

Assessing Different Remotely Sensed Wheat Biophysical Variables for Data Assimilation in a Crop Model for Yield Mapping Prediction

Riccardo Fazioli

`riccardo.fazioli@unitus.it`

Tuscia University

Federica Carucci

Tuscia University

Luca Marrone

Tuscia University

David Kovács

TU Wien

Davide Cammarano

Aarhus University

Stefano Lo Presti

Centrale Valutativa

Zhenhai Li

Shandong University of Science and Technology

Raffaele Casa

Tuscia University

Research Article

Keywords: Data assimilation, DSSAT-CERES-Wheat, Sentinel-2, Leaf Area Index, Aboveground biomass, Canopy Nitrogen content

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9682107/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

ASSESSING DIFFERENT REMOTELY SENSED WHEAT BIOPHYSICAL VARIABLES FOR DATA ASSIMILATION IN A CROP MODEL FOR YIELD MAPPING PREDICTION

Authors: Riccardo Fazioli ⁽¹⁾, Federica Carucci ⁽¹⁾, Luca Marrone ⁽¹⁾, David Kovács ⁽²⁾, Davide Cammarano ⁽³⁾, Stefano Lo Presti ⁽⁴⁾, Zhenhai Li ⁽⁵⁾, Raffaele Casa ⁽¹⁾

¹*Department of Agricultural and Forestry scieNcEs (DAFNE), University of Tuscia, Via San Camillo de Lellis, 01100 Viterbo, Italy*

²*Department of Geodesy and Geoinformation, TU Wien, Wiedner Hauptstrasse 8, Vienna, 1040, Austria*

³*Department of Agroecology, iClimate, CBIO, Aarhus University, Tjele, Denmark*

⁴*Centrale Valutativa, 00131 Roma, Italy*

⁵*College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, Shandong, China*

ABSTRACT

Purpose Process-based crop models are increasingly combined with satellite data to support agronomic management, but the implications of choosing different remotely sensed biophysical variables for data assimilation are still not fully understood. This study aimed to identify the most effective variables among leaf area index (LAI), aboveground biomass, and canopy nitrogen content for assimilation into the DSSAT CERES-Wheat model, by comparing single and multivariable schemes and testing their application at pixel scale for yield mapping.

Methods An optimization-based data assimilation framework was first tested in a synthetic experiment with different soils, growing seasons, and nitrogen fertilization levels. The same framework was then evaluated in winter wheat fields in Central Italy using Sentinel-2 derived LAI, aboveground biomass and canopy nitrogen content, first across 10 experimental sampling units and then at plot scale with 10 x 10 m resolution. Model performance was assessed using standard accuracy and agreement metrics and by analysing the spatial correspondence between simulated and observed yield.

Results Under synthetic conditions, assimilation reduced errors in simulated yield and grain nitrogen compared with the open-loop model. With field Sentinel-2 data, schemes including above-ground biomass improved yield prediction, whereas LAI-only and canopy nitrogen content-only schemes were less effective or even degraded performance. At pixel scale, assimilation produced realistic 10 × 10 m yield maps, reducing RMSE from 1415.40 to 576.53 kg DM ha⁻¹ and increasing codispersion with the observed yield map from -0.11 to 0.57.

35 **Conclusion** These results highlight above-ground biomass as the most robust remotely sensed
36 variable for constraining the model and underline the key role of satellite product quality in
37 multivariable data assimilation. Assimilating multiple satellite-derived biophysical variables into a
38 process-based crop model can improve wheat yield prediction and yield mapping at field scale,
39 providing practical guidance for crop-model–satellite integration in precision wheat management.

40 **Keywords:**

41 Data assimilation; DSSAT-CERES-Wheat; Sentinel-2; Leaf Area Index; Aboveground biomass;
42 Canopy Nitrogen content

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

1.INTRODUCTION

Wheat (*Triticum aestivum* L.) is the most widely cultivated cereal crop worldwide and represents a fundamental component of global food security due to its high contribution to human caloric and protein intake (Shiferaw et al. 2013). In this context, predicting accurate yield and grain quality before key agronomic decisions is essential for supporting the agronomic management decisions (e.g. in-season nitrogen fertilization). However, their estimation at field scale remains challenging due to the complexity of interaction among crop, environment and management, and the limited availability of spatially detailed observation (Basso and Liu 2019; Hao et al. 2021). Different methods have been developed for crop yield and quality estimation, differing in data requirements and model complexity. For example, empirical approaches, such as regression models, offer simplicity but often fail to capture yield variability under changing weather conditions. Machine learning (ML) techniques, ranging from random forests to deep learning, have shown high accuracy in data-rich settings by modeling nonlinear interactions among variables (Kamir et al. 2020; van Klompenburg et al. 2020). However, because they are largely data-driven, their predictive skill may decline when applied to conditions that are poorly represented in the training data, such as rare or extreme weather events, failing to estimate crop responses correctly under those circumstances (Morales and Villalobos 2023). In contrast, functional process-based models like DSSAT (Decision Support System for Agrothechnology), APSIM (Agricultural Production System sIMulator) or STICS (Simulateur mulTIdisciplinaire pour les Cultures Standard), simulate growth, development and yield using biophysical equations that simulate interactions between soil, plants, weather, management, quantifying water, nitrogen, and carbon balances as well as growth and yield stresses (Jones et al. 2017; Rosenzweig et al. 2014; Van Ittersum et al. 2013). These have been extensively applied to climate change impact assessments, gene-based modelling, agronomic trials, and precision agriculture (Basso et al. 2000; Cammarano et al. 2021; Messina et al. 2006). Their ability to predict yield under different “what-if” scenarios has expanded applications to areas such as precision farming (Fabbri et al. 2023) and risk assessment in the insurance sector (Weiss et al., 2020). However, uncertainties in input data and their high spatial variability still limit model accuracy (Asseng and Van Herwaarden 2003). Moreover, deterministic crop growth models are typically point-based (or field-based if inputs are considered uniform) and fail to take into account agricultural fields spatial variability unless particular and data demanding spatialization procedures are carried out (Pasquel et al. 2022). To overcome these issues, data assimilation has been proposed to improve consistency between simulated and observed data by integrating external measurements into crop models (Delrcolle et al. 1992; Maas 1988). Remote sensing data have proven useful for updating state variables or adjusting model parameters during the crop cycle. Assimilation methods and application

with crop growth models have been reviewed by (Dlamini et al. 2023; Huang et al. 2024; Jin et al. 2018; Luo et al. 2023) and can be classified as sequential and optimization techniques. Sequential techniques such as Kalman filters and its variants, e.g. Ensemble Kalman Filter or Particle Filter (Khodarahmi and Maihami 2023; Welch and Bishop 1997), involve updating the model during the simulation by using observation collected during the season and improving final predictions (Silvestro et al. 2021; Wang et al. 2024). Numerical optimization based approaches aim instead on minimizing the differences with the observed variable over a defined time window. One of the methods widely used for this approach is the Particle Swarm Optimization (PSO), which explores the parameter space using a population of candidate solutions (“particles”), each representing a specific set of model parameters. At each iteration positions are updated according to the best position found that minimize the difference with the observed variable, so the swarm progressively moves towards that position, trying to find better values (Wagner et al. 2020; Wang et al. 2018). Another commonly used scheme is the Nelder-Mead simplex algorithm, a local optimization method “derivate-free” which works on a small group of candidate solutions. These solutions form the vertices of a simplex in the parameter space, which is iteratively transformed until it converges to a local minimum (Nelder and Mead 1965). Both categories have shown potential to improve the consistency between simulation and observation, even when integrating satellite-derived information (Luo et al. 2023). Satellite observations are especially relevant for precision agriculture, as they provide spatially explicit and temporally frequent information on within-field variability. State variables such as the Leaf Area Index (LAI) have been often successfully assimilated with high accuracy (Richetti et al. 2019). Previous studies have focused on a single variable, typically LAI, for assimilation (Dlamini et al. 2023). However, combining multiple biophysical variables in assimilation processes, can capture complementary aspects of crop growth and physiology, enhancing model prediction performance (Li et al. 2015; Yu et al. 2022). Aboveground biomass, for example, integrates cumulative growth and is closely linked to yield, but its use in assimilation has been limited, mainly due to the difficulty of retrieving it accurately from remote sensing (Luo et al. 2023). Including biomass, along with LAI and canopy nitrogen content, could improve predictions by constraining multiple dimensions of crop development. Sentinel-2 imagery offers an optimal combination of frequency and resolution suitable for the retrieval of biophysical variables such as LAI. Hyperspectral sensors like PRISMA and ENMAP (Shaik et al. 2023), and the forthcoming CHIME mission (Nieke et al. 2023), enable improved remote sensing of variables like aboveground biomass and canopy nitrogen content due to their high spectral resolution. For example, (Wocher et al. 2022) , developed a hybrid retrieval workflow using EnMAP-like spectra to estimate crop canopy carbon content and aboveground biomass over cultivated areas, demonstrating the feasibility of biomass mapping from operational

imaging spectroscopy. (Verrelst et al. 2021) showed that PRISMA data can be used to map canopy nitrogen content at landscape scale with hybrid radiative transfer-machine learning approach and highlighted its relevance in the context of future missions such as CHIME. The hypothesis of this study is that by assimilating Leaf Area Index (LAI), aboveground biomass, and canopy nitrogen content into an existing wheat model the accuracy of the simulated yield will increase. The objective is to assess which biophysical variables, retrievable from satellite remote sensing, are most suitable for yield and grain nitrogen content estimation and whether assimilating multiple biophysical variables improves model predictions as compared to the assimilation of a single variable. Additionally, the objectives were also to assess the potential of this approach to generate a high-resolution yield map prediction at the field scale using satellite-derived biophysical variables.

2.MATERIALS AND METHODS

2.1 Experiment workflow

The proposed framework was tested under both under synthetic and real-world conditions, in the latter case using biophysical variables retrieved from remote sensing Sentinel-2 images using hybrid approaches based on physically based canopy reflectance models and machine learning algorithms (De Clerck et al. 2024; Verrelst et al. 2019), as the primary data source. The study was organized into three sequential phases. In the first phase, an *in-silico* experiment with synthetic data was carried out to assess the proposed data assimilation framework in ideal (controlled) conditions. The second phase involved evaluation of the approach using field data collected during the 2024 wheat season across multiple plots in the Rieti plain in Central Italy, by using Sentinel-2 derived observations for the assimilation. Finally, the third phase focused on generating a yield prediction map for a selected field by applying the assimilation framework at the pixel scale. The overall workflow of the study, including the synthetic experiment, the field validation and the yield mapping step, is summarized in Figure 1.

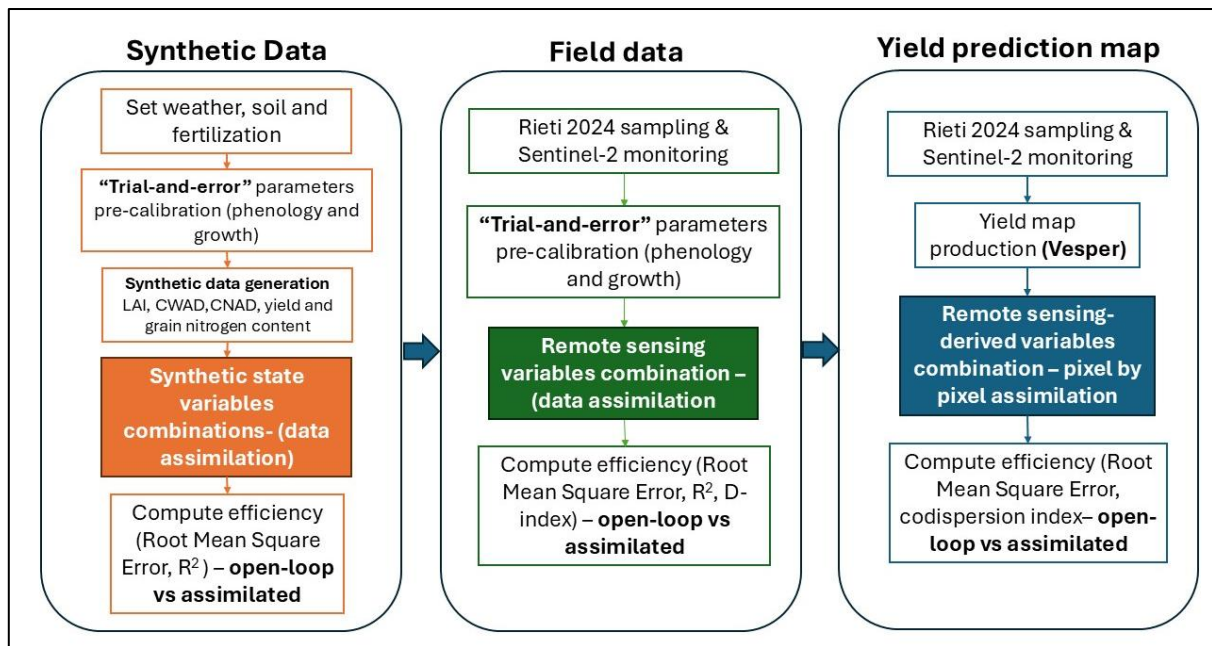


Figure 1 The framework consists of three main steps: (i) a synthetic data experiment to test the assimilation setup, (ii) validation with ground data and Sentinel-2 observations in the Rieti plain (2024 season), and (iii) pixel-scale yield mapping. In each step, pre-calibration was performed, followed by data assimilation (open-loop vs assimilated simulations) and evaluation of model efficiency.

2.2 Study Area

The study area is located in the Rieti Plain, Central Italy (lat. 42°25' N, long. 12°52' E, alt. 400 m a.s.l.), and covers approximately 90 km². It is a predominantly agricultural zone situated on an alluvial plateau, with soils primarily derived from alluvial deposits. The soils are classified as Eutric Fluvic Cambisols, Calcaric Cambisols and Regosols, according to the World Reference Base for Soil Resources (WRB). Sampling was conducted during the 2024 wheat growing seasons across three different fields which were all cultivated with winter wheat Figure 2. These included fields coded as: 1) LIDA15 of about 14 ha, characterized by loam to clay loam soils with about 32% clay; 2) TA7 of about 5.7 ha, presenting lighter texture, sandy loam to loam, with ~26% clay; 3) GITO6, of 26 ha, with more clay-rich soils, about 39.6% clay. Soil data were collected in a previous study, and the data used were obtained from Zanini et al. (2025), who performed point sampling at various sites across the plain and generated spatial soil maps through geostatistical techniques. Historical meteorological data (2011–2021) were obtained from the ARSIAL Rieti weather station (alt. 379 m a.s.l., lat. 42°25'22" N, long. 12°48'46" E), while for the 2023–2024 seasons, data were obtained from the GIAS

173 platform from Agronica S.R.L., through the Hypermetero (<https://www.hypermeteo.com>) station
174 (code IT0014244N1282E), which provided daily interpolated data from nearby stations.



175
176 *Figure 2. Study area in the Rieti Plain (Central Italy) showing the location and boundaries of the*
177 *three experimental plots: A – (LIDA15, ~14 ha), B – (TA7, ~5.7 ha) and C – (GITO6 ~26 ha) (red*
178 *outlines) and the Experimental Sampling Units (ESUs) where ground measurements were collected*
179 *during the 2024 growing seasons.*

180 **2.3 DSSAT model**

181 The crop model used in this study was the DSSAT v4.8 (Decision Support System for Agrotechnology
182 Transfer; (Hoogenboom et al. 2019) and within the suite of available wheat model, the CERES-Wheat
183 model was chosen to simulate crop growth, development and yield. The model is composed by
184 different and interconnected modules that simulate daily interactions between soil, plants, climate and
185 agronomic management, including soil water balance, crop water stress and nitrogen uptake. Soil
186 water balance is simulated by partitioning water into evaporation, rooting uptake, and drainage using
187 a tipping-bucket method. Crop water stress is calculated based on the ratio between potential
188 transpiration and plant-available water (Jones et al. 2003), while nitrogen uptake is simulated by
189 matching crop demand and soil supply following the framework described by Godwin and Singh
190 (1998). For each module, DSSAT requires daily weather data (minimum and maximum air
191 temperature, incoming solar radiation, rainfall), soil properties (texture, organic matter, mineral
192 nitrogen), and agronomic practices as inputs. These inputs determine the environmental and

management conditions under which crop growth and soil water and nitrogen dynamics are simulated throughout the season. In addition, an important set of inputs is the soil initial conditions (mineral nitrogen and available water) that need to be specified when the simulation starts (which might differ from the planting date). In this study, the initial conditions were determined by analysing historical weather data derived from the “Carlo Jucci” Apennine Centre (Rieti) managed by University of Perugia, and by conducting soil analyses during the 2023–2024 season, as reported by Zanini et al. (2025). Since the initial soil water and nitrogen status at the start of the experiment could not be directly reconstructed with sufficient detail, the model was initialized at least one year before sowing. This spin-up phase allowed DSSAT to simulate the temporal evolution of soil water and nitrogen under site-specific weather and soil conditions, thereby reducing dependence on arbitrary initialization values and providing more consistent initial states for the assimilation experiment. In the present study, cultivar-specific genetic coefficients were calibrated to reproduce observed phenology, biomass accumulation, and yield. These parameters govern key crop processes such as development rate, leaf appearance, and grain formation within the DSSAT CERES-Wheat framework.

2.4 Trial-and-error pre calibration process

Before the assimilation step, a calibration was performed to align key phenological and growth parameters of the DSSAT-CERES-Wheat model with the observed crop development under the specific environmental and management conditions of the study site. The calibration phase aimed to provide a realistic starting point for parameterization of the cultivar files ensuring that model simulation matched field or synthetic reference data before assimilation. This step is crucial in crop modelling workflows as it reduces the search space during the optimization and minimizes the risk of unrealistic parameter combinations (Seidel et al. 2018; Wallach et al. 2024). The parameters that were calibrated were the ones controlling phenology, leaf growth dynamics, photosynthetic efficiency, and biomass partitioning traits, not all of which were later included in the assimilation phase (Table 9). The synthetic-data and real-data experiments were conducted separately, as they used different reference information and had distinct objectives. The real-data experiment was adjusted to reproduce the average phenological and growth behaviour observed across the Experimental Sample Units (ESUs) in 2024 (see section 2.7 for details), whereas the synthetic-data experiment was tuned to generate plausible reference trajectories under the assumptions adopted for the multi-season synthetic setup. Consequently, different pre-calibrated cultivar parameter sets were retained for the two experiments. For real-data experiment, since measurements were collected across multiple ESUs, the aim of the calibration was to adapt model simulations to the average growth pattern across locations rather than overfitting to a single site. The pre-calibration relied on both ground-based

227 measurements (phenology, final yield) and remote sensing observations (LAI, aboveground biomass
228 from Sentinel-2). Parameters were adjusted manually through a trial-and-error approach, aiming at
229 reducing errors for simulated phenology and then growth. Specifically for phenology, the parameters
230 controlling sensitivity to vernalization and photoperiod (P1V and P1D in DSSAT) were adjusted to
231 minimized observed and simulated anthesis, while maturity the thermal sum from anthesis to maturity
232 (P5) was adjusted. For growth, PHINT was first adjusted to improve the timing and magnitude of the
233 simulated LAI curve. Then, ecotype coefficients controlling the duration of leaf growth (particularly
234 P1, P2 and P3) were modified to better reproduce the shape of the observed canopy development.
235 Finally, parameters related to radiation-use efficiency and grain formation (PARUE, PARU2, G1 and
236 G2) were refined to improve the agreement with observed biomass accumulation in final yield, while
237 retaining a parametrization able to reproduce the average growth trajectory observed across multiple
238 ESUs without overfitting any of them. This adjustment involved a broader set of coefficients than
239 those subsequently optimized during assimilation. Table 9 reports the subset of parameters retained
240 for the assimilation step.

241 For the synthetic-data experiment, calibration was guided primarily by Sentinel-2 LAI curves
242 extracted from ESUs located in winter wheat fields in Rieti plain over multiple seasons. ESUs were
243 selected in each year to correspond to parcels where winter wheat was present, as indicated by field
244 information and by the typical temporal pattern of LAI. Phenological stages in earlier seasons were
245 interfered from field observations collected in 2023 – 2024 and applied consistently across years.
246 Because detailed fertilization records were not available for all seasons, an assumed nitrogen
247 fertilization level of 210 kg N ha⁻¹ was used during the trial-and-error calibration of the synthetic-
248 data experiment to obtain LAI trajectories consistent with the multi-year Sentinel-2 observations.
249 Under these constraints, the trial-and-error pre-calibration focused mainly on phenology and LAI
250 dynamics, applying a compromise-based adjustment to avoid overfitting a single time series and to
251 reproduce the typical multi-year LAI behaviour of wheat in the study area.

252

253

254

255

256

257

258 *Table 9 Default values refer to DSSAT-CERES-Wheat standard settings for the reference cultivar;*
 259 *te.c. (trial-and-error calibration) values were obtained through the trial and error calibration process*
 260 *described in Section 2.4, separately for the real-data and synthetic-data experiments. The te.c. values*
 261 *are those applied during the assimilation process.*

Parameter	Description	Default Values	te.c. values (Real Data)	te.c. values (Synthetic data)
P1D	Photoperiod response (% reduction in rate/10 h drop in photoperiod)	75	150	50
P1V	Days, optimum vernalizing temperature, required for vernalization	5	60	15
P5	Grain filling phase duration (°C.d)	450	600	230
PHINT	Interval between successive leaf tip appearances (°C.d)	70	45	45
G1	Kernel number per unit canopy weight at anthesis (#/g)	30	30	30
G2	Standard kernel size under optimum conditions (mg)	35	40	35
PARUE	PAR conversion to dm ratio, after last leaf (g/MJ)	1.0	1.5	4.1
PARU2	PAR conversion to dm ratio, before last leaf stage (g/MJ)	6.0	2.5	1.8

262

263 **2.5 Assimilation setup**

264 Synthetic data are defined here as model-generated observations obtained under well-defined
 265 simulation scenarios. These data were used in this context to test the assimilation system and to assess
 266 the model’s sensitivity and effectiveness by following the procedure used by (Curnel et al. (2011)).

267 To generate synthetic data, several simulation scenarios were setup (Figure 20), with eleven nitrogen
 268 (N) fertilization rates ranging from 0 to 300 kg N ha⁻¹, with 30 kg N ha⁻¹ increments. In addition,
 269 simulations were performed across three growing seasons, representing different rainfall conditions
 270 based on cumulative precipitation during the wheat growth cycle: 2011–2012 (dry, 282 mm), 2014–
 271 2015 (average, 403 mm), and 2022–2023 (wet, 904 mm). Across all scenarios, the growth season was
 272 defined as winter wheat sown on January 5th to maturity which corresponding to the maturity date
 273 simulated by DSSAT. Additionally, three distinct soil types with a maximum depth of 2 m were used

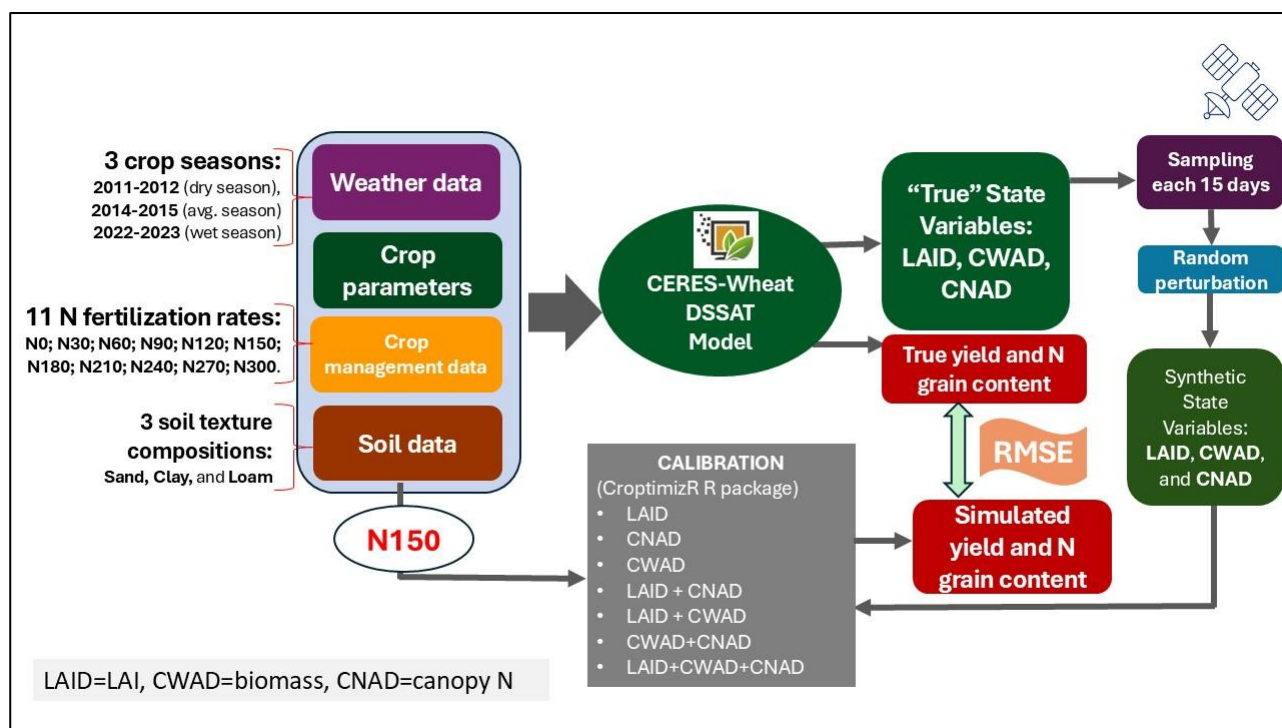
for the parametrization of the model: clay soil (60% clay, 30% silt), loam (32% clay, 25% silt), and sandy soil (30% clay, 10% silt).

A total of 99 factorial simulations were performed, covering all combinations of factors, i.e. nitrogen fertilization, meteorological and soil conditions (11x3x3). For each simulation, final yield and grain nitrogen content were retained and treated as the “synthetic reference” values. Each of these simulated datasets was accompanied by the corresponding simulated time series of LAI, aboveground biomass, and canopy nitrogen content (LAID, CWAD, CNAD in DSSAT notation), which were considered as synthetic state variables. From these daily simulations, data were extracted every 15 days (to match a realistic satellite observation frequency) and perturbed with random noise to simulate observation uncertainty. The perturbation followed the formula:

$$\text{StateVariable}_{\text{syn},i} = \text{StateVariable}_{\text{true},i} (1 + \eta_i) \text{ with } \eta_i \sim N(0; 0.3^2)$$

Where “ $\text{StateVariable}_{\text{syn},i}$ ” is the synthetic state variable at time i , $\text{StateVariable}_{\text{true},i}$ is the corresponding “true” state obtained from the previous simulation at time i and η_i is the relative random error term added at time i . The random error followed a normal distribution with zero mean and standard deviation equal to 0.3 corresponding to a coefficient of variation (CV) of 30%. This value was adopted as a pragmatic and standardized uncertainty scenario for the synthetic-data experiment, in order to compare the different assimilation setups under controlled conditions rather than to reproduce the exact empirical uncertainty of each variable.

Assimilation was performed using a calibration-based approach, always starting from the phenology, growth, and yield parameters. The scenario with 150 kg N ha⁻¹ was used as the baseline for calibration, as it represents an intermediate fertilization level and was considered agronomically reasonable to use this methodology. The assimilation was tested using different sets of variables: LAID only (Leaf Area Index), CNAD only (Canopy Nitrogen Content), CWAD (Aboveground biomass) only, LAID + CWAD, LAID + CNAD, CWAD + CNAD, and LAID + CWAD + CNAD. An additional “open-loop” simulation (without assimilation) was also performed using the same 150 kg N ha⁻¹ scenario, in order to serve as a baseline for comparison and to isolate the contribution of the assimilation process.



301

302 *Figure 20 Workflow of the synthetic data experiment. The DSSAT CERES – Wheat model was run*
 303 *across factorial combinations of seasons, soils, and nitrogen rates to generate “true” yield and state*
 304 *variables (LAID, CWAD, CNAD). Synthetic observations were created by adding random noise to*
 305 *these variables, and assimilation was tested on a baseline scenario using different variable*
 306 *combinations to evaluate the model performance*

307 2.7 Evaluation with ground data

308 A validation using real data was conducted using measurements from three different plots across the
 309 Rieti plain during the 2024 wheat season. The plots were subject to heterogeneous management
 310 practices, reflecting the specific choices and strategies adopted by each farmer. Three fertilizations
 311 events were performed using Variable Rate Application (VRA), adjusting the standard dose by $\pm 25\%$
 312 according to agronomic needs as shown in Table 2. Management zones were defined using the
 313 algorithm proposed by Holland and Schepers (2010), which is based on the calculation of a
 314 Sufficiency Index (SI) derived from Sentinel-2 imagery. The SI was computed using the CAB index,
 315 a proxy of canopy chlorophyll content, to assess spatial variability in nitrogen status and guide site-
 316 specific fertilization. summarizes the specific agronomic management practices adopted in each field.

317

318

319 *Table 2. Agronomic management practices in the three experimental fields during the 2024 wheat*
 320 *season, including tillage, sowing, fertilization (with VRA where applied), and harvest date. Dates are*
 321 *expressed as month/day/year (mm/dd/yyyy).*

Field	Tillage	Sowing (mm/dd/yyyy)	Fertilization dates	Fertilizer details (dose)	Harvest
TA 7	Subsoiling	12/24/2023 –	02/08/2024	Urea ammonium phosphate	07/22/2024
	0.45 m, cultivator 0.3 m	280 kg ha ⁻¹	03/20/2024	200 kg ha ⁻¹ – 38% N Urea 200 kg ha ⁻¹ – 46% N	
LIDA 15	Subsoiling	12/16/2023 –	02/25/2024	Ammonium phosphate	07/18/2024
	0.4 m, rotary harrow	280 kg ha ⁻¹	03/21/2024	150 kg ha ⁻¹ –27% N Calcium nitrate 150 kg ha ⁻¹ – 27% N - VRA	
	0.25 m		04/10/2024	Urea 200 kg ha ⁻¹ – 46% N - VRA	
GITO 6	Subsoiling	01/05/2024 –	02/22/2024	Urea ammonium nitrate	07/18/2024
	0.4 m, rotary harrow 0.1 m	300 kg ha ⁻¹	04/06/2024	300 kg ha ⁻¹ – 33% N - VRA Urea 200 kg ha ⁻¹ – 46% N	

322

323 Across the crop cycle, measurements of LAI, biomass, and canopy nitrogen content were carried out

324 through field sampling in 10 Experimental Sample Units (ESUs) of 30 × 30 m distributed across the

325 experimental plots (Figure 2). For LAI, in each ESU three representative sub-areas were selected

326 along the diagonals to capture within-plot variability. In each sub-area, LAI was measured non-

327 destructively using an LAI-3100 area meter (LI-COR Biosciences, USA) on a composite sample of

328 fully expanded green leaves collected from the canopy top. Aboveground biomass and yield

329 components were determined destructively by harvesting four 0.25 × 0.25 m quadrats placed along

330 the ESU diagonals, avoiding plot borders. Fresh biomass was weighed in the field, then samples were

331 oven-dried at 65 °C until constant weight to determine dry matter. At maturity, grain yield and its

332 components (spike number, grain number, thousand-grain weight) were measured from the same

333 ESUs. For canopy nitrogen content, subsamples of the oven-dried biomass were finely ground and

334 analysed in the laboratory using a portable near-infrared (NIR) spectrometer.

335 In order to generate a spatially continuous yield map for comparison with model simulations, an
336 additional sampling campaign was conducted in the LIDA 15 plot (Lat. 42°27'9" N, Long. 12°46'48"
337 E). Samples were collected every 30 m across the entire field, resulting in a total of 137 measurement
338 points, each 0.25 m x 0.25 m in size. The data were interpolated using Block Kriging, following the
339 construction of a variogram, and processed with the VESPER software (Whelan et al. 2002), resulting
340 in a 10 × 10 m resolution yield map that is consistent with the Sentinel-2 spatial resolution.

341 **2.8 Remote sensing data**

342 Sentinel-2 imagery (ESA Copernicus program) was used to extract biophysical variables during the
343 growing season and to provide input for the assimilation framework. For each Experimental Sampling
344 Unit (ESU), cloud-free Sentinel-2 scenes were processed to derive time series of Leaf Area Index
345 (LAI), canopy nitrogen content, and aboveground biomass at 10 × 10 m resolution. LAI was
346 calculated from Sentinel-2 images under clear sky conditions using the BV-NET algorithm (Weiss
347 and Baret, 2016), a robust neural-network-based method integrated in the ESA SNAP toolbox,
348 implemented within the EO Browser / Copernicus Algorithm Thematic Platform, specifically trained
349 to convert bottom-of-atmosphere reflectance into LAI values.

350 The canopy nitrogen content was estimated using the PyEOGPR package (De Clerck et al. 2024)
351 available within the openEO Algorithm Plaza of the Copernicus Data Space Ecosystem. The package
352 implements a Gaussian Process Regression (GPR) model trained on synthetic data and optimized
353 with a Euclidean-based Diversity Active Learning (EBD-AL) algorithm. The method showed good
354 agreement with field measurements in validation. For this study, the pre-trained model for canopy
355 nitrogen content was applied to Sentinel-2 surface reflectance data and to a set of vegetation indices
356 sensitive to canopy N. Predicted values represent canopy nitrogen content on a dry biomass basis,
357 obtained by multiplying predicted nitrogen concentration by biomass estimates.

358 Aboveground biomass was estimated using a Monteith-type model (Monteith 1972) implemented on
359 the Tethys platform, which combines Sentinel-2-derived FAPAR and the evaporative fraction from
360 SEBAL to account for water stress (Gobbo et al. 2019). Daily biomass accumulation was computed
361 from absorbed PAR and adjusted for radiation use efficiency under water-limited conditions. More
362 specifically, daily dry matter production was estimated from absorbed photosynthetically active
363 radiation and reduced when water stress, derived from evaporative fraction, limited crop growth. For
364 each ESU, a time series of these variables was extracted from Sentinel-2 images under clear-sky
365 conditions through these different algorithms. The temporal integration of daily biomass
366 accumulation provides a time series of aboveground biomass for each pixel. In addition to the ESUs,

for each pixel in the yield map (10 × 10 m), the whole time series of biophysical variables was extracted, creating a pixel-specific time series used for assimilation.

2.9 Statistical Analysis

Biophysical variables were assimilated into the DSSAT model following the same method used with synthetic data. The accuracy of the assimilation framework was evaluated by comparing model outputs with simulated and observed reference values, both in synthetic and real data. In the synthetic data experiment, the Root Mean Square Error (RMSE) was used to compare the synthetic “true values” of yield and grain nitrogen content and the values after the assimilation. RMSE was calculated by applying the formula:

Equation 1 Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (P_i - O_i)^2}$$

where P_i is the predicted (assimilated) value, O_i is the observed (true) value, and n is the number of observations. RMSE provides a measure of the average deviation between predicted and actual values, expressed in the same units as the variable considered. Lower RMSE values indicate better model accuracy. In addition, a linear regression analysis was conducted to evaluate the agreement between satellite-derived biophysical variables (aboveground biomass, LAI and canopy nitrogen content) and field observations across all ESUs. The coefficient of determination (R^2) and the RMSE were used to quantify the quality of fit.

In real-data analysis, the performance of model in predicting yield and grain nitrogen content was assessed by using the RMSE (Equation 1), the coefficient of determination (R^2), which quantifies the proportion of the variance explained by the model, and the Willmott and Wicks (1980) index of agreement or D-index evaluate how closely model predictions matched observed data by following the formula:

Equation 2 D-Index

$$d = 1 - \left[\sum_{i=1}^n (P_i - O_i)^2 / \sum_{i=1}^n (|P'_i| + |O'_i|)^2 \right]$$

Where $O'_i = O_i - \bar{O}$ and $P'_i = P_i - \bar{O}$. The D-index can range from 0 (no agreement) to 1 (perfect agreement).

In both synthetic and real-data experiments, it's been calculated the Assimilation Efficiency (AE) based on RMSE to quantify the relative improvement of given assimilation scenario compared to a baseline simulation (simulation without assimilation process) following a concept similar to the improvement measure used in (Curnel et al. 2011). The value of AE was calculated as follows:

Equation 3 Assimilation Efficiency (AE) - RMSE

$$AE_{RMSE}(\%) = 100 \times \left(1 - \frac{RMSE_{assim}}{RMSE_b}\right)$$

Where $RMSE_{assim}$ is the RMSE of the assimilation scenario and $RMSE_b$ is the RMSE of the reference simulation. Positive AE values indicate improved accuracy, zero indicates no change, and negative values indicate worse performance compared to the reference simulation.

In the real-data experiment, instead of representing changes in R^2 and Willmott's index of agreement (D-index) as percentages relative to the baseline, the overall enhancement was assessed by determining the absolute difference between open-loop and calibrated values.

For spatial comparison between simulated and observed yield map (10 x 10 m resolution, LIDA 15 field) the pixel-wise RMSE is calculated. In addition, the spatial association between the two maps was evaluated using the codispersion coefficient, which measures the degree to which two spatial variables vary together as a function of distance. Positive values indicate similar spatial patterns; negative values indicate opposite spatial patterns. Statistical significance of the codispersion coefficient was assessed using the modified t-test of Dutilleul (1993).

3. RESULTS

3.1 Synthetic data

The assimilation of synthetically generated state variables significantly improved the simulation of both yield and grain nitrogen content compared to the open-loop simulation. In the open-loop setup, RMSE was 2004 kg DM ha⁻¹ for grain yield and 0.681% for grain nitrogen content. Across all variable combinations, RMSE decreased significantly: the corresponding AE ranged from 64.4% to 76.2% for yield and from 59.8% to 65.8% for grain nitrogen content. However, the magnitude of the improvement depended on the specific assimilated variable(s) (Table 3). Among single-variable setups, LAID showed the best individual performance for yield (RMSE = 481.3 kg DM ha⁻¹), followed by CWAD (RMSE = 501 kg DM ha⁻¹). Instead CNAD alone was less accurate (RMSE = 714 kg DM ha⁻¹). For grain nitrogen content, the best performance was obtained when assimilating LAID + CNAD (RMSE = 0.233%), while CNAD alone yielded the highest error (RMSE = 0.274%).

Overall, multivariable configurations that include biomass (CWAD) tended to minimize yield errors (with LAID + CWAD providing the lowest yield RMSE, 477.28 kg DM ha⁻¹), whereas coupling structural and nitrogen-related information (LAID + CNAD) benefited grain N predictions. These outcomes are consistent with the scatterplots in the summary reported in Table 3.

Table 3. Results of the tests of DSSAT assimilation with synthetic data, using either a single biophysical variable or combinations of two or three biophysical variables. The table reports absolute RMSE values, as well as the assimilation efficiency (AE), i.e the improvement compared to the open-loop simulation in which no assimilation was performed.

Variable used in the assimilation	Grain yield (kg DM ha ⁻¹)		Grain nitrogen content (%)	
	RMSE	AE (%)	RMSE	AE (%)
None (Open-Loop)	2004	-	0.68	-
LAI	481.3	76	0.24	64.3
Canopy N	713.9	64.4	0.27	59.8
Biomass	501.5	47.9	0.26	62.3
LAI + Canopy N	513.6	74.4	0.23	65.8
LAI + Biomass	477.3	76.2	0.25	63.3
Biomass + Canopy N	527.3	73.7	0.25	63.4
LAI + Biomass + Canopy N	497.8	75.2	0.26	62

3.2 Remote Sensing data validation

Satellite-derived biophysical variables were validated against field observations across all ESUs using linear regression analysis (Figure 4). Biomass estimates showed a very strong agreement with measured values ($R^2 = 0.943$, $RMSE = 96.7$ g DM m⁻²), with points closely distributed around the 1:1 line and minimal dispersion across the full range of values. LAI estimates exhibited a moderate but consistent agreement ($R^2 = 0.855$, $RMSE = 1.12$ m²/m²) with a generally linear trend across observations. For canopy nitrogen content, the agreement was weaker ($R^2 = 0.405$, $RMSE = 30.1$ kg N ha⁻¹) with a visibly wider dispersion and a lower correspondence with the 1:1 line. These results provide a quantitative assessment of the accuracy of the remote sensing variables and support their integration into the data assimilation framework adopted in this study.

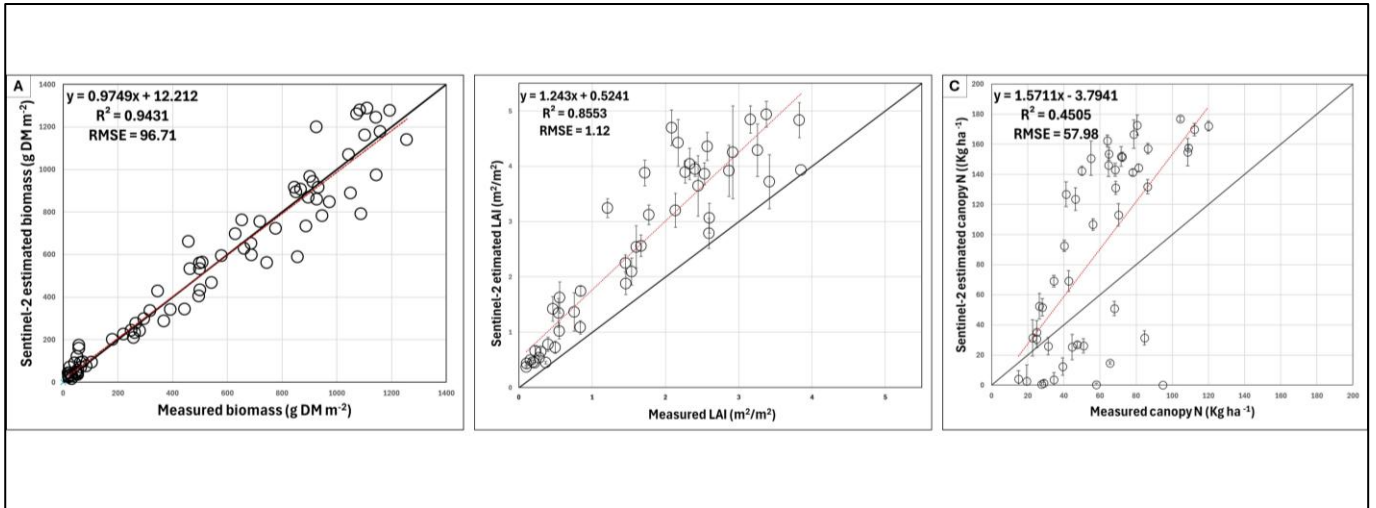


Figure 4 Validation of Sentinel-2-derived biophysical variables against field observations: biomass (A), Leaf Area Index - LAI (B), and canopy nitrogen content – CNC (C). Solid line: 1:1 line; dotted line: regression line; R^2 and Root Mean Square Error (RMSE) quantify the agreement.

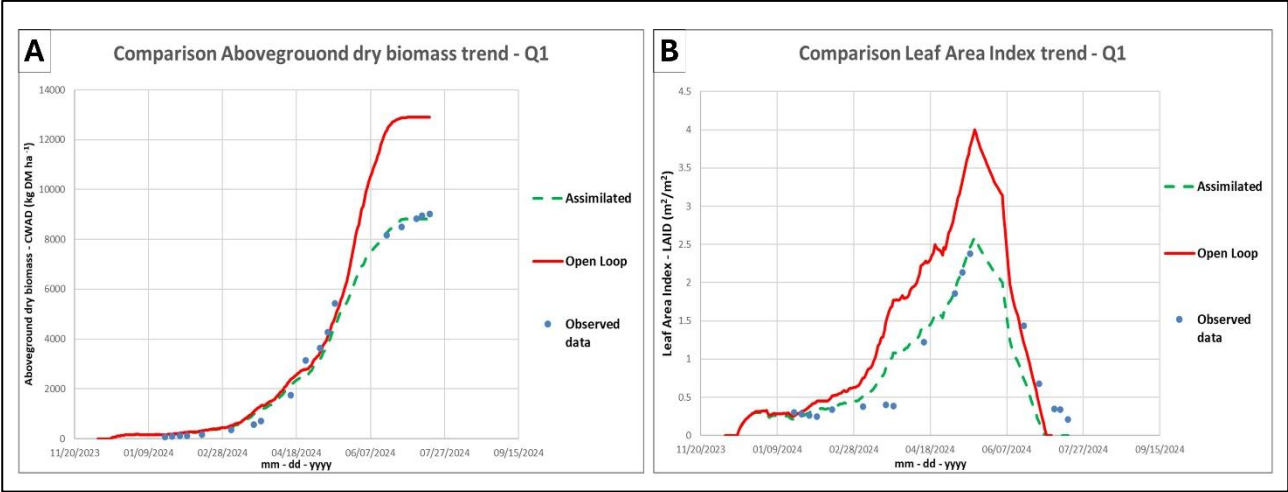
3.3 Field data

Figure 5 shows the comparisons between simulations for Leaf Area Index (LAI) and aboveground biomass (CWAD), using the Q1 Experimental Unit (ESU). This ESU was selected as an illustrative case because it showed one of the clearest responses to assimilation at temporal trajectory level, particularly for biomass under locally observed stress conditions. The purpose of this figure is not to summarize all ESU-level behaviours, which are quantitatively compared in Table 4, but to provide an example of how assimilation affected the seasonal dynamics of the simulated variables. In this case, assimilation was performed using a variable combination of LAI, CWAD, and CNAD. The assimilated CWAD curves demonstrate improved agreement with the observed Sentinel-2-derived data, particularly in the final stage of the season. For LAI, both simulations generally follow the observed data's trend; however, assimilation helped refine the timing and peak of the curve. While this example highlights a visible improvement, it is important to declare that this behaviour was also observed in other ESUs where the same combination of assimilated variables was applied, showing similar trends and improved consistency between simulations and observed data.

Figure 6 shows two scatterplots representing the correlation between the observed yield data and simulated data before (A) and after (B) the assimilation with the same variable combination described in Section 2.5. We can see an overall improvement on the several indices after the assimilation with a reduction of 31.47% on RMSE value, while the R^2 increased from 0.11 to 0.56 and the D-index from 0.30 to 0.69.

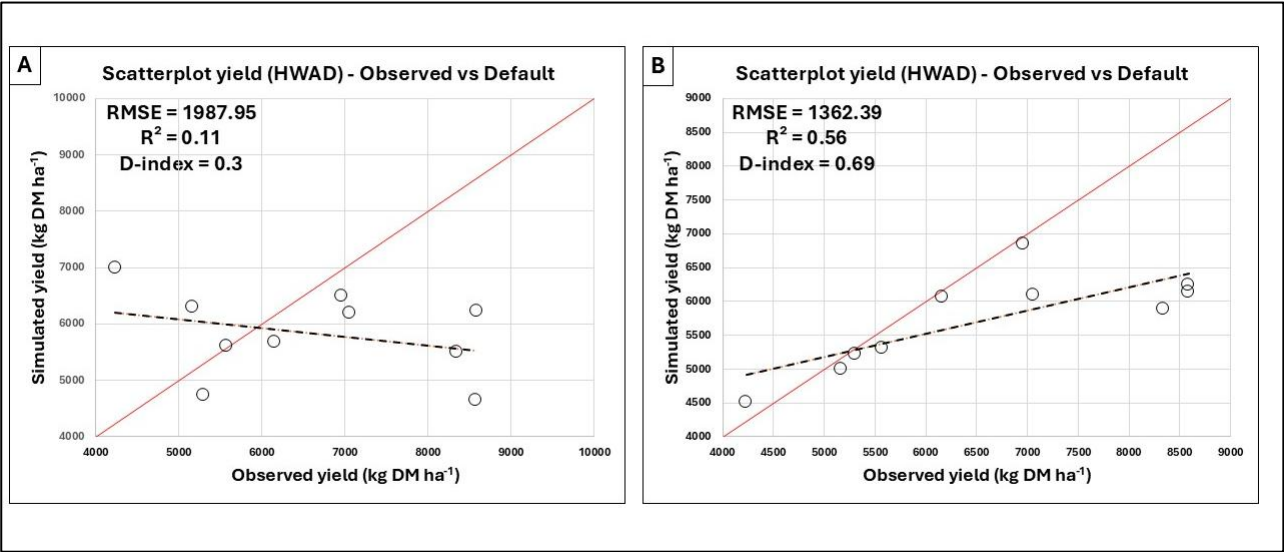
Results for different combinations of assimilated variables were reported in

468 Table 4. CWAD was the variable that minimize the errors in simulated results: RMSE is reduced by
 469 32 %, R^2 improvement (+0.48), and D-index increase (+0.41) compared to the open-loop simulation.
 470 Conversely, isolated assimilation of LAI or CNAD showed counterintuitive performance, worsening
 471 yield prediction with higher RMSE (3164 kg DM ha⁻¹ and 2661 kg DM ha⁻¹ respectively) and lower
 472 R^2 values (-0.04 and -0.06 respectively). The D-index values however showed in general a better
 473 performance with no increasing (+0 for LAI variable alone and + 0.22 for CNAD). Combining
 474 variables for the assimilation improved the yield prediction only when the CWAD variable is used.



476 *Figure 5. Comparison between default DSSAT simulations, assimilated simulations using LAID +*
 477 *CNAD + CWAD variable combination, and Sentinel-2 observations for CWAD (aboveground*
 478 *biomass, A) and LAID (leaf area index, B) during the 2024 growing season in Q1 ESU. Assimilation*
 479 *improved the agreement with observations more clearly for biomass, while also reducing LAI*
 480 *overestimation around the peak.*

481



482

483 *Figure 6. Comparison between simulated and observed yield of 10 ESUs before (A) and after (B) the*
484 *assimilation using LAID + CNAD + CWAD variable combination.*

485 *Table 4 Performance metrics of yield simulations under different assimilation scenarios (real data).*
486 *The table reports RMSE, coefficient of determination (R²), and Willmott's D-index for yield*
487 *predictions, along with their differences (Δ) relative to the open-loop simulation*

Variable Used	RMSE (kg ha-1)	AE RMSE (%)	R ²	ΔR ²	D-Index	ΔD-Index
Open-Loop	1987.95	-	0.11	-	0.30	-
LAID	3164.29	-59.2	0.07	- 0.04	0.30	+ 0
CWAD	1350.78	+32.1	0.59	+ 0.48	0.71	+ 0.41
CNAD	2661.66	-33.9	0.05	- 0.06	0.52	+ 0.22
LAID + CWAD	1337.30	+32.8	0.58	+ 0.47	0.70	+ 0.40
LAID + CNAD	2557.08	-28.6	0.01	- 0.10	0.38	+ 0.08
CWAD + CNAD	1348.95	+32.2	0.57	+ 0.46	0.71	+ 0.41
LAID + CWAD + CNAD	1362.39	+31.5	0.56	+ 0.45	0.69	+ 0.39

488

489

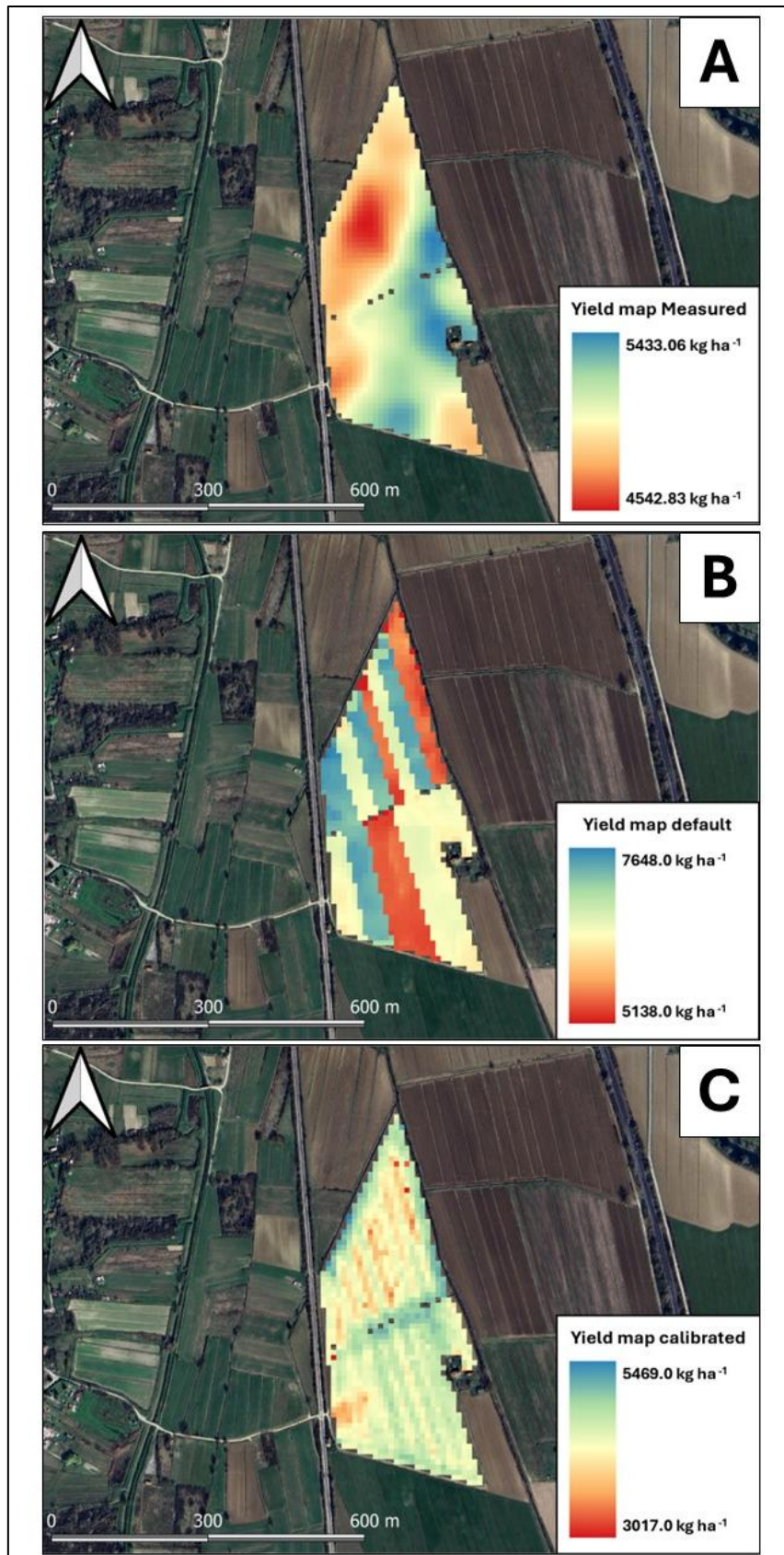
490

491

492 3.4 Yield Map

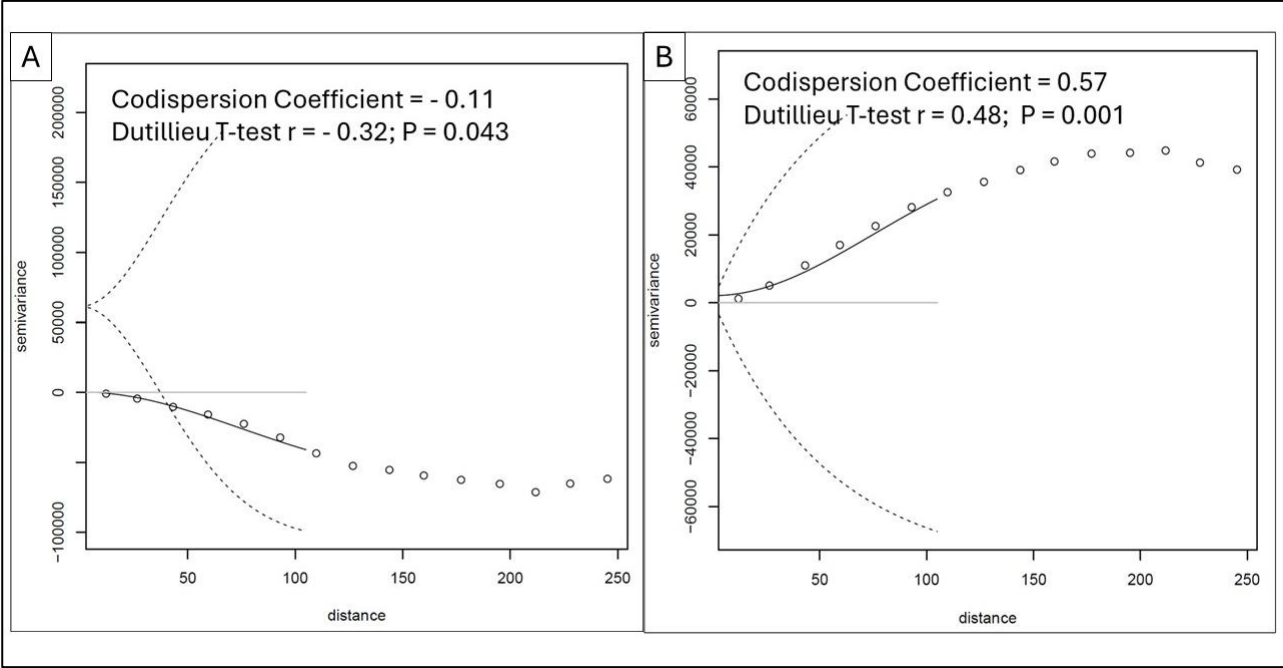
493 Based on the results reported in

494 Table 4, spatial assimilation was performed using LAID, CWAD and CNAD combination, selected
495 as a robust multivariable configuration for pixel-scale application. Figure 7 shows the observed yield
496 map (A), the simulated map without assimilation (B), and the calibrated map after assimilation (C).
497 The open-loop map shows a segmented yield distribution directly reflecting the variable rate
498 fertilization treatments applied in the field, with three distinct nitrogen fertilization levels. However,
499 this pattern does not match the actual distribution, as indicated by the interpolated map from the 137
500 ground truth sampling points. Moreover, absolute yield values in the open-loop simulation are
501 overestimated, often exceeding 7000 kg DM ha⁻¹. In contrast, the assimilated map shows a more
502 continuous spatial pattern and better agreement with the observed field-scale yield. Absolute values
503 are also more in line with field observations, ranging from 4500 to 5400 kg DM ha⁻¹. Supporting the
504 visual evidence, statistical performance indices also show a clear improvement following
505 assimilation: RMSE decreased from 1415 to 576 kg DM ha⁻¹, and the codispersion coefficient (Figure
506 8) increased from -0.11 to 0.57. In particular, the codispersion analysis highlighted clear differences
507 in the spatial agreement between measured and simulated yield maps. In Figure 9A, the codispersion
508 coefficient between the observed yield map and the open-loop map was weakly negative and only
509 marginally significant, indicating poor spatial consistency between the two maps. In contrast, Figure
510 8B showed a strong and highly significant positive codispersion after calibration, demonstrating that
511 the assimilation procedure substantially improved the spatial correspondence between simulated and
512 observed yield patterns.



513

514 *Figure 7 Yield maps for the LIDA15 field (2024 season) comparing measured and simulated yields at*
 515 *10×10 m resolution. (A) Observed yield map interpolated. (B) Simulated yield map without*
 516 *assimilation. (C) Simulated yield map after assimilation of LAID, CWAD, and CNAD.*



518

519 *Figure 8 Codispersion analysis of spatial agreement between simulated and observed yield maps in*
520 *the LIDA15 field. (A) open-loop simulation vs. observed yield. (B) Assimilated simulation vs.*
521 *observed yield.*

522 **DISCUSSION**

523 The objectives of this work were to assess which remotely sensed biophysical variables, in particular
524 LAI, above-ground biomass and canopy nitrogen content, are most suitable for assimilation into the
525 DSSAT-CERES-Wheat model in order to improve its predictive performance for a wheat crop. We
526 examined the advantages of a multivariate approach compared with approaches that use a single
527 variable, testing it with synthetic data and then using field data. Finally, the possibility of reproducing
528 the framework at pixel scale over a specific field to produce a simulated yield map was assessed and
529 evaluated.

530 The experiments with synthetic data confirm the expectations outlined in the introductory section:
531 the integration of external observations into the model considerably improves the simulation of yield
532 and grain nitrogen content compared with the “open-loop” simulation (Dlamini et al. 2023; Huang et
533 al. 2024; Jin et al. 2018). It must be noted, however, that this experiment was conducted under
534 idealised conditions in which the “true state” of the system is known and the observational uncertainty
535 is simulated by adding a controlled error to the assimilated variables. In this way, the assimilation
536 process can fully exploit the information contained in the observations, thus providing very marked
537 improvements, with assimilation efficiency between 48 and 76 % for yield and 60–66 % for grain

538 nitrogen content. The differences between the various configurations of assimilated variables show
539 that LAI and biomass are particularly effective in guiding growth dynamics and yield formation,
540 while information related to canopy nitrogen becomes more relevant when the qualitative component
541 of production is considered, especially when combined with LAI, as indicated by the lowest RMSE
542 for grain nitrogen in the LAI + canopy N scheme. Given the role of these variables in controlling crop
543 growth and nutritional status, the observed behaviour is overall coherent with expectations. In this
544 synthetic experiment, multivariate combinations of structural and nutritional variables achieved
545 performances that were comparable to the best single-variable schemes for yield and slightly
546 improved grain nitrogen simulations, particularly for the LAI + canopy N configuration. This
547 behaviour is consistent with what has already been reported in the literature, namely that the joint
548 assimilation of multiple variables can represent crop growth and development better than the use of
549 LAI alone (Dlamini et al. 2023; Li et al. 2015; Yu et al. 2022).

550 The shift to real data introduces the typical complexities of an operational scenario in which the
551 assimilated variables are derived from actual Sentinel-2 images and are therefore unavoidably
552 affected by estimation errors (Huang et al. 2024; Luo et al. 2023; Richetti et al. 2019). In this context,
553 the importance of above-ground biomass clearly emerges. Among the variables considered, it is the
554 one showing the best agreement with the measurements collected in the field. This characteristic
555 translates into better model performances when this variable is used in the assimilation. On the other
556 hand, setups that relied only on LAI or canopy nitrogen content didn't show results as good for
557 predicting yield, and sometimes these approaches were even less efficient than the "open-loop"
558 model. This behaviour is open to multiple interpretations: for example, aboveground biomass
559 represents an integrated measure of growth and, in the model, is closely linked to yield, which to a
560 first approximation can be seen as a fraction of it; constraining the model on this variable therefore
561 means acting on a quantity that is very close to the target. In contrast, LAI and canopy nitrogen
562 content, while describing essential aspects of crop structure and nutritional status, provide more
563 indirect information on yield, as they do not directly quantify the amount of dry matter transferred to
564 the grain. The LAI in DSSAT is also simulated as expansive growth process of individual leaves, that
565 is if phenology is well simulated the accumulated biomass is partitioned among organs at different
566 degrees. The simulation of biomass accumulation and individual leaf area/growth is made by
567 equations that consider potential conditions, following the Wageningen concepts of production (Van
568 Ittersum et al. 2003). Then the actual biomass and leaf area are minimized by stress factor like water,
569 nitrogen, temperature and CO₂. Therefore, LAI is not a value that is directly simulated but it is scaled
570 up from individual leaf growth that depends on the accumulated biomass and the environmental
571 stressors. For example, expansive growth processes (like the individual leaf growth) are very sensitive

572 to water stress. Therefore, the initial conditions like the initial soil water content and soil mineral
573 nitrogen becomes important input because if they are not correct the model would simulate the wrong
574 water balance causing more or less stress affecting leaf growth. In addition to this, uncertainties also
575 affect the estimation of remotely sensed variables: biomass showed the closest agreement with field
576 measurements, whereas LAI and canopy nitrogen content showed weaker correlations (Figure 4).
577 This aspect is particularly relevant for LAI, which is also affected by uncertainty in both field
578 measurements and satellite retrievals. Optically derived vegetation signals may become less sensitive
579 under dense canopies, so that LAI estimation becomes more uncertain when canopy development is
580 high (Gao et al. 2023). This combination of modelling and observational uncertainty may partly
581 explain why LAI alone provided a weaker constraint on model trajectories than aboveground
582 biomass.

583 Another aspect that may help explain the real-data results is illustrated by the graph in Figure 5, where
584 the biomass fit improves considerably after assimilation. Q1 was chosen as an illustrative case
585 because it showed a distinct assimilation response at the trajectory level under field conditions likely
586 influenced by local heterogeneity not fully captured by the model inputs. Since the soil information
587 used in this study was derived from previously spatialized soil maps rather than a dedicated ESU-
588 scale soil characterization (Zanini et al. 2025), small-scale differences affecting water dynamics may
589 not have been fully represented in the model inputs. In addition, unobserved differences in local initial
590 soil condition, including soil water status, may also have contributed to the mismatch between
591 simulated and observed biomass, as these initial states are known to influence crop model simulations
592 under water limited conditions (Yao et al. 2025). However, because common input assumptions were
593 adopted across all ESUs, this factor alone is unlikely to explain the pattern observed only in Q1 and
594 may instead have interacted with local heterogeneity not represented in the available inputs. This
595 uncertainty in the soil input data may have contributed to the difficulty of DSSAT-CERES-wheat in
596 representing anomalous or severe stress conditions. In our case, the improved fit after assimilation
597 should be interpreted with caution, since part of the mismatch may have originated from uncertainties
598 in the available inputs rather than from the assimilated variables themselves. However, these results
599 suggest that aboveground biomass can be a particularly informative variable for assimilation under
600 real-field conditions, especially in situations where local variability may not be fully represented by
601 the available inputs (Dlamini et al. 2023; Huang et al. 2024).

602 Moving from these point-based cases to the scale of the entire field, the pixel-level assimilation
603 experiment shows that a similar contrast between open-loop and assimilated simulations also emerges
604 in the representation of spatial variability. In the open-loop simulation, the yield map reflects mainly
605 the differences among nitrogen fertilization levels, with strong contrast between management zones

606 and limited variability within each zone. This results in a rigid spatial pattern that does not reproduce
607 the more complex variability observed in the field (Pasquel et al. 2022). By contrast, the assimilation
608 of remotely sensed biophysical variables makes the simulated map more consistent with the spatial
609 pattern of measured yield, reducing dominant effect of fertilization differences and recovering part of
610 the variability (Basso and Liu 2019; Fabbri et al. 2023). This improvement should nevertheless be
611 interpreted as the effect of stronger observational constraints on simulated crop states, rather than
612 evidence that all underlying field-scale processes were fully represented.

613 These results should be interpreted considering several limitations. The study was conducted in a
614 single area and over only one growing season, so the robustness of the proposed framework remains
615 to be tested under different conditions and pedoclimatic contexts. Only one crop model and one
616 optimization-based assimilation scheme were evaluated, without exploring the potential of alternative
617 models or sequential approaches better suited to in-season updating (Jin et al. 2018). Third, the
618 uncertainty associated with the remotely sensed variables was not explicitly incorporated into the
619 assimilation process, even though the results clearly indicate that differences in retrieval quality
620 among biomass, LAI and canopy nitrogen content strongly affected the performance of tested
621 schemes. Future developments should include multi-year and multi-site applications, comparison
622 among different crop models and assimilation algorithms, and the explicit integration of observational
623 uncertainty, for example through differentiated weighting strategies or Bayesian approaches (Dlamini
624 et al. 2023; Luo et al. 2023). A further direction concerns the use of sequential schemes to dynamically
625 update yield forecasts during the season, exploiting satellite observations as they become available
626 (Khodarahmi and Maihami 2023; Silvestro et al. 2021; Wang et al. 2024). Finally, linking the
627 simulated maps to decision rules for nitrogen management will make it possible to assess to what
628 extent the benefits at the modelling level translate into concrete advantages for precision agriculture
629 (Fabbri et al. 2023)

630 **CONCLUSION**

631 Overall, the results obtained provide some useful indications for the combined use of crop models
632 and remote sensing data in wheat management. It appears that the choice of variables to be assimilated
633 cannot be based only on their availability but must consider both their physiological link with the
634 outputs of interest and the quality of their remote sensing estimates (Li et al. 2015; Yu et al. 2022). In
635 this study, above-ground biomass proved to be the most suitable variable to constrain the model and
636 improve yield estimation. The experiments with synthetic data indicated that, under idealised
637 conditions, some multivariable schemes can slightly improve the joint prediction of yield and grain
638 nitrogen compared with the best single-variable configurations. When the framework was applied to

639 real Sentinel-2 products, however, the additional benefit of multivariable assimilation over the
 640 scheme based solely on biomass was limited and not systematic, reflecting the lower quality and
 641 higher noise affecting LAI and canopy nitrogen in this dataset. These findings support the central role
 642 of aboveground biomass in assimilation schemes for yield prediction, while highlighting that
 643 contribution of LAI and canopy nitrogen content depends strongly on the quality of their satellite
 644 retrievals. The extension of the framework to the pixel scale and the possibility of obtaining high-
 645 resolution yield maps from a limited number of ground measurements indicate that this type of
 646 approach can support the delineation of management zones and the ex post evaluation of
 647 differentiated fertilisation strategies and, in perspective, the design of variable-rate applications
 648 (Dlamini et al. 2023; Luo et al. 2023).

649 REFERENCES

- 650 Asseng, S., & Van Herwaarden, A. F. (2003). *Analysis of the benefits to wheat yield from*
 651 *assimilates stored prior to grain filling in a range of environments*. *Plant and Soil* (Vol. 256).
 652 www.apsim-help.tag.csiro.au.
- 653 Basso, B., Ritchie, J. T., Pierce, F. J., Braga, R. P., & Jones, J. W. (2001). Spatial validation of crop
 654 models for precision agriculture. *Agricultural Systems*, 68(2), 97–112.
- 655 Basso, B. & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. In
 656 *Advances in Agronomy* (Vol. 154). <https://doi.org/10.1016/bs.agron.2018.11.002>
- 657 Cammarano, D., Basso, B., Holland, J., Gianinetti, A., Baronchelli, M., & Ronga, D. (2021).
 658 Modeling spatial and temporal optimal N fertilizer rates to reduce nitrate leaching while
 659 improving grain yield and quality in malting barley. *Computers and Electronics in Agriculture*,
 660 182. <https://doi.org/10.1016/j.compag.2021.105997>
- 661 Curnel, Y., de Wit, A. J. W., Duveiller, G., & Defourny, P. (2011). Potential performances of
 662 remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment.
 663 *Agricultural and Forest Meteorology*, 151(12), 1843–1855.
 664 <https://doi.org/10.1016/j.agrformet.2011.08.002>
- 665 De Clerck, E., D.Kovács, D., Berger, K., Schlerf, M., & Verrelst, J. (2024). Optimizing hybrid
 666 models for canopy nitrogen mapping from Sentinel-2 in Google Earth Engine. *ISPRS Journal*
 667 *of Photogrammetry and Remote Sensing*, 218, 530–545.
 668 <https://doi.org/10.1016/j.isprsjprs.2024.11.005>
- 669 Delrcolle, R., Maas, S. J., Gurrit, M., & Baret, F. (1992). *Remote sensing and crop production*
 670 *models: present trends*. *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 47).
- 671 Dlamini, L., Crespo, O., van Dam, J., & Kooistra, L. (2023, August 1). A Global Systematic Review
 672 of Improving Crop Model Estimations by Assimilating Remote Sensing Data: Implications for
 673 Small-Scale Agricultural Systems. *Remote Sensing*. Multidisciplinary Digital Publishing
 674 Institute (MDPI). <https://doi.org/10.3390/rs15164066>

675 Fabbri, C., Basso, B., Napoli, M., Dalla Marta, A., Orlandini, S., & Martinez-Feria, R. A. (2023).
676 Developing a tactical nitrogen fertilizer management strategy for sustainable wheat production.
677 *European Journal of Agronomy*, 144. <https://doi.org/10.1016/j.eja.2023.126746>

678 Gao, S., Zhong, R., Yan, K., Ma, X., Chen, X., Pu, J., et al. (2023). Evaluating the saturation effect
679 of vegetation indices in forests using 3D radiative transfer simulations and satellite
680 observations. *Remote Sensing of Environment*, 295. <https://doi.org/10.1016/j.rse.2023.113665>

681 Gobbo, S., Lo Presti, S., Martello, M., Panunzi, L., Berti, A., & Morari, F. (2019). Integrating
682 SEBAL with in-field crop water status measurement for precision irrigation applications-a case
683 study. *Remote Sensing*, 11(17). <https://doi.org/10.3390/rs11172069>

684 Godwin, D.C., Singh, U. (1998). Nitrogen balance and crop response to nitrogen in upland and
685 lowland cropping systems. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (eds)
686 Understanding Options for Agricultural Production. Systems Approaches for Sustainable
687 Agricultural Development, vol 7. Springer, Dordrecht. [https://doi.org/10.1007/978-94-017-](https://doi.org/10.1007/978-94-017-3624-4_4)
688 3624-4_4

689 Hao, S., Ryu, D., Western, A., Perry, E., Bogen, H., & Franssen, H. J. H. (2021). Performance of a
690 wheat yield prediction model and factors influencing the performance: A review and meta-
691 analysis. *Agricultural Systems*. <https://doi.org/10.1016/j.agsy.2021.103278>

692 Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White, J. W.,
693 Asseng, S., Lizaso, J. I., Moreno, L. P., Pavan, W., Ogoshi, R., Hunt, L. A., Tsuji, G. Y., &
694 Jones, J. W. (2019). The DSSAT crop modeling ecosystem. In K. J. Boote (Ed.), *Advances in*
695 *Crop Modeling for a Sustainable Agriculture* (pp. 173–216). Burleigh Dodds Science
696 Publishing. <https://doi.org/10.19103/AS.2019.0061.10>

697 Huang, J., Song, J., Huang, H., Zhuo, W., Niu, Q., Wu, S., et al. (2024, December 1). Progress and
698 perspectives in data assimilation algorithms for remote sensing and crop growth model.
699 *Science of Remote Sensing*. Elsevier B.V. <https://doi.org/10.1016/j.srs.2024.100146>

700 Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018, January 1). A review of
701 data assimilation of remote sensing and crop models. *European Journal of Agronomy*. Elsevier
702 B.V. <https://doi.org/10.1016/j.eja.2017.11.002>

703 Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P.
704 W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model.
705 *European Journal of Agronomy*, 18(3–4), 235–265. [https://doi.org/10.1016/S1161-](https://doi.org/10.1016/S1161-0301(02)00107-7)
706 0301(02)00107-7.

707 Jones, James W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., et al. (2017). Toward
708 a new generation of agricultural system data, models, and knowledge products: State of
709 agricultural systems science. *Agricultural Systems*, 155, 269–288.
710 <https://doi.org/10.1016/j.agsy.2016.09.021>

711 Kamir, E., Waldner, F., & Hochman, Z. (2020). Estimating wheat yields in Australia using climate
712 records, satellite image time series and machine learning methods. *ISPRS Journal of*

713 *Photogrammetry and Remote Sensing*, 160, 124–135.
714 <https://doi.org/10.1016/j.isprsjprs.2019.11.008>

715 Khodarahmi, M., & Maihami, V. (2023, January 1). A Review on Kalman Filter Models. *Archives of*
716 *Computational Methods in Engineering*. Springer Science and Business Media B.V.
717 <https://doi.org/10.1007/s11831-022-09815-7>

718 Li, Z., Jin, X., Zhao, C., Wang, J., Xu, X., Yang, G., et al. (2015). Estimating wheat yield and
719 quality by coupling the DSSAT-CERES model and proximal remote sensing. *European*
720 *Journal of Agronomy*, 71, 53–62. <https://doi.org/10.1016/j.eja.2015.08.006>

721 Li, Z., Wang, J., Xu, X., Zhao, C., Jin, X., Yang, G., & Feng, H. (2015). Assimilation of two
722 variables derived from hyperspectral data into the DSSAT-CERES model for grain yield and
723 quality estimation. *Remote Sensing*, 7(9), 12400–12418. <https://doi.org/10.3390/rs70912400>

724 Luo, L., Sun, S., Xue, J., Gao, Z., Zhao, J., Yin, Y., et al. (2023, August 1). Crop yield estimation
725 based on assimilation of crop models and remote sensing data: A systematic evaluation.
726 *Agricultural Systems*. Elsevier Ltd. <https://doi.org/10.1016/j.agsy.2023.103711>

727 Maas, S. J. (1988). Use of remotely-sensed information in agricultural crop growth models.
728 *Ecological Modelling*, 41(3–4), 247–268. [https://doi.org/10.1016/0304-3800\(88\)90031-2](https://doi.org/10.1016/0304-3800(88)90031-2)

729 Messina, C. D., Jones, J. W., Boote, K. J., & Vallejos, C. E. (2006). A gene-based model to simulate
730 soybean development and yield responses to environment. *Crop Science*, 46(1), 456–466.
731 <https://doi.org/10.2135/cropsci2005.04-0372>

732 Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *Journal of Applied*
733 *Ecology*, 9(3), 747–766. <https://doi.org/10.2307/2401901>.

734 Morales, A., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past
735 or the future. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1128388>

736 Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer*
737 *Journal*, 7(4). <https://doi.org/10.1093/comjnl/7.4.308>

738 Nieke, J., Despoisse, L., Gabriele, A., Weber, H., Strese, H., Ghasemi, N., et al. (2023). The
739 copernicus hyperspectral imaging mission for the environment (CHIME): an overview of its
740 mission, system and planning status. <https://doi.org/10.1117/12.2679977>

741 Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022, August 1). A
742 review of methods to evaluate crop model performance at multiple and changing spatial scales.
743 *Precision Agriculture*. Springer. <https://doi.org/10.1007/s11119-022-09885-4>

744 Richetti, J., Boote, K. J., Hoogenboom, G., Judge, J., Johann, J. A., & Uribe-Opazo, M. A. (2019).
745 Remotely sensed vegetation index and LAI for parameter determination of the CSM-
746 CROPGRO-Soybean model when in situ data are not available. *International Journal of*
747 *Applied Earth Observation and Geoinformation*, 79, 110–115.
748 <https://doi.org/10.1016/j.jag.2019.03.007>

749 Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., et al. (2014). Assessing
750 agricultural risks of climate change in the 21st century in a global gridded crop model

751 intercomparison. *Proceedings of the National Academy of Sciences of the United States of*
752 *America*, 111(9), 3268–3273. <https://doi.org/10.1073/pnas.1222463110>

753 Seidel, S. J., Palosuo, T., Thorburn, P., & Wallach, D. (2018). Towards improved calibration of crop
754 models – Where are we now and where should we go? *European Journal of Agronomy*, 94.
755 <https://doi.org/10.1016/j.eja.2018.01.006>

756 Shaik, R. U., Periasamy, S., & Zeng, W. (2023). Potential Assessment of PRISMA Hyperspectral
757 Imagery for Remote Sensing Applications. *Remote Sensing*, 15(5).
758 <https://doi.org/10.3390/rs15051378>

759 Shiferaw, B., Smale, M., Braun, H. J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops
760 that feed the world 10. Past successes and future challenges to the role played by wheat in
761 global food security. *Food Security*, 5(3). <https://doi.org/10.1007/s12571-013-0263-y>

762 Silvestro, P. C., Casa, R., Hanuš, J., Koetz, B., Rascher, U., Schuettemeyer, D., et al. (2021).
763 Synergistic use of multispectral data and crop growth modelling for spatial and temporal
764 evapotranspiration estimations. *Remote Sensing*, 13(11). <https://doi.org/10.3390/rs13112138>

765 Van Ittersum, M K, Leffelaar, P. A., Van Keulen, H., Kropff, M. J., Bastiaans, L., & Goudriaan, J.
766 (2003). *On approaches and applications of the Wageningen crop models*.
767 <http://www.dpw.wageningen-ur.nl/pp>

768 Van Ittersum, Martin K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013).
769 Yield gap analysis with local to global relevance-A review. *Field Crops Research*, 143, 4–17.
770 <https://doi.org/10.1016/j.fcr.2012.09.009>

771 van Klompenburg, T., Kassahun, A., & Catal, C. (2020, October 1). Crop yield prediction using
772 machine learning: A systematic literature review. *Computers and Electronics in Agriculture*.
773 Elsevier B.V. <https://doi.org/10.1016/j.compag.2020.105709>

774 Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J. P., Lewis, P.,
775 et al. (2019, May 1). Quantifying Vegetation Biophysical Variables from Imaging
776 Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics*. Springer Science
777 and Business Media B.V. <https://doi.org/10.1007/s10712-018-9478-y>

778 Verrelst, J., Rivera-Caicedo, J. P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., et al.
779 (2021). Mapping landscape canopy nitrogen content from space using PRISMA data. *ISPRS*
780 *Journal of Photogrammetry and Remote Sensing*, 178, 382–395.
781 <https://doi.org/10.1016/j.isprsjprs.2021.06.017>

782 Wagner, M. P., Slawig, T., Taravat, A., & Oppelt, N. (2020). Remote sensing data assimilation in
783 dynamic crop models using particle swarm optimization. *ISPRS International Journal of Geo-*
784 *Information*, 9(2). <https://doi.org/10.3390/ijgi9020105>

785 Wallach, D., Buis, S., Seserman, D. M., Palosuo, T., Thorburn, P. J., Mielenz, H., et al. (2024). A
786 calibration protocol for soil-crop models. *Environmental Modelling and Software*, 180.
787 <https://doi.org/10.1016/j.envsoft.2024.106147>

- 788 Wang, C., Gao, C., Zhang, J., Hao, Y., Jia, Y., Liu, Z., et al. (2023). *Field scale assessment of winter*
789 *wheat yield in Loess Plateau of 2 China by assimilating the DSSAT model and Sentinel-2 data*
790 *3 with EnKF algorithm*. <https://ssrn.com/abstract=4668779>
- 791 Wang, D., Tan, D., & Liu, L. (2018). Particle swarm optimization algorithm: an overview. *Soft*
792 *Computing*, 22(2), 387–408. <https://doi.org/10.1007/s00500-016-2474-6>
- 793 Wang, J., Wang, Y., & Qi, Z. (2024, September 1). Remote Sensing Data Assimilation in Crop
794 Growth Modeling from an Agricultural Perspective: New Insights on Challenges and
795 Prospects. *Agronomy*. Multidisciplinary Digital Publishing Institute (MDPI).
796 <https://doi.org/10.3390/agronomy14091920>
- 797 Welch, G., & Bishop, G. (1997). *An Introduction to the Kalman Filter*. <http://www.cs.unc.edu/~gb>
- 798 Whelan, B. M., McBratney, A. B., & Minasny, B. (2002). VESPER 1.5: Spatial prediction software
799 for precision agriculture. In Proceedings of the 6th International Conference on Precision
800 Agriculture.
- 801 Woche, M., Berger, K., Verrelst, J., & Hank, T. (2022). Retrieval of carbon content and biomass
802 from hyperspectral imagery over cultivated areas. *ISPRS Journal of Photogrammetry and*
803 *Remote Sensing*, 193, 104–114. <https://doi.org/10.1016/j.isprsjprs.2022.09.003>
- 804 Yao, N., Wei, Y., Jiang, K., Liu, J., Li, Y., Ran, H., et al. (2025). Nonlinear water stress response
805 functions can improve the performance of the DSSAT-CERES-Wheat model under water
806 deficit conditions. *Agricultural Water Management*, 307.
807 <https://doi.org/10.1016/j.agwat.2024.109235>
- 808 Yu, D., Zha, Y., Shi, L., Ye, H., & Zhang, Y. (2022). Improving sugarcane growth simulations by
809 integrating multi-source observations into a crop model. *European Journal of Agronomy*, 132.
810 <https://doi.org/10.1016/j.eja.2021.126410>
- 811 Zanini, M., Heiden, U., Pace, L., Casa, R., & Priori, S. (2025). Soil Reflectance Composite for
812 Digital Soil Mapping in a Mediterranean Cropland District. *Remote Sensing*, 17(1).
813 <https://doi.org/10.3390/rs17010089>