Conference Proceedings Paper

Estimation of sunflower yields at a decametric spatial scale - A statistical approach based on multi-temporal satellite images

Remy Fieuzal [1,*], Vincent Bustillo [1,2], David Collado [2] and Gerard Dedieu [1]

[1] Centre d'Études de la BIOsphère (CESBIO), Université de Toulouse, CNES/CNRS/INRA/IRD/UPS, Toulouse, France; gerard.dedieu@cesbio.cnes.fr
[2] IUT Paul Sabatier, 24 rue d'Embaquès, Auch, France; vincent.bustillo@iut-tlse3.fr; david.collado@univ-tlse3.fr
* Correspondence: fieuzal.remy@gmail.com; Tel.: +xx-xxx-xxx-xxxx

Published: date
Academic Editor: name

Abstract: Recent advances in sensors onboard harvesting machines allow accessing the intra-plot variability of yields, spatial scale fully compatible with numerous on-going satellite missions. The aim of this study is to estimate the sunflower yield at the intra-plot spatial scale using the multi-temporal images provided by the Landsat-8 and Sentinel-2 missions. The proposed approach is based on a statistical algorithm, testing different sampling strategies to partition the dataset into independent training and testing sets: a random selection (testing different ratio), a systematic selection (focusing on different plots), and a forecast procedure (using an increasing number of images). Emphasis is put on the use of high spatial and temporal resolution satellite data acquired throughout two agricultural seasons, on a study site located in southwestern France. Ground measurements consist in intra-plot yields collected by a surveying harvesting machine with GPS system on track mode. The forecast of yield throughout the agricultural season provides early accurate estimation two months before the harvest, with R^2 equal to 0.59 or 0.66 and RMSE of 4.7 or 3.4 q ha^{-1}, for the agricultural seasons 2016 and 2017 respectively. Results obtained with the random selection or the systematic selection will be developed later, in a longer paper.

Keywords: sunflower; yield estimates; forecast; sampling strategy; Landsat-8; Sentinel-2; random forest

1. Introduction

Over the last 50 years, the world production and the areas allocated to sunflower have both increased steadily, with positive trends of around 0.6 million tones and 0.4 million hectares per year, respectively (trends derived from the statistics of [1]). In France, the culture occupies a large part of the useful agricultural area (behind wheat, barley, rapeseed and maize). The spatial distribution of the sunflower is highly disparate considering a department benchmark. With an average of 75,000 hectares in the last ten years, the Gers department ranks first, gathering more than 10% of the national area allocated to sunflower [2]. In view of the considered surfaces, the challenge is to identify suitable tools to monitor the culture, able to meet the constraints related to the crop growth cycle (agricultural season for several months) and the organization of the landscape (irregular and fragmented parcels).

The surface observation capabilities provided by the satellite missions constitute a useful tool, allowing to access to repetitive information on the surface states. They are conditioned by the characteristics of the embedded sensors operating in specific wavelengths (*e.g.*, visible, near or medium infrared, thermal or microwave) and delivering products at different spatial scales (pixel sizes ranging from meters to several kilometers). The contribution of satellite imagery for the monitoring of agricultural areas has been previously demonstrated, as evidence by the large range of applications dealing with various topics as the classification of land uses, the monitoring of the crop or the soil status (through the estimation of target parameters), the mapping of cultural practices or the detection of crop damage zones [3-5]. In the context of yield estimates, optical images have been widely used, providing a regular status of the photosynthetic activity of canopy. Estimates of yields were obtained using different approaches at spatial scale ranging from the region to the field [6-7]. Nevertheless only few studies deal with the monitoring of intra-plot variability of yields and rarely with the real-time aspect.

The objective of this study is to take both advantage of optical decametric satellite missions (by combining acquisitions performed by Landsat-8 and Sentinel-2A) together with ground data collected by sensors onboard harvesting machines to estimate yields of sunflower at the intra-plot scale (*i.e.*, spatial resolution of 30 m). The network of plots where ground measurements and satellite data were available is fully described in section 2. The proposed approach is based on random forest, considering reflectance as predictive variables and crop yields as target. The results are analyzed and discussed (sections 3 and 4), focusing on the estimates throughout the agricultural season which addresses the potential of real-time estimates.

2. Experiments

2.1. Materials

2.1.1. Study site

The study area is located in southwestern France in the Gers County. Surrounded by valleys, the territory is characterized by a great diversity of landscapes and types of soil comprising ustic luvisols, limestone, clay-limestone or more sandy soils. The county is subject to oceanic and mediterranean climatic influences, with a precipitation regime spatially and annually variable. The useful agricultural area occupies 71% of the territory (or 447 223 ha), being mainly dedicated to the cultivation of seasonal crops (cereal for 44.5% or oleaginous and proteinaceous for 24%) or forage crops and evergreen surfaces for 19% [2]. The present paper focuses on sunflower, for which the agricultural season delineated by the sowing and harvesting periods is observed from spring to autumn.

2.1.2. Intra-plot yield data

A network of 12 and 10 field plots of sunflower (representing 117 and 140 hectares (ha) respectively) were monitored to collect agricultural practices and the value of yields from farmers, during two successive agricultural seasons. Sizes of these fields ranged from 3.2 to 28.6 ha. The sunflower was sown during the spring, mostly during the month of April, and was harvested during the month of September. The mean values of yield ranged from 18.1 to 31.0 q ha^{-1} for 2016 and from and 16.9 to 24.1 q ha^{-1} for 2017, showing a variability depending on the considered plot, as evidence by the coefficients of variation (*i.e.*, CV = 100 × standard deviation / mean) ranged from 18 to 36%.

The yield values were derived from the data collected by the surveying harvesting machine with GPS system on track mode, namely the distance, the width of the cutting bar, the flux and the humidity of grain. The distance and the width of the cutting bar were first combined to obtain the area matching with the grain flux. The harvested yields were then computed and dry yields were last calculated by accounting for the humidity of grain. All the measurements performed in a pixel with a spatial resolution of 30 m were aggregated, avoiding the extreme values (*i.e.*, average plus three sigma or 99.7% of the values). Those maps of yields constitute the targeted variable of the statistical algorithm.

2.1.3. Optical satellite images

Table 1 presents an overview of the satellite images acquired during the two agricultural seasons. From April to September, regular high spatial resolution images were provided by Sentinel-2 (4 and 11 images for the years 2016 and 2017 respectively) and Landsat-8 (6 images for the year 2016).

Table 1. Characteristics of the satellite remote sensing data

Years	2016		2017
Satellites	Sentinel-2	Landsat-8	Sentinel-2
Dates (M-D)	05-21 ; 06-20	04-15 ; 06-09 ; 07-04	04-06 ; 05-06 ; 05-16
	07-10 ; 07-30	08-12 ; 09-06 ; 09-13	05-26 ; 06-05 ; 06-25
			07-05 ; 08-04 ; 08-14
			08-24 ; 09-13

The time series of Landsat-8 and Sentinel-2 images were provided by the Theia land data center. The images were processed using the software developed by [8], delivering level 2A products characterized by ortho-rectified surface reflectance. The data were first corrected from atmospheric effects and provided with a mask of clouds and their shadows on the ground (using a multi-temporal algorithm). All the images were finally resized at the same spatial resolution of 30 m.

In the present study, focus is on the comparable bands considering the wavelength, that is signals acquired in blue, green, red, near infrared and short wavelength infrared. The satellite images constitute the input data of the statistical algorithm described hereinafter, considering two cases: the widely used Normalized Difference Vegetation Index or the combination of the six reflectances.

2.2. Methods

The multi-temporal satellite acquisitions are used to estimate the yields throughout the agricultural season of crops. Beginning with the first image acquired after the period of sowing (April), the estimates are then performed with a cumulative number of successive images (*i.e.*, 1 to 10 or 11 images for the years 2016 and 2017, respectively), until the harvest of crops (September). For each estimation of yields, the dataset is partitioned into independent training and testing sets, using a ratio of 50/50.

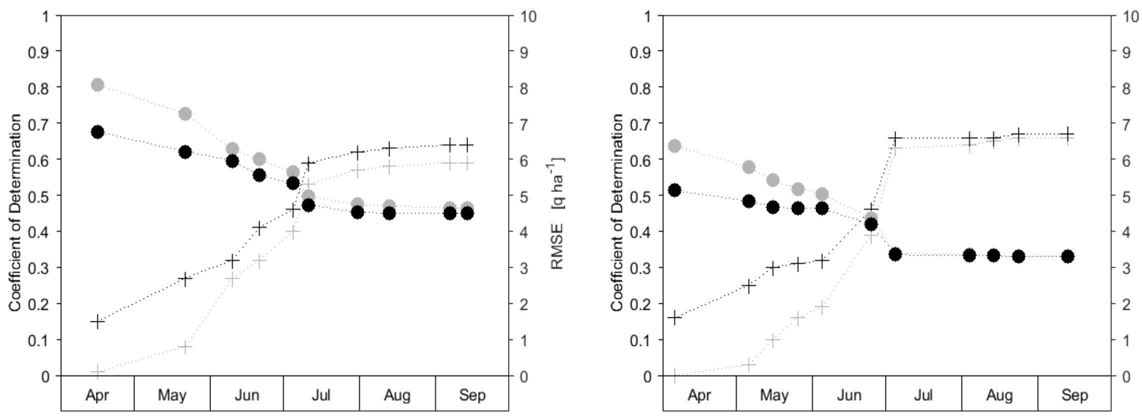
The estimation of yield is based on the statistical algorithm proposed by [9]. Random forest has been widely used in different fields providing accurate estimates of both qualitative (through classification) and quantitative (through regression) variables. This non-parametric approach consists in combining an ensemble of independent decision trees trained on different set of samples, through a procedure called bagging (abbreviation of bootstrap aggregating). Each decision tree is first trained on a subset of randomized samples derived from the initial dataset using bootstrap procedure, and used to provide estimates for the remaining independent samples. The decision trees are finally aggregated through the weighted mean of the ensemble of estimations, providing an estimate of the targeted variable. Unlike other statistical methods that may have limitations related to problems of over-adjustment, noise influence on data, or stability of results, random forests are particularly appropriate in multi-factorial context to account for non-linear relationships.

Coefficient of determination (R^2) and root mean square error (RMSE) are finally derived from the comparison between the observed and estimated yields. The analysis of results presented in the following section focuses on the independent testing set. Similar procedure was tested using artificial neural networks, showing slightly lower performances For the sake of concisness, only results obtained using random forest are presented hereinafter.

3. Results

3.1. Multi-temporal estimation of yields

The temporal evolution of statistical indexes associated to the yield retrieval is illustrated for the years 2016 and 2017 (Figure 1 a and b respectively), focusing on estimates based on NDVI and on the combination of different satellite reflectances. The statistical performance observed throughout the crop's agricultural season shows comparable general behavior, regardless the year or the satellite data considered. A strong increase of accuracy is first observed with the cumulative number of satellite acquisitions used for estimating yields. In 2016, the 6 images acquired from April to July (days 106 to 192) allow the R^2 to increase from 0.15 to 0.59, while RMSE decrease from 6.7 to 4.7 q.h^{-1} (considering the estimates based on the 6 reflectances). In 2017, 7 images are acquired during the same period (days 96 to 186) and statistics show higher performances, the R^2 increasing from 0.16 to 0.66 and RMSE decreasing from 5.1 to 3.4 q.h^{-1}. However, a notable difference is observed between the two considered years regarding the gain of accuracy. Indeed, the gain appears progressive in 2016, while the maximal increase is associated to the two images acquired at the end of June and at the start of July in 2017(days 176 and 186). Such behavior is closely related to the growth dynamic of sunflower, which can vary from one year to the other through the combination of agricultural practices (especially the dates of sowing) and climatic conditions (e.g., cold or warm conditions which can reduce or accelerate the growth rate). Then, performances saturate at specific phenological stage (flowering), and only slight gain of accuracy is obtained by the addition of new satellite images.

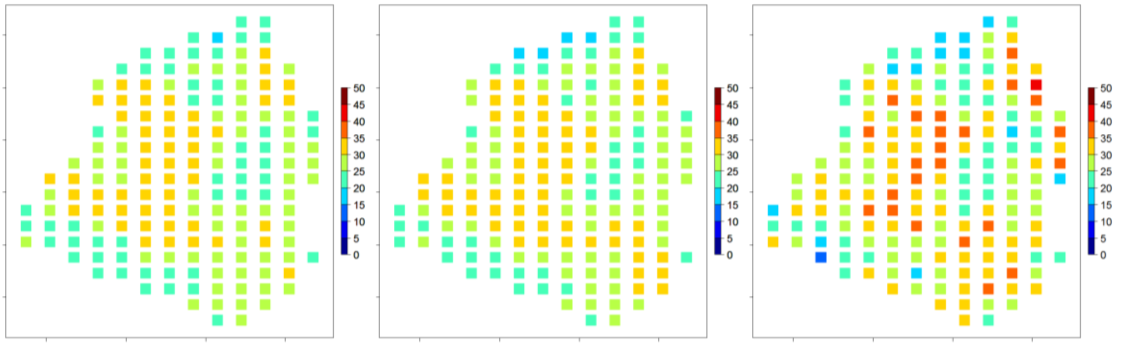


(**a**) (**b**)

Figure 1. Temporal evolution of the statistical performance (coefficients of determination and root mean square errors, crosses and dots respectively) associated to sunflower yield forecast using NDVI (in grey) or the combination of six reflectances (in black), for the years 2016 (left) and 2017 (right).

3.2. Mapping of yields at the intra-plot spatial scale

Finally, two maps of yield obtained during the agricultural season 2016 are compared to the intra-plot measurements (Figure 2). The estimated maps of yield observed during the months of July and September are derived from the 6 or 10 successively acquired images, respectively. The maps present similar intra-plot spatial patterns of low and high values of yield, and only few difference are observed between estimates performed two months before the harvest and those obtained just before the harvest (differences inferior to 2.8 q.ha^{-1}, with a mean value of 0.7 q.ha^{-1}). Such observation is confirmed considering all the pixel of the plot through the values of the averages of estimated yields (28.1 and 28.5 q.ha^{-1}), the standard deviations (3.2 and 3.2 q.ha^{-1}), or the range (from 20.0 to 33.3 q.ha^{-1} and 20.2 to 34.1 q.ha^{-1}). These two maps provide accurate estimates of the targeted variable (with relative error lower than 13% compared to measurements), nevertheless similar bias are observed, that is, extreme low and high measured values are not well reproduced by the statistical approach.



(**a**) (**b**) (**c**)

Figure 2. Maps of sunflower yield estimated two months before the harvest (**a**) and just before the harvest (**b**), together with measurements (**c**) collected on a plot dedicated to the cultivation of sunflower in 2016.

4. Discussion

The ability of obtaining accurate early estimates of yield has been demonstrated in previous agronomic studies focused on wheat or corn [10-12]. Interesting performance are thus observed during specific crop phenological stages, *i.e.*, during the elongation of the main stem of wheat or when the central stem of corn develops. For the sunflower, the accurate in-season estimates are observed during the first half of July, whatever the considered year. During this period, the fields cultivated with sunflower are observed at two characteristic phenological stages (BBCH scale numbers 5 and 6, [13]), corresponding to the inflorescence emergence and flowering. Moreover, the in-season performance observed for sunflower appear consistent with other studies based on the use of successive acquired optical and radar image [14,15], even if the yield estimates are not performed at the same spatial scale (intra-scale scale in the present study *vs.* field scale in the previous papers). In those studies, the levels of accuracy depend on both the considered crop (*e.g.*, $R^2 = 0.76$ and RMSE $= 7.0$ q ha^{-1} for wheat, R^2 of 0.69 and an RMSE of 7.0 q ha^{-1} for corn) and on the configuration of the satellite signals used as input variable of the statistical algorithm (*i.e.*, several combinations of frequencies and polarizations combinations were tested). Nevertheless, the statistics were obtained with a limited number of ground truth (due to the difficulty to obtain precise information on yield, the number of fields ~30) and for a single agricultural season. In the present study, the robustness of the approach is tested considering two agricultural seasons and large dataset of more than thousand of measurements for each studied year.

5. Conclusions

The proposed study addresses the potential of using multi-temporal optical images (Landsat-8 and Sentinel-2A) for the estimation of sunflower yields at the intra-plot spatial scale. The statistical approach takes advantage of both regular decametric satellite images acquired throughout two agricultural seasons and yield measurements collected on a network of plots. Random forest are implemented on independent training and testing sets, considering a forecast of yield throughout the agricultural season.

In the present study, the data were collected by a surveying harvesting machine with GPS system on track mode which presents the following advantages: (i) the ability of working at a spatial scale consistent with the size of pixels and thus considering the intra-field variations of yield (which are merged when working at the field or at the regional scale), and (ii) obtaining a large dataset, useful for testing the robustness of the proposed approaches (*i.e.*, the algorithm being trained and validated on more than one thousand of measurements). Moreover, the proposed approaches were solely based on series of optical satellite images and tested on two successive agricultural seasons, showing comparable trends and stability of the results regarding the levels of accuracy.

The estimation of yield throughout the agricultural season provide a demonstration of the potential of real-time approaches by considering an increasing number of successive satellite images. Accurate in-season estimation of yields were observed two months before the harvest (R^2 of 0.59 or 0.66 and RMSE of 4.7 or 3.4 q ha^{-1} for the years 2016 and 2017). Moreover, the map of yield obtained during the crop flowering presented spatial patterns consistent with those estimated just before harvest (correlation close to 0.96 between the two estimated maps).

Acknowledgments: The authors wish to thank the Theia land data center for providing the level 2A Landsat-8 and Sentinel-2 satellite images. In addition, the authors wish to thank the agricultural cooperative AGRO D'OC, especially Mrs. Guigues, Mr. Gambini and Mr. Hypolite, for their time and precious discussion, and the farmers who helped to collect the ground data.

References

1. Food and Agriculture Organization of the United Nations or FAO. Available online: http://faostat.fao.org/ (accessed on 25 February 2019).
2. Agreste. Available online: http://agreste.agriculture.gouv.fr/ (accessed on 25 February 2019).
3. Fieuzal, R.; Baup, F. Estimation of leaf area index and crop height of sunflowers using multi-temporal optical and SAR satellite data. *Int. J. Remote Sens.* **2016**, *37*, 2780–2809.
4. Marais Sicre, C.; Inglada, J.; Fieuzal, R.; Baup, F.; Valero, S.; Cros, J.; Huc, M.; Demarez,V. Early detection of summer crops using high spatio-temporal resolution optical images time series. *Remote Sens.* **2016**, *8*, 591.
5. Yang, H.; Chen, E.; Li, Z.; Zhao, C.; Yang, G.; Pignatti, S.; Casa, R.; Zhao, L. Wheat lodging monitoring using polarimetric index from RADARSAT-2 data. *Int. J.Appl. Earth Obs. Geoinf.* **2015**, *34*, 157–166.
6. Fieuzal, R.; Marais Sicre, C.; Baup, F. Estimation of sunflower yield using a simplified agrometeorological model controlled by optical and SAR satellite data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 5412–5422.
7. Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. of Environ.* **2010**, *114*, 1312-1323.
8. Hagolle, O.; Huc, M.; Villa Pascual, D.; Dedieu, G. A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VENμS and Sentinel-2 Images. *Remote Sens.* **2015**, *7*, 2668-2691.
9. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
10. Bushong, J.T.; Mullock, J.L.; Miller, E.C.; Raun, W.R.; Klatt, A.R.; Arnall, D.B. Development of an in-season estimate of yield potential utilizing optical crop sensors and soil moisture data for winter wheat. *Precis. Agric.* **2016**, *17* (4), 451–469.
11. Sharma, L.K.; Franzen, D.W. Use of corn height to improve the relationship between active opticals sensor readings and yield estimates. *Precis. Agric.* **2014**, *15*,331–345.
12. Yin, X.; Hayes, R.M.; McClure, M.A.; Savoy, H.J. Assessment of plant biomass and nitrogen nutrition with plant height in early-to mid-season corn. *J. Sci.Food Agric.* **2012**, *92*, 2611–2617.
13. Meier, U. Stades Phénologiques des Mono- et Dico-tylédones Cultivées. Centre Fédéral de Recherches Biologiques pour l`Agriculture et les Forêts, 2001. www.agroedieurope.fr/ref/doc/BBCH.pdf
14. Fieuzal, R.; Baup, F. Forecast of wheat yield throughout the agricultural season using optical and radar satellite images. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *59*, 147-156.
15. Fieuzal, R.; Marais-Sicre, C.; Baup, F. Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *57*, 14-23.