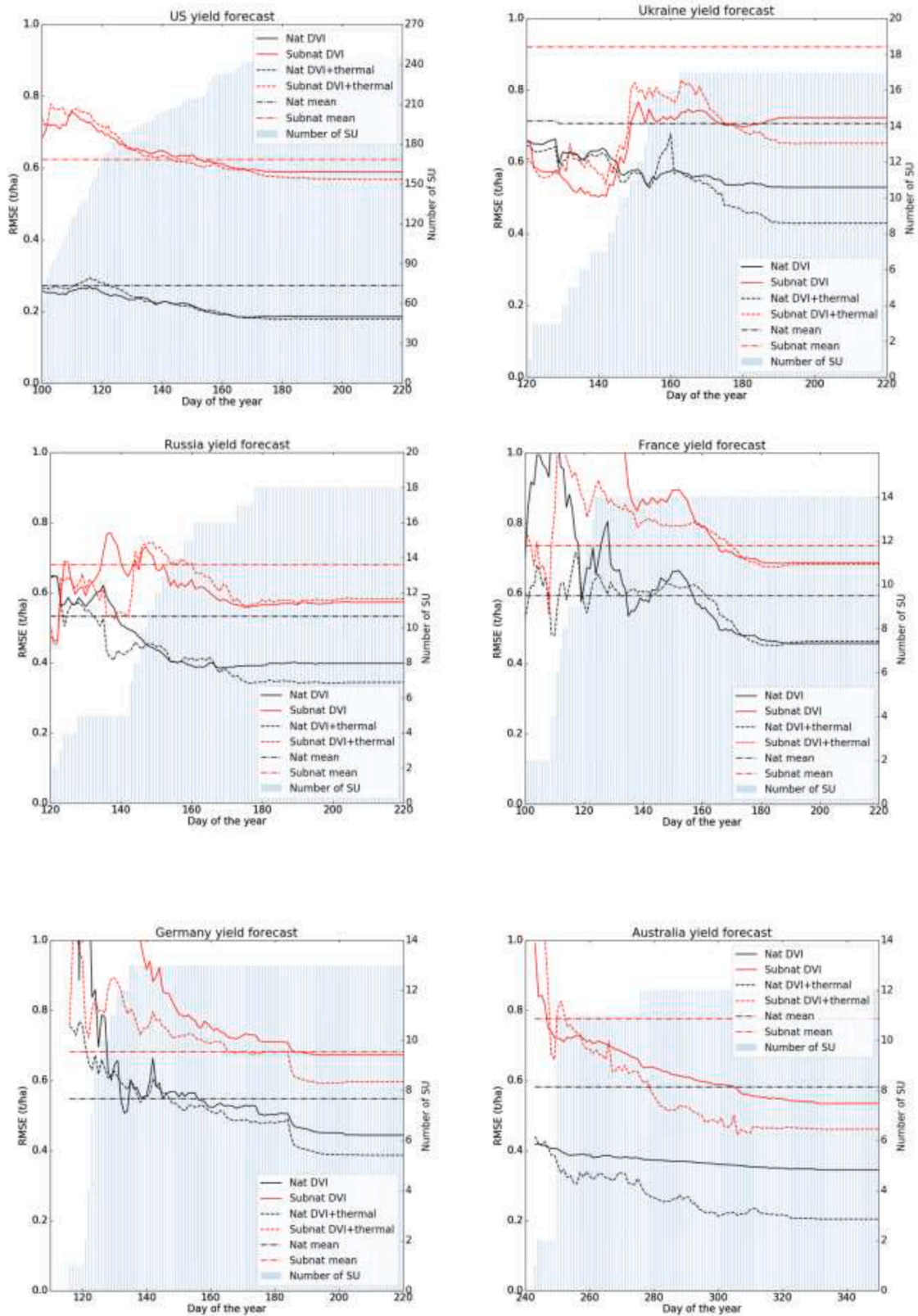
Fig. 8. Forecast error (%) evolution depending on the forecast date at subnational (red) and national level (black) for each country. The solid line represents ARYA performance when applying Eq. (8) and the dashed line when applying Eq. (9). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
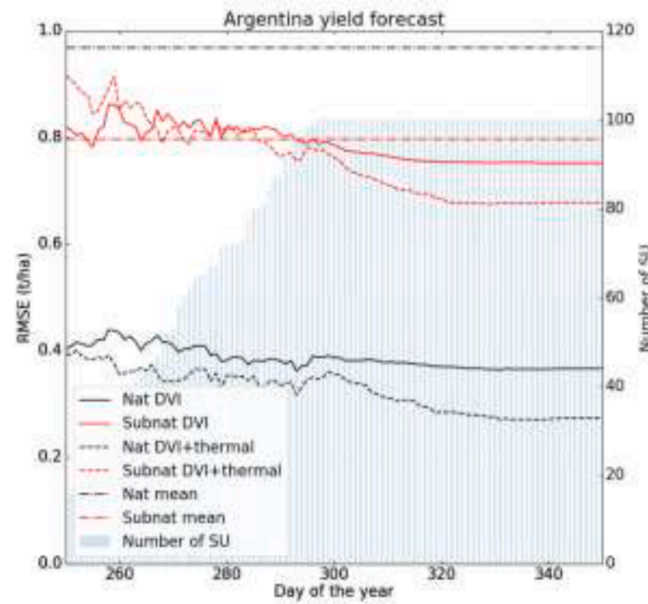
Fig. 8. (*continued*).

Table 3
Forecasts' RMSE (t/ha) and relative error (%) at national level for different DOY. Values in red highlight the timing of the national average wheat peak date while the yellow represents the date when in average the latest AU reaches the peak.

RMSE (t/ha) / RRMSE (%)	DOY130	DOY140	DOY150	DOY160	DOY170	DOY180	DOY190	DOY200
USA	0.26/8.6	0.22/7.3	0.20/6.7	0.18/6.1	0.17/5.6	0.17/5.5	0.17/5.5	0.17/5.5
Russia	0.54/18.3	0.42/14.2	0.44/14.9	0.41/13.8	0.37/12.6	0.34/11.5	0.34/11.5	0.34/11.5
Ukraine	0.59/18.2	0.60/18.4	0.57/17.6	0.70/21.6	0.55/17.0	0.46/14.3	0.42/13.1	0.42/13.1
France	0.61/8.6	0.60/8.4	0.61/8.7	0.60/8.4	0.52/7.3	0.46/6.3	0.46/6.4	0.46/6.4
Germany	0.62/8.3	0.55/7.3	0.52/7.0	0.51/6.8	0.48/6.4	0.47/6.3	0.40/5.3	0.39/5.1
	DOY260	DOY270	DOY280	DOY290	DOY300	DOY310	DOY320	DOY330
Australia	0.31/19.0	0.30/18.3	0.28/16.9	0.25/15.1	0.24/14.4	0.22/13.1	0.22/13.2	0.21/13.0
Argentina	0.35/14.6	0.30/12.4	0.33/13.5	0.32/13.1	0.34/14.1	0.30/12.4	0.28/11.5	0.27/11.0

aimed in the operational context both at national and subnational level, providing timely but detailed estimations, and contributing to address gap #3.

The work presented builds on our previous work (Franch et al., 2015) by using the GDD as a parameter to forecast the NDVI peak. However, said work assumes a linear regression of the forecasting day and the peak and in this work, we use the gaussian fitting to forecast both the peak and the width of the DVI curve. Additionally, Franch et al. (2015) focus on the 5% purest pixels in each AU to implement a unique regression algorithm for all AU just corrected by the maximum purity retrieved

(Becker-Reshef et al., 2010). Deriving a global model (unique model for all regions), might be more robust thanks to higher cardinality of data and in Franch et al. (2015) paper it was shown that the same model could be applied to the US, Ukraine and China by just adding a calibration factor. However, that might not be the case for other regions. In the present work we unmix the DVI signal from all pixels and calibrate a different regression algorithm in each AU which is more efficient to capture the particularities of each region. Nevertheless, we need to keep in mind that the model is trained for each AU with limited data (compared to the number of regressors) that is reduced further to derive
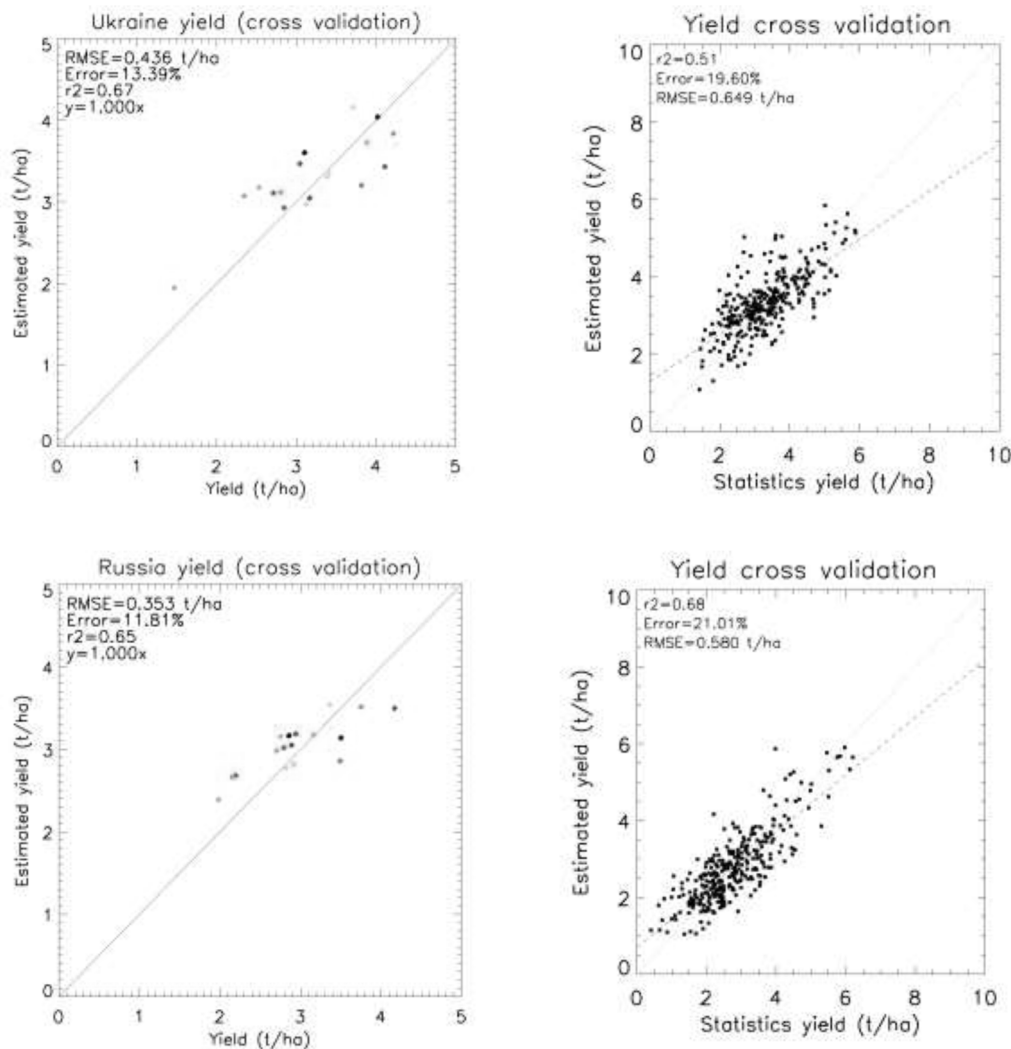
Fig. 9. EOS yield forecast validation at national (left) and subnational (right) level in each country analyzed.

the cross-validation statistics (each year's performance is evaluated excluding it from the calibration of the model). This kind of validation provides a good representation of the model performance and robustness. In practice, however, yield forecasts for upcoming years will use data from available previous years, which might be useful to better characterize the effect of past extreme events on crop yields. This is the case for the late frost that impacted Kansas in 2007 or Ukraine's drought in 2003. When evaluating these particular years' performance, the error is high since those are unique events in each countries' time series (2001–2019). One solution that will be explored in future research to enhance the robustness of the model will be integrating all the AU with similar performance in terms of yield within the same calibration equation. This can be done using geographically weighted neural networks (Crane-Droesch, 2018).

Comparing this work with Franch et al. (2019), there are several differences:

(1) ARYA is a forecasting algorithm while Franch et al. (2019) can be just implemented after the EOS.

(2) In this work we directly use the difference between the LST and the air temperature in the model while our previous work considers the average Evaporative Fraction (EF) 30 days after the DVI peak. Note that the EF was estimated based on the S-SEBI method (Roerink et al., 2000) that uses LST vs albedo diagrams of the area

considered to derive the dry and wet conditions. This limits the applicability of the method to days when the cloud conditions allowed a good representation of the area conditions. This worked when averaging this value over a 30 day period, but in the operational context of a forecasting method, it might create important gaps in the timeseries. This was addressed in this work by directly using the LST combined with the air temperature from reanalysis products.

Comparing the ARYA performance metrics with previous works, the results show some improvements. The forecasts remain low from 2 to 2.5 months before harvest. This agrees with the results presented in Franch et al. (2015). However, said work just focused on forecasting two countries (the US and Ukraine) at national level and the error (<10%) refers to calibration validation results. Compared to Franch et al. (2019), this work provides equivalent results at the EOS in Ukraine and the US by directly using the LST and the air temperature. Note that said work was based on the DVI peak and width and the average EF 30 days after the peak and is not a forecasting method. Compared to Kogan et al. (2013) and despite including more years in our analysis (19 versus 12 years), our results show very similar performance metrics with errors at subnational level ranging between 0.5 and 0.8 t/ha up to 3 months before harvest. In Europe, our results show larger errors in France and Germany (6–8% and 5–7% respectively) compared to Pagani et al.
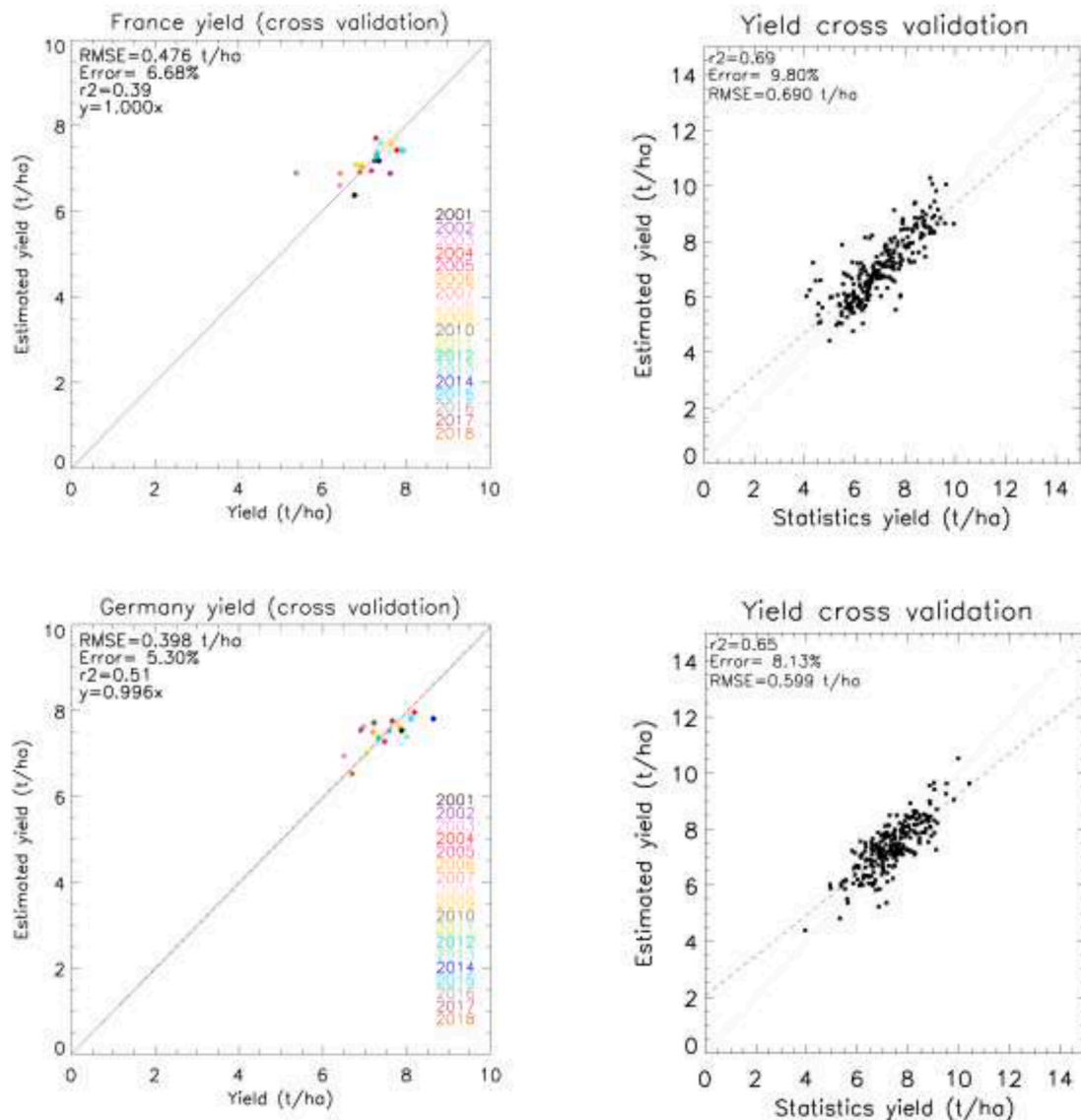
Fig. 9. (*continued*).

(2017) (4.2–5.6% and 3.8–4.6% respectively). However, said study, that also analyzed 19 years (from 1995 to 2013) does not include especially complex years. Particularly, year 2016 in France, when a conjunction of abnormally warm temperatures in late autumn and abnormally wet conditions in the following spring led to the most extreme yield loss in recent history that was not anticipated by any public forecasting system (Ben-Ari et al., 2018). In fact, this year is not well captured by ARYA. Note that the only factor in ARYA that accounts for meteorological conditions (not captured by the DVI) is the accumulated temperature difference (TDif). However, the underperformance of France in 2016 is related to extreme wet conditions in spring (April to July). Thus, to solve the underperformance of ARYA in France 2016 (Fig. 9), in future works we will look into integrating additional meteorological parameters that could capture said conditions such as the accumulated precipitation. Another complex year in Europe that has been included in our analysis is 2018 in Germany, when crop losses (48.9%) and early harvesting (61.3%) happened as a result of the compound effects of the soil moisture drought and a heatwave (de Brito et al., 2020). Though this year is well captured at the EOS (Fig. 9), the earlier forecasts' larger errors might be impacted by such events. Finally, compared to Kamir et al. (2020) that reported an EOS error of 0.59 t/ha at subnational level in

Australia, our results show better performance metrics of 0.21 t/ha at national and 0.47 t/ha at subnational level.

Comparing the different countries' performance, the US shows the lowest errors. This might be a consequence of two error sources. One is the number of AU considered, which is largest for the US. In fact, Fig. 12 shows the ratio of the RMSE at subnational and national level versus the square root of the number of AU considered in each country. Except for Australia, that has a low RMSE with a limited number of AU (13), the other countries' error ratio show a linear dependency with the square root AU. This is a consequence of the error propagation theory: the national error depends both on the subnational error and de square root of number of subnational units.

The other source of error is the crop type maps. The US is the only country with moderate resolution crop type maps (30 m resolution from CDL), while ARYA implementation to other countries relies on crop type maps derived directly from MODIS 250 m data.

During this work we learned that the LST is a very useful parameter to monitor the crop yield. The improvement of estimates when including the thermal information versus just using the optical information (DVI) is stronger in Australia, Argentina, Germany, Ukraine and Russia and is more limited or even provides very similar performance metrics in
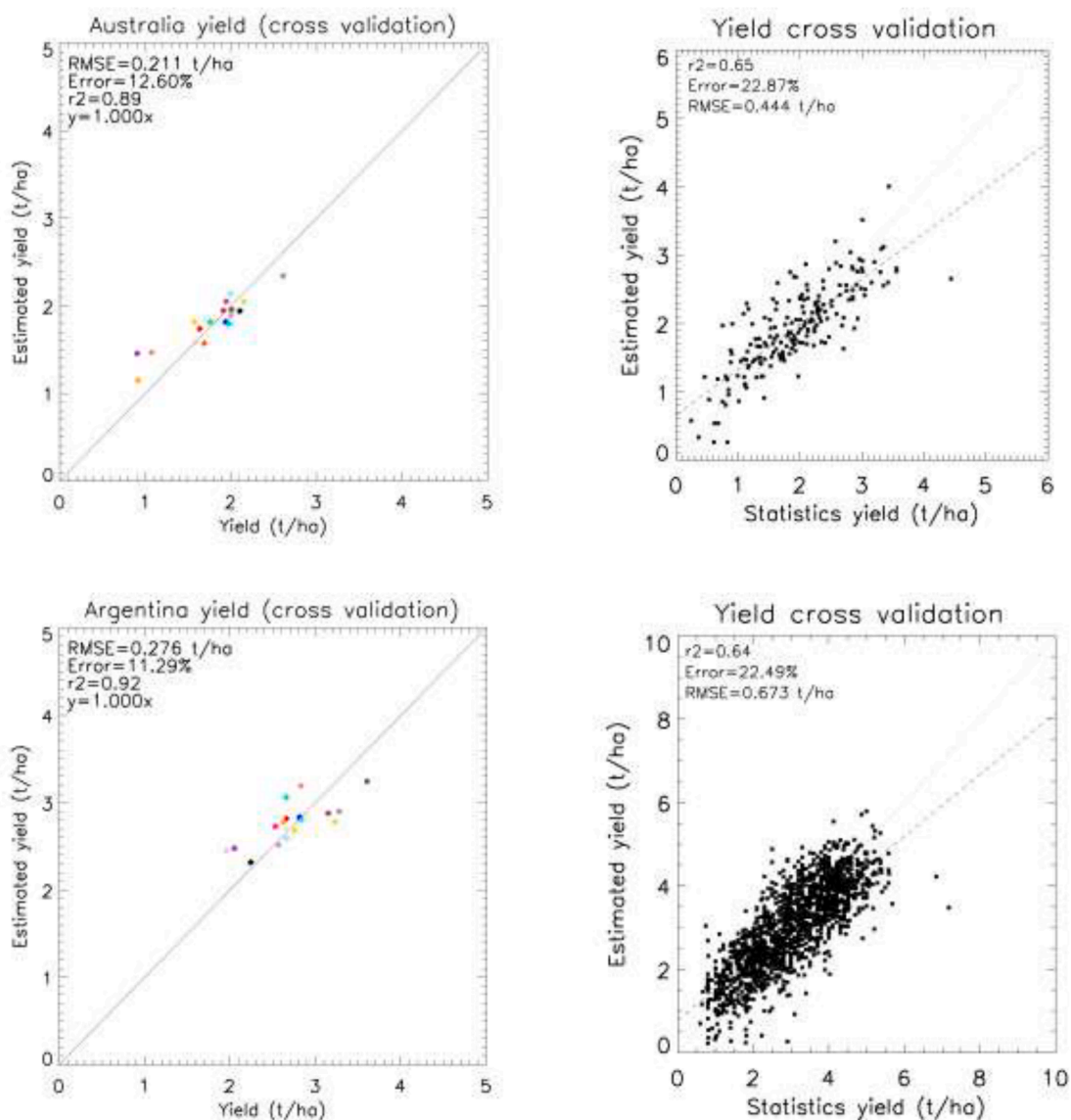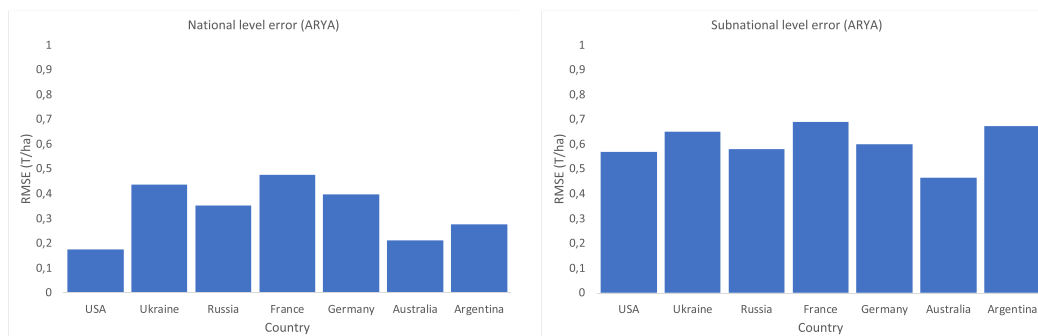
Fig. 9. (*continued*).

Fig. 10. Errors at the EOS in each country at national (left) and subnational (right) level.

France and the US. However, the thermal information retrieval presents some limitations that we must consider in order to exploit its benefits. Fig. 2 showed that depending on the direction of the cloud, the LST of the underlying surface can change dramatically. In this work, we minimized the impact of such cases by dilating the original cloud mask

with a 5 pixels buffer and by integrating both Aqua and Terra temperature differences. However, after these processes the data presented gaps caused by clouds in some regions that were filled out interpolating linearly the LST-Tair cloud free estimations. This approach might be causing some inaccuracies in some regions prone to cloud cover.
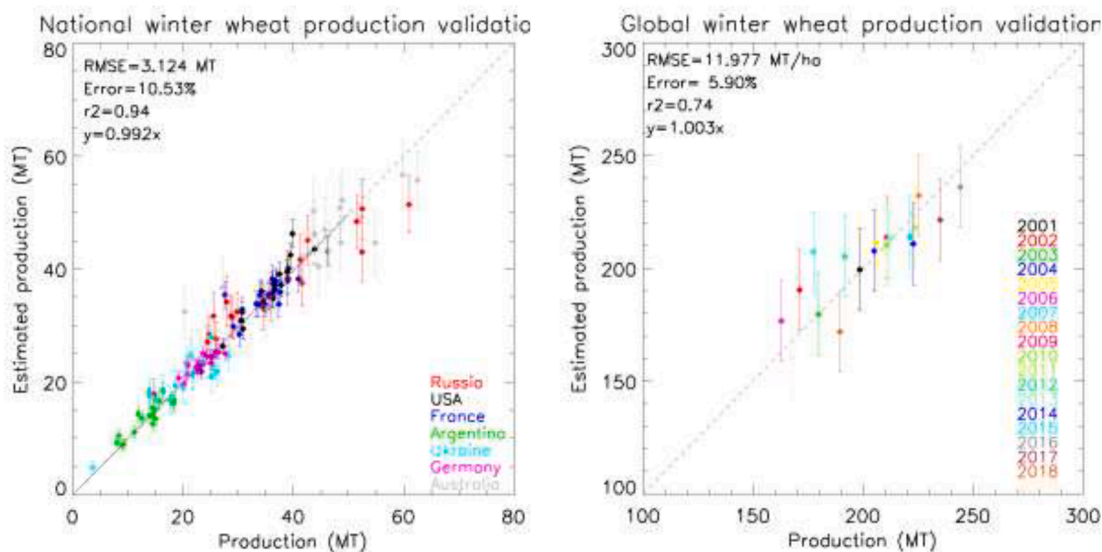
Fig. 11. Production validation at national level and at global level (based on the seven countries analyzed).
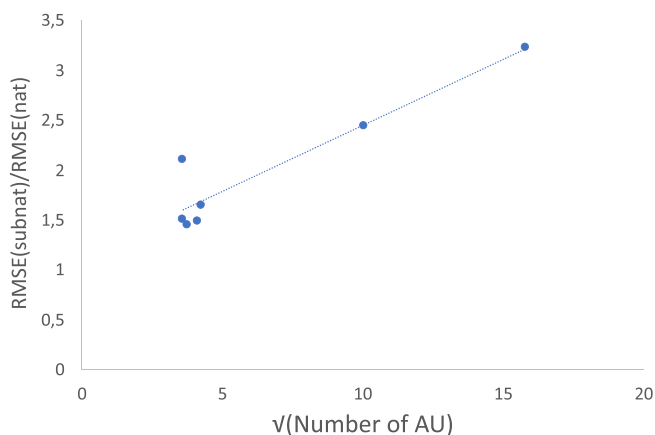

Fig. 12. RMSE at subnational level divided by the RMSE at national level versus the square root of the number of AU considered in each country.

Therefore, a potential solution to this problem, would be using geostationary data to analyze the clouds direction and exploit the large number of LST images per day.

6. Conclusions

This study presents the ARYA forecasting algorithm which is based on analyzing the DVI evolution with the accumulated GDD information and the difference between LST and the air temperature. The algorithm is successfully applied to the seven major wheat exporting countries which account for over 70% of the wheat exports worldwide. The results show that ARYA provides good estimations with RMSE within 0,2–0,4 t/ha at national and 0,5–0,7 t/ha at subnational level after DOY 140 (mid of May) in the Northern Hemisphere and DOY 280 (beginning of October) in the Southern Hemisphere. This means that ARYA can provide reliable estimates 2 to 2.5 months prior to harvest. Additionally, the study provides some insite of LST importance to monitor crop yield and can provide improvements the yield uncertainties up to 0.2 t/ha.

CRediT authorship contribution statement

B. Franch: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Supervision. **E. Vermote:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – review & editing. **S. Skakun:** Conceptualization, Formal analysis, Investigation, Data curation, Writing – review & editing. **A. Santamaria-Artigas:** Data curation, Writing – review & editing. **N. Kalecinski:** Data curation, Writing – review & editing. **J.-C. Roger:** Conceptualization, Writing – review & editing. **I. Becker-Reshef:** Conceptualization, Writing – review & editing. **B. Barker:** Data curation, Writing – review & editing. **C. Justice:** Writing – review & editing. **J.A. Sobrino:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Altmann, A., Tolo\csi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), 1340–1347.

Anderson, M.C., Hain, C., Wardlow, B., Pimstein, A., Mecikalski, J.R., Kustas, W.P., 2011. Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the Continental United States. J. Climate 24(8), 2025–2044. 10.1175/2010JCLI3812.1.

Anderson, M.C., Zolin, C.A., Sentelhas, P.C., Hain, C.R., Semmens, K., Tugrul Yilmaz, M., Gao, F., Otkin, J.A., Tetrault, R., 2016. The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. Remote Sens. Environ. 174, 82–99. https://doi.org/10.1016/j.rse.2015.11.034.

Anderson, S.M., Velez, J.J., Thompson, G.W., 2014. An analysis of K-12 Teachers' conceptions of agriculture prior to and during engagement in an agricultural literacy program. J. Agric. Educ. 55 (3), 132–146.

Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sens. Environ. 35 (2-3), 161–173.

Baruth, B., Royer, A., Klisch, A., Genovese, G., 2008. The use of remote sensing within the MARS crop yield monitoring system of the European Commission. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci 37, 935–940.

Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., McGaughey, K., Justice, C., Baruth, B., Wu, B., Prakash, A., Abdolreza, A., Jarvis, I., 2019. The GEOGLAM crop monitor for AMIS: Assessing crop conditions in the context of global markets. Global Food Security 23, 173–181. https://doi.org/10.1016/j.gfs.2019.04.010.

Becker-Reshef, I., Justice, C., Whitcraft, A.K., Jarvis, I., 2018. Geoglam: A Geo Initiative on Global Agricultural Monitoring. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 8155–8157.

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 (6), 1312–1323.

Ben-Ari, T., Boé, J., Ciais, P., Lecerf, R., van der Velde, M., Makowski, D., 2018. Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France. Nat. Commun. 9 (1), 1627. https://doi.org/10.1038/s41467-018-04087-x.

Breon, F.-M., Vermote, E., Murphy, E.F., Franch, B., 2015. Measuring the directional variations of land surface reflectance from MODIS. IEEE Trans. Geosci. Remote Sens. 53 (8), 4638–4649. https://doi.org/10.1109/TGRS.2015.2405344.

Buschmann, C., Nagel, E., 1993. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. Int. J. Remote Sens. 14 (4), 711–722.

Crane-Droesch, A., 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ. Res. Lett. 13 (11), 114003. https://doi.org/10.1088/1748-9326/aae159.

de Brito, M.M., Kuhlicke, C., Marx, A., 2020. Near-real-time drought impact assessment: a text mining approach on the 2018/19 drought in Germany. Environ. Res. Lett. 15 (10), 1040a9.

Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. Remote Sens. Environ. 97 (2), 192–202.

Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C., Sobrino, J.A., 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 Sens. Environ. 161, 131–148. https://doi.org/10.1016/j.rse.2015.02.014.

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., Devadiga, S., 2017. A 30+ year AVHRR land surface reflectance climate data record and its application to wheat yield monitoring. Remote Sens. 9 (3), 296. https://doi.org/10.3390/rs9030296.

Franch, B., Vermote, E.F., Skakun, S., Roger, J.C., Becker-Reshef, I., Murphy, E., Justice, C., 2019. Remote sensing based grain monitoring: Application to winter wheat in United States and Ukraine. Int. J. Appl. Earth Obs. Geoinf. 76, 112–127. https://doi.org/10.1016/j.jag.2018.11.012.

Franch, B., Vermote, E.F., Sobrino, J.A., Julien, Y., 2014. Retrieval of surface albedo on a daily basis: Application to MODIS data. IEEE Trans. Geosci. Remote Sens. 52 (12), 7549–7558. https://doi.org/10.1109/TGRS.2014.2313842.

Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., Mücher, S., Tetrault, R., Moorthy, I., McCallum, I., 2019. A comparison of global agricultural monitoring systems and current gaps. Agric. Syst. 168, 258–272. https://doi.org/10.1016/j.agsy.2018.05.010.

Gallego, J., J., M., B., B., & Craig M., F. S. (2008). GEOSS (2009). Best practices for crop area estimation with Remote Sensing. Ispra, June 5-6.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30 (5), n/a–n/a.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorolog. Soc. 146 (730), 1999–2049.

Idso, S.B., Hatfield, J.L., Jackson, R.D., Reginato, R.J., 1979. Grain yield prediction: extending the stress-degree-day approach to accommodate climatic variability. Remote Sens. Environ. 8 (3), 267–272.

Janetos, A., Justice, C., Jahn, M., Obersteiner, M., Glauber, J., Mulhern, W., 2017. The risks of multiple breadbasket failures in the 21st century: a science research agenda. Boston University Frederick S. Pardee Center for the Study of the Longer~….

Johnson, D.M., Mueller, R., 2010. The 2009 Cropland Data Layer. Photogramm. Eng. Remote Sens. 76 (11), 1201–1205.

Johnson, D.M., 2014. An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141, 116–128.

Kamir, E., Waldner, F., Hochman, Z., 2020. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J. Photogramm. Remote Sens. 160, 124–135. https://doi.org/10.1016/j.isprsjprs.2019.11.008.

Kanamitsu, M., et al., 2002. NCEP-DOE AMIP-II reanalysis (R-2). Bull. Amer. Meteoror. Soc. 83 (11), 1631–1643.

Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., Lavrenyuk, A., 2013. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. Int. J. Appl. Earth Obs. Geoinf. 23, 192–203. https://doi.org/10.1016/j.jag.2013.01.002.

Kogan, F.N., 1997. Global drought watch from space. Bull. Am. Meteorol. Soc. 78 (4), 621–636.

Li, X., Strahler, A., 1986. Geometric-optical bidirectional reflectance modeling of a conifer forest canopy. IEEE Trans. Geosci. Remote Sens. GE-24 (6), 906–919.

López-Lozano, R., Duveiller, G., Seguini, L., Meroni, M., García-Condado, S., Hooker, J., Leo, O., Baruth, B., 2015. Towards regional grain yield forecasting with 1 km-resolution EO biophysical products: Strengths and limitations at pan-European level. Agric. For. Meteorol. 206, 12–32.

Lopresti, M.F., di Bella, C.M., Degioanni, A.J., 2015. Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. Inform. Process. Agric. 2 (2), 73–84. https://doi.org/10.1016/j.inpa.2015.06.001.

Lucht, W., 1998. Expected retrieval accuracies of bidirectional reflectance and albedo from EOS-MODIS and MISR angular sampling. J. Geophys. Res.: Atmos. 103 (D8), 8763–8778.

Macdonald, R.B., 1984. A summary of the history of the development of automated remote sensing for agricultural applications. IEEE Trans. Geosci. Remote Sens. GE-22 (6), 473–482.

Maignan, F., Bréon, F.-M., Lacaze, R., 2004. Bidirectional reflectance of Earth targets: Evaluation of analytical models using a large set of spaceborne measurements with emphasis on the Hot Spot. Remote Sens. Environ. 90 (2), 210–220.

Mateo-Sanchis, A., Piles, M., Muñoz-Marí, J., Adsuara, J.E., Pérez-Suay, A., Camps-Valls, G., 2019. Synergistic integration of optical and microwave satellite data for crop yield estimation. Remote Sens. Environ. 234, 111460. https://doi.org/10.1016/j.rse.2019.111460.

McMaster, G.S., Wilhelm, W.W., 1997. Growing degree-days: one equation, two interpretations. Agric. For. Meteorol. 87 (4), 291–300.

Pagani, V., Guarneri, T., Fumagalli, D., Movedi, E., Testi, L., Klein, T., Calanca, P., Villalobos, F., Lopez-Bernal, A., Niemeyer, S., Bellocchi, G., Confalonieri, R., 2017. Improving cereal yield forecasts in Europe – The impact of weather extremes. Eur. J. Agron. 89, 97–106. https://doi.org/10.1016/j.eja.2017.06.010.

Rembold, F., Meroni, M., Urbano, F., Csak, G., Kerdiles, H., Perez-Hoyos, A., Lemoine, G., Leo, O., Negre, T., 2019. ASAP: A new global early warning system to detect anomaly hot spots of agricultural production for food security analysis. Agric. Syst. 168, 247–257.

Roerink, G.J., Su, Z., Menenti, M., 2000. S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. Phys. Chem. Earth Part B 25 (2), 147–157.

Rojas, O., 2015. Protocol for Country-Level ASIS: Calibration and National Adaptation Process. Food and Agriculture Organization of the United Nations.

Rosenzweig, C., Iglesias, A., Yang, X.B., Epstein, P.R., Chivian, E., 2001. Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests. Global Change and Human Health 2 (2), 90–104. https://doi.org/10.1023/A:1015086831467.

Roujean, J.-L., Leroy, M., Deschamps, P.-Y., 1992. A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. J. Geophys. Res.: Atmos. 97 (D18), 20455–20468.

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sens. Environ. 131, 215–231. https://doi.org/10.1016/j.rse.2012.12.017.

Santamaria-Artigas, A.E., Franch, B., Guillevic, P., Roger, J.-C., Vermote, E.F., Skakun, S., 2019. Evaluation of near-surface air temperature from reanalysis over the United States and Ukraine: Application to winter wheat yield forecasting. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (7), 2260–2269. https://doi.org/10.1109/JSTARS.460944310.1109/JSTARS.2019.2902479.

Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proc. Natl. Acad. Sci. 106 (37), 15594–15598. https://doi.org/10.1073/pnas.0906865106.

Sellers, P.J., 1985. Canopy reflectance, photosynthesis and transpiration. Int. J. Remote Sens. 6 (8), 1335–1372.

Skakun, S., Franch, B., Vermote, E., Roger, J.-C., Becker-Reshef, I., Justice, C., Kussul, N., 2017. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. Remote Sens. Environ. 195, 244–258. https://doi.org/10.1016/j.rse.2017.04.026.

Tadesse, G., Algieri, B., Kalkuhl, M., von Braun, J., 2014. Drivers and triggers of international food price spikes and volatility. Food Policy 47, 117–128. https://doi.org/10.1016/j.foodpol.2013.08.014.

USDA-NASS, 2019. Kansas wheat history. News Release. https://www.nass.usda.gov/Statistics_by_State/Kansas/Publications/Cooperative_Projects/KS-wheat-history19.pdf.

Vermote, E., Justice, C.O., Bréon, F.M., 2008. Towards a generalized approach for correction of the BRDF effect in MODIS directional reflectances. IEEE Transactions on Geoscience and Remote Sensing 47 (3), 898–908.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. 236, 111402. https://doi.org/10.1016/j.rse.2019.111402.

Wu, Bingfang, Meng, Jihua, Li, Qiangzi, Yan, Nana, Du, Xin, Zhang, Miao, 2014. Remote sensing-based global crop monitoring: experiences with China's CropWatch system. Int. J. Digital Earth 7 (2), 113–137.

URL 1. www.amis-outlook.org.

URL 2. www.geoglam.org.

URL 3. https://cropmonitor.org/.