Data assimilation of LAI improved crop growth modeling: Comparison between in-situ measurements and satellite estimations⋆

Marija Kopanja*1,*,†*, Gordan Mimić*1,†*, Hilde Vaessen*2,†*, Nevena Stevanović*1,†*, Nikola Stanković*1,†*, Maša Buđen*1,†*, Dragana Blagojević*1,†*, Bernardo Maestrini*2,†*, Nataša Ljubičić*1,†*, Sanja Brdar*1,†* and Frits van Evert*2,†*

¹ BioSense Institute, Zorana Đinđića 1, Novi Sad, 21000, Serbia

² Wageningen University & Research, Droevendaalsesteeg 4, Wageningen, 6708 PB, The Netherlands

Abstract

A digital twin of crops was created by integrating real-time measurements into existing crop growth models. The study compared the performance of the WOFOST model in predicting yield for maize and soybean, with and without data assimilation of leaf area index (LAI) measurements. Data assimilation improved the model's performance, especially when using in-situ measured LAI, leading to more accurate yield predictions.

Keywords

Digital twin, Data assimilation, WOFOST, Ensemble Kalman Filter, LAI

1. Introduction

A concept of digital twin originates from engineering, and it has been developed to describe physical and chemical processes in the machines. Recently, it is finding the application in the green life sciences, and particularly it can be used in digital farming [1]. Gathering data on weather, soil conditions, and key plant growth parameters throughout the season enables running models for predicting crop growth and yield under specific abiotic conditions. A digital twin of a crop can be created by the integration of near real-time measurements from various sensors into existing crop growth models. The purpose of the digital twin in digital farming is to simulate crop growth in near-real time and adjust the state variables to reflect the situation in the field.

A systematic review about the current status and future perspectives of assimilation of remote sensing data into crop growth models gave insight into a number of challenges ahead, some of which are coming from the observation [2]. The first challenge is to combine observations from different sensors, with different spatial and temporal resolution, in a harmonized way, and the second one is to provide accurate estimates of the uncertainty of the retrieved variables. However, it is shown that data assimilation (DA) is a valuable technique for estimating variables related to crop growth, such as: soil moisture, biomass and leaf area index (LAI), by using satellite remote sensing data [3, 4].

Some authors developed and successfully validated a process-based soybean model for a prediction of within-field yield variability by coupling LAI retrieval from Sentinel-2 into the crop model, and with that approach they managed to optimize some soil-related parameters [5]. In [6], the authors assimilated sensor data in a crop growth model of potato, using Planet topsoil moisture C-band data and canopy reflectance from Sentinel-2 to calculate LAI, and as a result the final yield estimate is being corrected to a value closer to what was measured in the field at the end of the season.

In this study, the ensemble Kalman filter was exploited. We used data collected at field level during the 2023 growing season and the WOFOST crop simulation model for a prediction of maize and soybean yield. We investigated the performance of a model after DA of LAI, by comparing in-situ measurements and satellite estimations.

2. Data acquisition

The experiment was conducted in Čenej, Vojvodina (Serbia), covering an area of 56 hectares, where soybeans and maize were cultivated, each occupying 28 hectares. The field (latitude=45.410290, longitude=19.804531) was divided by partnering agricultural company into 20 different management zones based on multi-year average values of the NDVI in growing season and the zones were used as stratification for the soil texture sampling. In this study, six zones were designated for each crop. The field topography is fairly plain and the existing soil types are as follows: loam (medium), clay loam (medium fine) and clay (fine). Before planting, soil samples were taken from each zone to conduct a basic soil chemical analysis. The planting of both crops was carried out on April 27, 2023. Automatic weather station (PrecMet, Kite ZRT, Hungary) was positioned at the edge of the field to gather daily data on air temperature, precipitation, air humidity, wind speed and direction. Surface solar radiation was downloaded from ERA-5 reanalysis dataset available at Climate Data Store [7].

To track the progress of plant development, the LAI of soybean and maize was consistently monitored throughout the growing season. For soybean, measurements were taken from the V4 stage until harvest, while for maize, measurements were conducted from the V6-V7 stages until harvest, with readings taken biweekly, resulting in four observed values. In each zone, plant samples were collected from a 1 m² area and the whole sample was passed through an area meter (LI - 3100C) to measure LAI. At the corresponding time points LAI was also estimated using cloud-free Sentinel-2 images processed through Sentinel Application Platform (SNAP), resulting in 10 values during the growing season. Harvesting of crops was performed with the John Deere S770 combine equipped with the John Deere Harvest Monitor™ system for yield monitoring during harvest. This system is integrated with GPS, enabling us to track reliable yield data at appropriate locations in the field and thoroughly analyze our management zones (Figure 1). For each zone, the average yield was calculated and further compared with the results of the crop growth model.
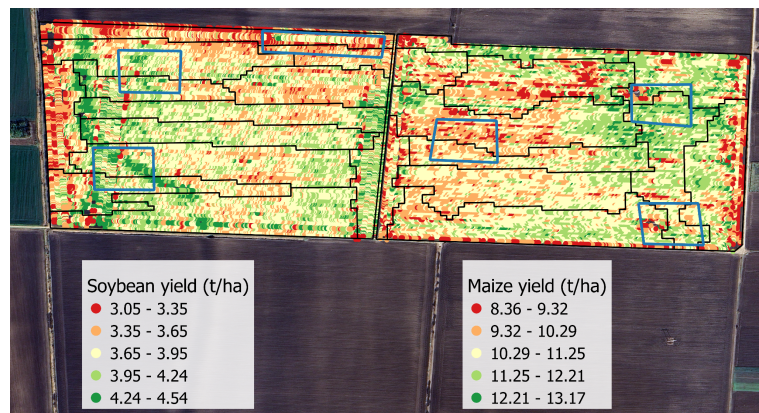

Figure 1: Yield observed in the zones of soybean (left) and maize (right), separated with a straight line.

3. Methodology

3.1. WOFOST

The WOFOST is a dynamic, explanatory model that simulates crop growth with a temporal resolution of one day. The crop growth is simulated based on several eco-physiological processes, including phenological development, CO_2-assimilation, transpiration etc. In general, three different hierarchical levels of crop growth (potential, limited, reduced) can be distinguished that correspond to three levels of crop production [8]. In WOFOST implementation there are three levels of crop production: potential, water-limited, and nutrient-limited. The focus of our work was on the water-limited crop growth model due to environmental constraints regarding the availability of water. For that, the implementation of WOFOST version 7.2 is used in the Python Crop Simulation Environment (PCSE) 5.4.0. Accordingly, a yield estimation is retrieved from the variable Total Weight Storage Organs (TWSO) that represent the harvestable product of the crop as dry weight in kg/ha.

In our analysis, the WOFOST water-limited crop growth model is used for two crops: maize, and soybean. The modeling concept is the same for both crops, while the differences are expressed by differences in the value of model parameters. Three different types of model parameters are required for running the WOFOST. First group of parameters consist of a set of crop parameters, a set of soil parameters, and a set of site parameters. Second group is related to weather data. Lastly, parameters related to agro management actions are required to specify the farm activities that will take place on the field.

The crop parameter files Grain_maize_203 and Soybean_902 are chosen for maize and soybean from a given list of pre-defined crop files, appropriate to the varieties used in the field experiment, based on the growing season duration. Variables related to soil moisture required by WOFOST enable a connection between the WOFOST crop simulation model and the underlying soil model. The knowledge of the agronomists is used to determine the most appropriate soil file from the list of available soil types. In the data acquisition stage it is ensured that soil types are known for each zone. Soil data files are named EC-1 (coarse), EC-2 (medium), EC-3 (medium fine), EC-4 (fine), and EC-5 (very fine). For our experimental analysis it is determined that EC-2, EC-3 and EC-4 correspond to the soil types in the field. More precisely, for six zones of maize are used EC-2 and EC-3, while for six zones of soybean are used EC-2, EC-3 and EC-4.

Additionally, two parameters that needed to be defined before running WOFOST are maximum soil-limited rooting depth (RDMSOL) and initial amount of available water in total rootable zone (WAV). A RDMSOL was set to be 150 cm based on knowledge of local pedology from the field agronomist. The sensitivity analysis is used to determine WAV and a value of 40 is obtained for both maize and soybean.

3.2. Ensemble Kalman filter and data assimilation

Ensemble Kalman filter is a sequential DA method, meaning the update of the state variables of the crop model is done for each date until an observation is available. The study focused on the LAI, and investigated the performance of the crop growth model after DA of LAI, obtained in two different ways: in-situ measured and estimated from satellite images. Adjustment of each model state was based on standard deviation of the observed and the simulated LAI, as a measure of the uncertainty of the observation and model state.

DA was performed using the ensemble Kalman filter, with 50 ensemble members, varying the total initial dry weight (TDWI) as initial condition, and life span of leaves (SPAN) as a parameter, both related to the leaves. The higher value of SPAN has a positive effect on LAI during the final growth period (crop maturity) while on the other hand, during the initial phase of crop establishment, LAI at emergence is calculated from TDWI [8]. The TDWI and SPAN are treated as Gaussian random variables with a mean value and a standard deviation which are determined from several runs of the model for each crop individually. For maize, the mean value and standard deviation of the TDWI distribution are 50 and 10, respectively, while for soybean corresponding values are 90 and 10. Regarding the distribution of SPAN, for maize the corresponding values of mean and standard deviation are 40 and 3, whereas for soybean the mean and standard deviation are 50 and 3.

4. Results and discussion

The WOFOST model was run for each zone and the results for yield prediction were compared in three different cases: simulation without DA, DA of in-situ measured LAI (M), and DA of LAI estimated from satellite images (E). Examples for the ensemble Kalman filter DA of in-situ measured LAI and satellite LAI are given in Figure 3 for maize and soybean.

Performance of the WOFOST model was evaluated for the final yield, comparing it to the observed yield, by using standard error metrics such as mean absolute error (MAE) and root mean squared error (RMSE). The average yield for maize is 10 670 kg/ha, while the average yield for soybean is 3 800 kg/ha. In Table 1 are reported averaged results over all subzones for a given crop. Notably, for maize the results indicate that model with DA of LAI estimated from satellite images overperformed two other cases, while for soybean the model with in-situ measured LAI performed better compared to the model without DA and model with DA of satellite LAI.

Table 1
Error metrics of the yield prediction for maize and soybean, with and without data assimilation (DA)

	Maize		Soybean	
	MAE	RMSE	MAE	RMSE
without DA	1198,50 kg/ha	1492,57 kg/ha	240,71 kg/ha	275,76 kg/ha
DA in-situ LAI	947,27 kg/ha	1011,03 kg/ha	217,43 kg/ha	231,92 kg/ha
DA satelliteLAI	543,69 kg/ha	632,59 kg/ha	227,61 kg/ha	256,27 kg/ha

Visual representation of the yield estimation using WOFOST simulation with/without DA versus field measured yield for each subzone (annotated next to corresponding dot) and both crops, maize, and soybean, is given in Figure 2.
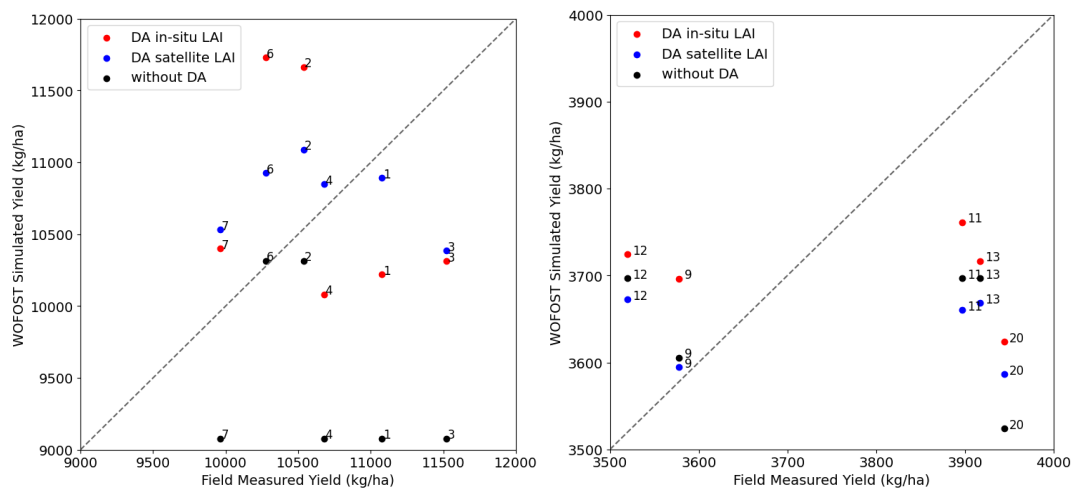

Figure 2: The performance of different data assimilation strategies for WOFOST simulation of maize (left) and soybean (right).

It is shown in Table 1 that in both cases DA improved the performance of the WOFOST model. DA adjusted the state variable such as LAI to reflect the situation in the field. This can be seen in Figure 3 where in both cases the first assimilated measurement of LAI decreased unrealistically high values in the early stages of the growing season. Thus, in this case the adjustment of LAI partially corrected the wrong emergence date estimation from the model. Nonetheless the impact of a wrong emergence date extends beyond the LAI curve, and it involves calculation of thermal time. Thus, having a correct estimate of emergence day remains of major importance when modeling crop growth.

Our study demonstrates one successful application and validation of the WOFOST crop growth model with DA for two important crops in Serbia: maize and soybean. Experimental work for the

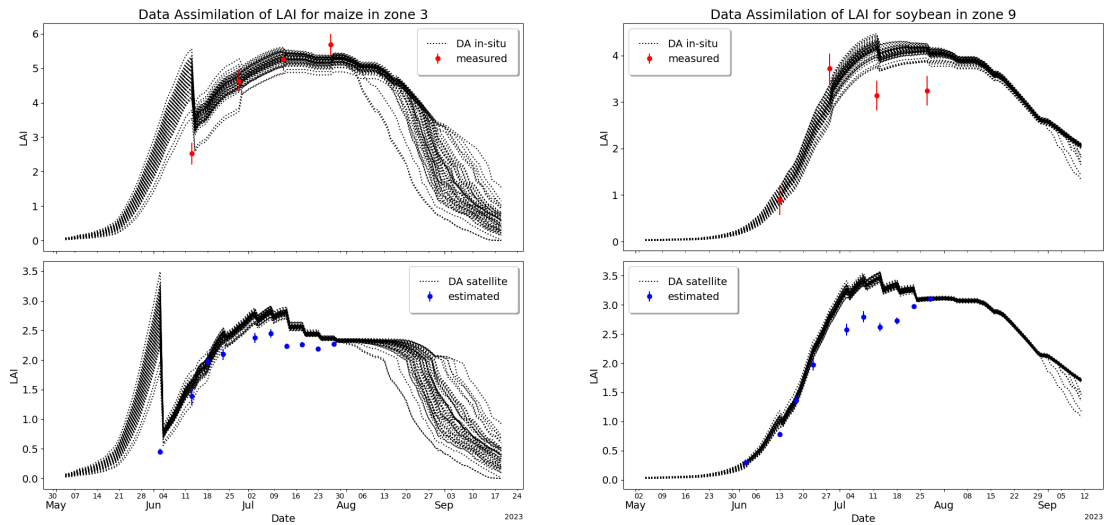soybean is generally important as there is limited testing and validation of the WOFOST model for this crop [1].

Figure 3: Data assimilation of in-situ LAI (upper) and satellite LAI (lower) for maize (left) and for soybean (right).

Acknowledgements

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Knibbe, W.J., Afman, L., Boersma, S., Bogaardt, M.-J., Evers, J., Evert, F., van der Heide, J. et al. (2022) Digital Twins in the Green Life Sciences. NJAS: Impact in Agricultural and Life Sciences 94, no. 1, 249–79. doi:10.1080/27685241.2022.2150571.

[2] Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., Xie, X. (2019) Assimilation of remote sensing into crop growth models: Current status and perspectives, Agricultural and Forest Meteorology, 276–277, 107609, https://doi.org/10.1016/j.agrformet.2019.06.008.

[3] Jindo, K., Kozan, O., de Wit, A. (2023) Data Assimilation of Remote Sensing Data into a Crop Growth Model. In: Cammarano, D., van Evert, F.K., Kempenaar, C. (eds) Precision Agriculture: Modelling. Progress in Precision Agriculture. Springer, Cham. https://doi.org/10.1007/978-3-031-15258-0_8

[4] Gaso, D.V., Paudel, D., de Wit, A., Puntel, L.A., Mullissa, A., Kooistra, L. (2024) Beyond assimilation of leaf area index: Leveraging additional spectral information using machine learning for site-specific soybean yield prediction, Agricultural and Forest Meteorology, 351, 110022, https://doi.org/10.1016/j.agrformet.2024.110022.

[5] Gaso, D.V., de Wit, A. Berger, A.G., Kooistra, L. (2021) Predicting within-field soybean yield variability by coupling Sentinel-2 leaf area index with a crop growth model, Agricultural and Forest Meteorology, 308–309, 108553, https://doi.org/10.1016/j.agrformet.2021.108553.

[6] Maestrini, B., Vaessen, H. M., van Hoeven, A. J. A., Sijbrandij, F. D., van Oort, P. A. J., Kempenaar, C., & van Evert, F. K. (2023) Using assimilation of sensor data in crop growth models. https://edepot.wur.nl/638921.

[7] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023) ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 (Accessed on 29-4-2024).

[8] de Wit, A., Boogaard, H. 2021. A gentle introduction to WOFOST. Wageningen Environmental Research.