Correlation Between Satellite Multispectral Imagery and Combine Harvester CANbus Data for Corn Yield Assessment

Published Version:
Miralles, G.S., Biglia, A., Mattetti, M., Tortia, C., Gay, P., Comba, L. (2025). Correlation Between Satellite Multispectral Imagery and Combine Harvester CANbus Data for Corn Yield Assessment. Cham : Springer [10.1007/978-3-031-84212-2_75].

Availability:
This version is available at: https://hdl.handle.net/11585/1030730 since: 2025-11-30

Published:
DOI: http://doi.org/10.1007/978-3-031-84212-2_75

(Article begins on next page)

Correlation between satellite multispectral imagery and Combine Harvester CANbus data for corn yield assessment

Gica Stefanescu Miralles[1,2*][0009-0008-6472-707X], Alessandro Biglia[1][0000-0002-4256-095X], Michelle Mattetti[3][0000-0002-2473-5250], Cristina Tortia[1][0000-0002-8479-886X], Paolo Gay[1][0000-0002-8342-425X], Lorenzo Comba[1][0000-0001-5466-0918]

[1] Department of Agricultural, Forest and Food Sciences (DiSAFA) – Università degli Studi di Torino, Largo Paolo Braccini 2, 10095 Grugliasco (TO), Italy
2 Department of Computer Science (DI) – Università degli Studi di Torino, Via Pessinetto 12, 10149 Torino (TO), Italy
[3] Department of Agricultural and Food Sciences, Alma Mater Studiorum – University of Bologna, Viale G. Fanin 50, 40127 Bologna (BO), Italy

gica.stefanescumiralles@unito.it

Abstract. Adoption of digital tools has led to relevant advancements in the agricultural sector. In particular, precision agriculture techniques have shown effective applications in farming practices. Integrating multisource data is a crucial task for improving agricultural management, and in this context the reliable assessment of crop yield by remote sensed imagery plays a relevant role. In this work, the correlation between satellite multispectral imagery and corn yield was investigated. Sentinel-2 satellite mission was selected as source of multispectral data, with 10-meter spatial resolution and 4-6 days revisit time. A corn field in Ferrara was considered as a case study, with the dataset consisting of 17 multispectral images acquired in days with cloud coverage under 7%. The reference yield map was computed using CANbus data from a combine harvester, and its correlation with NDVI and GRVI indices was explored throughout the whole corn life cycle. In addition, the effect of applying a gaussian filter in the raster CY spatial distribution was explored. Results showed an overall good correlation between remotely sensed tiles and in-field farm data. The Pearson correlation coefficients showed a sharp increase during the vegetative stage of the crop (May and June), followed by a slowly decreasing plateau in July and August dates. Imagery from early June provided the highest correlations. The application of a gaussian filter in the CY map revealed an enhance in correlation of more than 25%. These results pave the path to the development of machine learning based methods exploiting multi-band-multi-temporal data for CY estimation.

Keywords: Precision Agriculture, Remote sensing, CANbus, Yield assessment, Machine Learning.

1 Introduction

Effects of global warming, such as the increase in drought severity or the trend towards warmer summers, together with the demands of a growing population, dictate huge challenges in the agricultural sector [1,2]. Farmers are pushed to increase the productivity and the efficiency of their practices [3], thus, becoming crop yield one of the essential features to assess [4].

In this context, automatic digitalization of field activities and Earth Observation Systems (EOS) represent key technologies, providing huge amounts of data with potentially relevant information [5,6]. The recent improvement of capabilities in EOS (better resolutions and increased revisit times) has resulted in more opportunities to apply precision agriculture in crop monitoring and management [7]. The usage of CANbus logger sensors installed into farm machinery, has shown to be an effective solution to collect reliable ground data, rich of information [8,9]. Merging accurate real-world ground data and modern satellite imagery, suitable datasets can be obtained to be exploited by machine learning methods, with the final aim to estimates relevant crop parameters [10]. However, before applying complex tool on relevant amount of data, an investigation of the correlation between data acquired from different sources have to be performed. A significant example of such approach is the assessment of the correlation between corn yield (CY) and remotely sensed multispectral imagery (SAT), which can lead to Machine Learning (ML) models able to predict crop yield in advance [11].

A crop of interest to which studies of this type should be focused is corn (*Zea mays*), a major cereal grain cultivated worldwide, that accounts for a production at European level around 56 million tons, corresponding to the 23% of the total cereal production in the region [12]. Corn represents a major summer cultivation, and its life cycle can be divided into vegetative and reproductive stages [13]. During the vegetative phase, the plant progress through leaf stages, and usually lasts around 60 days since its emergence. The reproductive phase follows, starting with silking and pollination, followed with kernel development. When kernels are filled to full capacity the maturity is reached. The reproductive phase extends for a period of approximately two months as well. Few studies have been published up to date on specific CY assessment/prediction using SAT, especially concerning regions in the south of Europe.

At country level in the US, Kang *et al* [14] showed that the XGBoost model achieves reasonably good CY estimations, within 3-4 month prior to harvest. Another case-study in Malawi, East Africa, tested PlanetScope and Senitnel-2 satellite imagery on CY prediction at farm-level [15], reaching however moderate estimates (R^2 ~ 0.52). The state-of-the-art claims for a deeper understanding of the correlation between in-field CY and SAT data, setting the basis to construct featured engineered datasets to be used by ML models to estimate CY.

In this work, the correlation trends inferred between Sentinel-2 multispectral imagery and CY maps, derived by CANbus data is investigated. To this aim, the study explores different preprocessing steps to build up a coherent and representative dataset, so that each pixel of satellite imagery, with spectral data, has its respective local CY data. This investigation sets a basis to develop further studies aimed at robust and reliable CY estimation from SAT imagery, exploiting ML based models.

2 Materials

2.1 Study Area

A corn field located in Emilia Romagna (44°41'46'' N - 11°33'36'' E, North Italy), at an altitude of approximately 30 m AMSL, with an extension of 9.3 ha, was selected as case study. The study considered the 2020 corn growing season, with field sowing at the beginning of April and harvesting in September. The environmental condition in the region during the considered period were characterized by an average rainfall of 822 mm and an average temperature of 13.4°C in the region.

2.2 Multispectral Satellite Imagery

In this work, Sentinel-2 mission was chosen as source for the satellite multispectral imagery. Sentinel-2 presents a high spatial resolution (up to 10 meters) and a short revisit time (up to 2/3 days) [16], that together with its open-source platform makes it a valuable solution for accessible precision agriculture solutions. In addition, the spectral bands provided by the mission have proved to be effective in agricultural applications, due to being explicitly sensitive to changes in chlorophyll content and canopy structures [17].

For the case study field, a total of 17 multispectral tiles was downloaded from Copernicus Open Access hub [18]. The acquired images date over all the 2020 corn season, from sowing until harvesting, and include only days with cloud coverage under 7%. Each tile was georeferenced in the UTM/WGS84 projection and was composed of 10 raw bands, located in a range including visible, red edge, near infrared and shortwave infrared wavelengths. In Fig.1, a graphical representation of the multispectral imagery of the considered field in the visible spectrum (Red, Green and Blu bands) is reported.

2.3 Combine Harvester yield data

A custom-made CANbus data-logger developed described in [19] was installed on a combine harvester New Holland CR7.90. This tractor is particularly rich in its embedded sensors, which allows for comprehensive identification of the operating state. The data logger was set up to automatically record all CANbus messages anytime the machine was turned on, thus, the operators were not responsible for the recording process. For the purpose of this study, the signals of interest recorded by the logger were the vehicle Ground Speed (km/h) and the Crop Flow (kg/s), which are used, together with the cutter bars Work Width (mm), to compute the CY as defined by Eq. (1).

$$CY = Crop\ Flow\ /\ (Ground\ Speed\ *\ Work\ Width) \quad (1)$$

Moreover, a GNSS (global navigation satellite system) receiver with an update rate of 10 Hz was installed in the tractor to monitor its position (Latitude and Longitude). In Fig.1, the vectorial GNSS path of the combine harvester is represented, recorded during the harvesting task. The combine harvester was monitored, and the data recorded during the 9th and 10th of September 2020, when the harvest of the case study field took place.

3 Methods

3.1 Data pre-processing

The information conveyed from the CANbus logger allowed computing the reference CY using Eq. (1). However, the vectorial format of the reference CY required a rasterization in order to make it compatible with the SAT imagery. For each pixel of the SAT dataset, a point-to-raster method was applied, in which the respective raster CY value was computed by averaging within the subset of vectorial data within each pixel area.

The CY raster map resulting from this analysis (CY_{raw}) consisted of a set of 1569 pixels, being each pixel value paired with its respective spectral band values in the SAT tiles. A graphical representation of the CY raster map in false colour is reported in Fig. 2b.

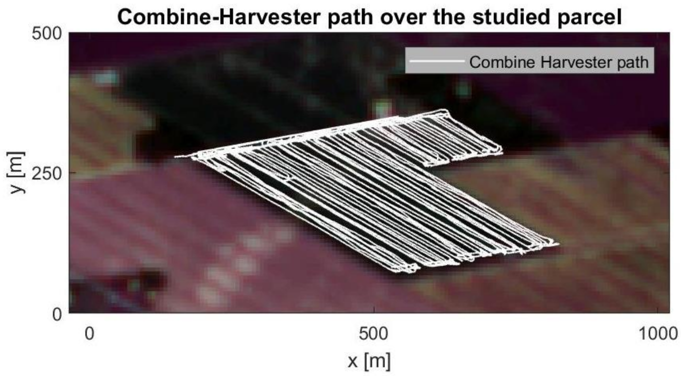

Fig. 1. RGB visualization of the multispectral satellite imagery (11th August 2020) of the case study field, with the vectorial GNSS position of the combine harvester path during the harvesting tasks on top as a white line.

3.2 Gaussian filter application

A further step was considered to refine the CY_{raw} map. The adopted rasterization method results in a spatial distribution presenting certain discontinuities, that might be mainly linked to GNSS sensor uncertainty [20]. Accuracy of yield mapping can be increased by refining phases, e.g. by empirical likelihood probability weights [21]. In this work, a gaussian filter was used as spatially smoothing moment robust to outliers, to refine the CY_{raw} map and enhance possible correlations with the spectral imagery from Sentinel-2.

The filter was applied on the whole parcel data, and the resulting map (CY_{filt}) was considered parallelly to the original one (CY_{raw}) for further correlation assessments. In Fig. 2c, a graphical representation of the CY_{raw} map in false colour is reported. In addition, the frequency distributions of CY values of both the CY maps are displayed, together with the associated gaussian curve fits.

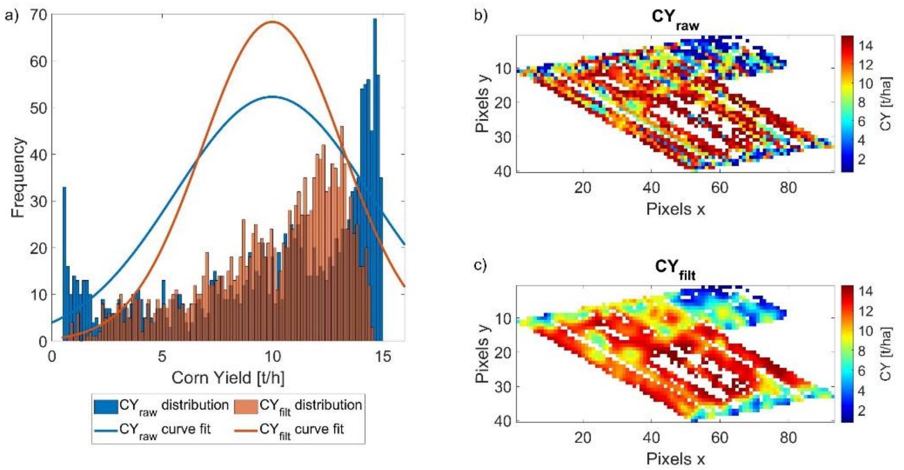

Fig. 2. (a): Corn Yield distribution of the case study parcel, both before and after the gaussian filter application. The coloured lines correspond to the gaussian curve fits of the respective distributions. (b) and (c): Corn Yield raster map of the case study parcel before (b) and after (c) the gaussian filter application.

3.3 Vegetation Index maps

To explore the potential informative content provided by the spectral data, the NDVI and GRVI vegetation indices (VIs) were computed using the raw multispectral imagery. Those indices, selected on the base of state-of-the-art works on crop yield assessment [10,22], are defined in Eq. (2) and Eq. (3), respectively.

$$NDVI = (NIR - RED) / (NIR + RED) \quad (2)$$

$$GRVI = (GREEN - RED) / (GREEN + RED) \quad (3)$$

nm). The GREEN band in Eq. (3) correspond to the band 3 (560 nm). The NDVI and GRVI maps was computed for all the available Sentinel-2 tiles. Of course, due to the definition of CY_{raw} and CY_{filt} spatial distributions, each pixel of VIs maps has its respective CY value. The correlation between the VIs maps and the CY was investigated computing the Pearson coefficient of determination (R^2), and the NDVI-CY_{raw}, NDVI-CY_{filt}, GRVI -CY_{raw}, and GRVI -CY_{filt} pairs have been considered.

4 Results

In Fig. 3a, the temporal evolution of the R^2 between NDVI and CY_{raw}/CY_{filt} is plotted for all dates available. Equivalently, Fig. 3b displays the same evolution for the case of the GRVI.

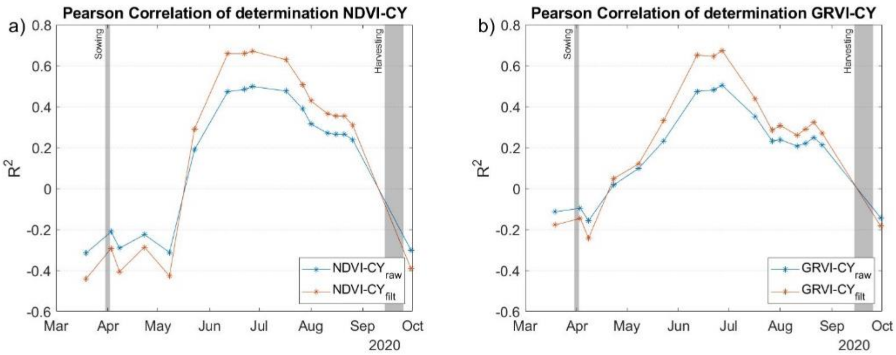

Fig. 3. (a) Pearson Correlation Coefficient (R^2) between NDVI maps and CY_{raw}, (blue) and CY_{filt} (red) and (b) between GRVI maps and CY_{raw}, (blue) and CY_{filt} (red). The sowing and harvesting stages are displayed as labelled grey areas.

The results showed how the values of R^2 vary significantly depending on the specific moment of the corn life cycle in which Sentinel-2 tiles were acquired. A general trend is observed both for NDVI and GRVI indices: during the first weeks after sowing, R^2 remains low (from -0.2 to -0.4 for NDVI and from -0.2 to 0 for GRVI). After approximately one and a half month after sowing (a bit earlier for the GRVI index), the R^2 values present a sudden and marked positive increase, very sharp in the case of the NDVI and slightly more progressive for the GRVI. Such an increase reaches its maximum in mid-June dates, coinciding with the conclusion of the vegetative period of the crop, with R^2 values up to 0.67. The Pearson correlation coefficients remain in a plateau-like shape over the beginning of July, followed by a progressive decrease during August and September. Nevertheless, R^2 values remain positive on these dates. Finally, a marked decrease in R^2 values is observed after the harvesting, returning to values equivalent to the standards observed in the sowing phase. This trend is similar to those described in other works [9, 12] and suggests the idea that before emergence of the plant, the spectral response is ground-like and does not provide relevant information to link with CY. However, after emergence and during the vegetative period, the corn canopy presents a spectral response consistent with its vigour and, in this study, this crop characteristic showed a good correlation with the obtained local yield. During the reproductive phase (July, August and September), the canopy size remains constant and the plants gradually dry, resulting in a spectral response less correlated with the CY.

The CY filtering phase was found to lead to positive effect on the relation between SAT data e crop yield: the Pearson correlation coefficient was found to be increased, considering both NDVI and GRVI indices. More in details, in the case of NDVI, the correlation with was, on average, about 25% higher when considering the filtered CY_{filt} distribution, while for the case of the GRVI, the average increment was up to 34%. This result justifies the usage of spatial filtering of CANbus derived data and presents the gaussian filter as a proper candidate for what concerns the presented specific application.

5 Conclusions

The presented work propose a procedure to jointly exploit CANbus ground truth corn yield data and multispectral imagery from Sentinel-2 mission. A rasterization method is introduced to make the multisource data compatible, and a filtering phase using a gaussian filter is applied to enhance yield mapping accuracy. NDVI and GRVI indices, computed from the raw satellite spectral data, have shown a good correlation with CY during the months corresponding to the vegetative period of the crop. Such correlation has been further increased by considering the filtered yield data.

This work sets a basis to obtain suitable multi-source datasets, that might be used as inputs for ML models, with the final aim to accurately estimate corn yield. In addition, the observed trends infer the possibility to predict corn yield even several weeks in advance to harvest period, which could translate to a relevant tool for farm management decisions.

Funding

This research was partially funded by the project “DATA-BUS - Digital Agriculture Technology to Achieve data to Build User-friendly Sustainability indicators” (PRIN 2020, Prot. 2020SCNF4L) and was partially carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors’ views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

1. 1. Vicente-Serrano S, Lopez-Moreno J I, Beguería S, Lorenzo-Lacruz J, Sanchez-Lorenzo A, *et al.* (2014). Evidence of increasing drought severity caused by temperature rise in southern Europe. Environmental Research Letters 9(4).
2. Sue Wing I, De Cian E, Mistry M N (2021). Global vulnerability of crop yields to climate change. Journal of Environmental Economics and Management 109:102462.
3. Varani, M., *et al.*, (2023) Correlation between power harrow energy demand and tilled soil aggregate dimensions. Biosystems Engineering, 225, 54–68.

4. Rosegrant M W (2003). Global food security: challenges and policies. Science 302:5652.
5. Comba L, *et al.,* (2020). Semantic interpretation and complexity reduction of 3D point clouds of vineyards. Biosystems Engineering, 197, 216-230.
6. Kohrs R, Lazzara M, Robaidek J, Santek D, Knuth S *et al.,* (2014). Global satellite comp sites 20 years of evolution. Atmospheric Research 8(34):135-136.
7. Mazzia V, Comba L, Khaliq A, Chiaberge M, Gay P (2020) UAV and Machine Learning Based Refinement of a Satellite-Driven Vegetation Index for Precision Agriculture. Sensors 20(9):2530-2546.
8. Mattetti M, Maraldi M *et al.,* (2021). Outlining the mission profile of agricultural tractors through CAN-BUS data analytics. Computers and Electronics in Agriculture 184
9. Angelucci L, Mattetti M (2024) The development of reference working cycles for agricultural tractors. Biosystems Engineering, 242:29-37.
10. Marshall M, Belgiu M, Stein A *et al.,* (2022). Field-level crop yield estimation with PRISMA and Sentinel-2. ISPRS Journal of Photogrammetry and Remote Sensing 187:191210.
11. Segarra J *et al.,* (2022). Farming and Earth Observation: Sentinel-2 data to estimate with field wheat grain yield. International Journal of Applied Earth Observation and Geoinformation 107:102697.
12. European Parliament, 2019. www.europarl.europa.eu/thinktank/en/document /EPRSBRI(2019)640143 last accessed 14/03/2024.
13. Abendroth LJ, Elmore RW, Boyer MJ, Marlay SK *et al.,* (2011). Corn growth and development. Ames (IA): Iowa State University Cooperative Extension Service. PMR 1009.
14. Kang Y *et al.,* (2020). Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. Environmental Research-Letters 15(6):064005.
15. Li C, Lu Y, Anghileri D *et al.,* (2022). Maize Yield Estimation in Intercropped Smallholder Fields Using Satellite Data in Southern Malawi. Remote Sensing 14(10):2458.
16. Khaliq A, Comba L, Biglia A, Ricauda Aimonino, D, Chiaberge M, Gay P (2019) Comparison of satellite and UAV-based multispectral imagery for vineyard variability assessment. Remote Sensing 11:436-454.
17. Frampton W J, Dash J, Watmough G, Milton E J (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS Journal of Photogrammetry and Remote Sensing 82:83-92.
18. Copernicus European Union ©, 2023. https://browser.dataspace.copernicus.eu last accessed 30/03/2023
19. Mattetti M, Medici M, Canavari M, Varani M (2022) CANBUS-enabled activity-based costing for leveraging farm management. Computers and Electronics in Agriculture, 194:106792.
20. Reitz P, Kutzbach H.D. (1996). Investigations on a particular yield mapping system for combine harvesters. Computers and Electronics in Agriculture 14(2):137-150.
21. Wen K, Wu X, Leatham, D.J. *et al.,* (2021). Spatially Smoothed Kernel Densities with Application to Crop Yield Distributions. JABES 26:349–366.
22. Primicerio J, Gay P, Ricauda D, Comba L, Matese A, Di Gennaro S.F. (2015). NDVI-based vigour maps production using automatic detection of vine rows in ultra-high resolution aerial images. Precision Agriculture 465-470.