TERRAENVISION 2025
Vol. 5 TNV-3430
Granada, Spain, 8-11 July, 2025

Analysis of Tomato Yield Predictions with Sentinel-2 and UAV-Based NDVI: a case study in Northern Italian farms

[1]Avinash Kafle, [2]Andrea Caporrella, [2]Federica Carmagnola, [2]Giuseppe Caporrella and [3]Fabio Dell'Acqua

[1]*Department of Environmental and Earth Sciences - University of Pavia, Pavia*
[2]*Innova-Tech-Agro*
[3]*Department of Electrical, Computer, Biomedical Engineering - University of Pavia, Pavia*

Abstract

This study presents a technique for assessing the Sentinel-2 and UAV-derived NDVI competence in monitoring and predicting yield in three different tomato fields of Northern Italy: "Castello" (29.15 hectares) in the province of Piacenza, "Andrea" (4.47 hectares) and "CEO" (2.67 hectares) in the province of Alessandria.
Sentinel-2 Level-2A tiles were used to derive spaceborne NDVI on the concerned fields. Additionally, a DJI MAVIC 3M UAV equipped with a multispectral sensor at 2×2 m^2 resolution was used to create airborne-NDVI maps of the fields from the day of the anticipated top vegetative stage and from the day before harvest. Nine different 2×2 m^2 grids at each field, three each from low-, medium- and high-NDVI field zones were selected to conduct yield sampling. In-depth relations between the NDVI of the sample grid and the sample grid yield were extracted using both remotely sensed data from July to September 2024.
Coefficient of determination (Adj. R^2), Root mean square error (RMSE) and Relative root mean square error (RRMSE) were computed as cost functions of prediction techniques. NDVI values from specific dates that had achieved the best Adj. R^2 to yield were selected to create a linear regression for yield prediction at the sample level. That same relation was then used to predict yield at the field level. Relating sample yield with predicted sample yield and total field yield with predicted field yields, Relative Root Mean Square Error (RRMSE) was calculated to define comparative prediction accuracy at sample and field level.
It was found that, for all three fields, both spaceborne- and airborne-NDVI yield prediction scored Adj R^2 above 50 % from the early flowering to the late fruiting stage, confirming that NDVI is a good indicator of yield during this period. Specifically, the "Castello" field featured the best spaceborne-NDVI-to-yield relationship 30 days prior to harvest; then, tomatoes were harmed by excessive rainfall during the harvesting stage. This unexpected factor resulted in an overprediction of 5.68% at sample level and 16.07% at field level as measured by RRMSE. This significant inaccuracy reflected the damage

caused to tomatoes by excess rainfall. Fields “Andrea” and “CEO”, which didn’t suffer late rainfall damage, had best spaceborne-NDVI-to-yield relation at 22 and 15 days before harvest, respectively, with more accurate predictions: on “Andrea” the overprediction of yield was measured by a RRMSE of 4.89% at the sample level and 10.13% at the field level, while for “CEO” it was 12% at sample level and 9.36% at field level. At the same time, it was found that airborne-NDVI caused underprediction of yield at all three experimental fields and performed better than spaceborne-NDVI only for “CEO”. Here, UAV underpredicted the yield by RRMSE of 6.05% at the sample level and 7.85% at the field level. It was found RRMSE derived from airborne data sources was smaller and more consistent between sample to field level, 6.05% at sample level and 7.85% at the field level to spaceborne data source, 12% at sample level and 9.36% at field level. “Andrea” and “Castello” reported higher RRMSE and less consistence with spaceborne data.

In conclusion, this study reveals some error patterns due to inherent limitations of each data source: scarce spatial resolution for Sentinel-2 and scarce temporal resolution for UAV. In general, Sentinel-2 data lead to overprediction and greater inconsistency with respect to UAV data. However, more frequent UAV campaigns would be advisable to increase accuracy. Still, relationship between NDVI and yield remained strong between early flowering to late fruiting stage, and the ideal time window to anticipate tomato yield is between 15 to 30 days prior to harvest.

Keywords: Yield prediction, tomatoes, multispectral data, airborne sensing, spaceborne sensing, NDVI

Acknowledgments: This publication was partly funded by the European Union - NextGenerationEU, Mission 4 Component 1.5 - ECS00000036 - CUP F17G22000190007”