100. Potential of hyperspectral satellite biophysical variables in a crop model assimilation scenario

R. Fazioli[1], F. Carucci[1], L. Marrone[1], S. Pignatti[2], S. Mirazei[2], F. Rossi[2], S. Lo Presti[3], D. Kovács[4] and R. Casa[1,*]

[1]*Department of Agricultural and Forestry scieNcEs (DAFNE), University of Tuscia, Via San Camillo de Lellis, 01100 Viterbo, Italy; * rcasa@unitus.it*

[2]*Institute of Methodologies for Environmental Analysis (IMAA), National Research Council of Italy (CNR), Via del Fosso del Cavaliere, 100, 00133 Roma, Italy*

[3]*Centrale Valutativa, 00131 Roma, Italy*

[4] *Image Processing Laboratory (IPL) –University of Valencia, Catedrático Agustín Scardino Benlloch 9, Paterna, 46980, Spain*

Abstract

This study evaluated the use of remotely sensed biophysical variables for estimation of wheat yield and grain nitrogen, in the DSSAT model. In a synthetic data experiment, seasonal patterns of LAI, biomass and canopy nitrogen were generated by model simulation, across various scenarios. Calibration of the DSSAT model with the assimilated biomass and canopy N significantly enhanced the grain yield and N estimation accuracy of the model. Validation with real data from Sentinel-2 generally confirmed these results, albeit with reduced accuracy. This work indicates that the use of hyperspectral satellite data allowed better estimation of biomass and canopy N, improving yield estimation from data assimilation with the DSSAT model.

Keywords: aboveground biomass, canopy nitrogen content, crop model, DSSAT-CERES-Wheat, Sentinel-2

Introduction

The assimilation of biophysical variables estimated from remote sensing into crop simulation models has a large potential for providing spatially and temporally consistent yield and quality maps at the field scale. This information is particularly valuable for a range of applications, such as precision agronomic management, or for the crop insurance market. Data assimilation into crop growth models has the advantage of being a more generalizable approach as compared to data driven methods, such as those based on machine learning, which are strongly dependent on the training data and cannot be easily extrapolated.

In most of the previous studies, only leaf area index (LAI), typically estimated from multispectral data, has been used for data assimilation into crop models. The recent availability of hyperspectral satellites, such as PRISMA and EnMAP (Shaik *et al.*, 2023) and the forthcoming launch of CHIME (Celesti *et al.*, 2022), extends the capability of estimation of biophysical variables, as compared to what was possible using multispectral data, in particular providing better estimates of canopy nitrogen and biomass (Verrelst *et al.*, 2021).

This study aimed to assess the potential advantage of assimilating crop aboveground biomass and canopy nitrogen content, in addition to LAI, into the DSSAT-CERES-Wheat model, to enhance wheat grain yield and nitrogen content predictions. The underlying research hypothesis was that the availability of these variables from hyperspectral remote sensing would lead to better yield and quality prediction. To better understand the behaviour of the DSSAT model in an ideal situation, a first step of the study was based on a methodology which involved the use of synthetic data, i.e.

generated by the DSSAT model, instead of real data, using the approach of Curnel *et al.* (2011). Further tests were then carried out using real data collected in wheat in Central Italy.

Materials and methods

Crop growth model

The CERES-Wheat simulation model, which is part of the Decision Support System for Agrotechnology Transfer (DSSAT) version 4.8.2 (Hoogenboom *et al.*, 2019), was used. This model simulates daily crop growth and development based on cultivar, environmental, and management factors. The analysis described in this paper was carried out by implementing DSSAT within R Studio v. 4.2.0 (https://rstudio.com/), using the R packages DSSAT (Alderman, 2021) and CroptimizR (Buis *et al.*, 2023).

Ground data and site information

Data collected from field monitoring campaigns conducted in Rieti (Central Italy) during the 2022–2023 wheat season were used to ensure that synthetic observations generated from DSSAT simulations reflected realistic conditions. Additionally, data from the 2023-2024 field monitoring campaign in Rieti were used to validate the results obtained with synthetic data.
The study site is located in the Rieti Plain in Central Italy (lat. 42°25′ N, long. 12°52′ E, alt. 400 m). Monitoring campaigns were carried out in wheat fields during the 2022–2023 and 2023–2024 seasons, collecting data from respectively 20 and 10 elementary sampling units (ESU), each consisting of a square of 30 m×30 m each. During the wheat growing cycle, approximately every two weeks, crop aboveground biomass was harvested from four sampling areas within each ESU, by cutting at the ground level all the plants falling within a 0.25 m^2 ground sampling frame. Direct LAI determination was carried out using the LI-3100C area meter (LI-COR) on a subsample of the biomass. At the end of the crop cycle, the final yield and grain nitrogen content were measured for each ESU. Canopy and grain N was assessed after grinding dry samples, using the Aurora NIR (Grainit) instrument.

Synthetic data experiment

Figure 1 presents an overview of the synthetic data experiment (Curnel *et al.*, 2011).

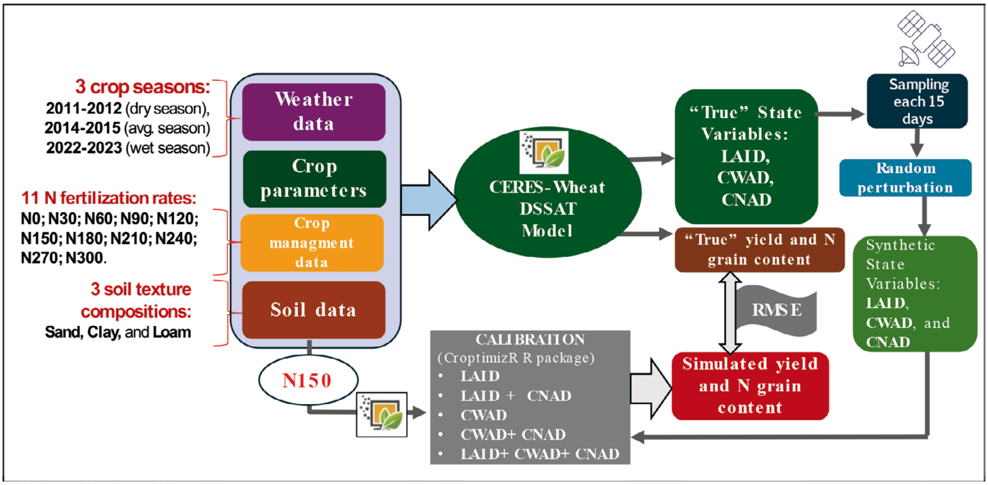

Figure 1. Overview of the synthetic data experiment. LAID, LAI (leaf area index); CWAD, crop aboveground biomass, CNAD, nitrogen canopy content, RMSE, root mean squared error.

Different scenarios were simulation with DSSAT, varying N fertilisation rates from 0 to 300 kg N/ha with a 30 kg N/ha step, under contrasting weather and soil conditions. A dry (2011–2012), average (2014–2015) and wet (2022–2023) year were chosen from Rieti rainfall records with respectively 282, 403 and 904 mm during the wheat growth season. Three contrasting soil profiles, respectively clay (60% clay, 30% silt), loam (32% clay, 25% silt), and sand (30% clay, 10% silt) texture were also used. DSSAT simulations for all these scenarios were used to generate synthetic observations of LAI (called LAID in DSSAT), aboveground biomass (CWAD), and Nitrogen Canopy Content (CNAD), by sampling the variables at 15 day intervals throughout the season (assumed as a likely frequency of satellite data). A random perturbation was then added, by sampling from a normal distribution with a zero mean and a coefficient of variation (CV) of 30%. For each simulated scenario, these data were considered as synthetic satellite observations of biophysical variables. The corresponding final grain yield and grain N simulated for each of these scenarios were considered as "true" values. Data assimilation was then carried out, by a calibration employing the synthetic seasonal biophysical variables, for all scenarios, always starting from the same initial conditions with a non-calibrated parametrization of the model, assuming a rate of 150 kg N/ha. Each time a different set of observed synthetic variables was used for the calibration (i.e. only LAID, LAID+CNAD, only CWAD, CWAD+CNAD, and LAID+CWAD+CNAD). The CroptimizR package (Buis *et al.*, 2023), was used for the calibration, exploiting the Nelder-Mead simplex optimization algorithm. The DSSAT parameters that were varied during the calibration were PARUE and PARU2, that express respectively the potential radiation use efficiency of the crop (RUE) and the influence of stress factors on the RUE. At the end of the procedure, final yields and N grain content estimations obtained after the calibration runs were compared with the "true" values, to assess the performance of the different in-season synthetic observed variables for model calibration by computing the root mean square error (RMSE).

Tests with real data

For assessing the assimilation of DSSAT using real data, biophysical variables estimated from Sentinel-2 during the 2024 wheat season in the Rieti plain were used, as a surrogate for the estimates from hyperspectral satellites, as a sufficient time series of PRISMA images was not available. LAI was obtained for each clear sky Sentinel-2 image available during the 2023-2024 wheat growth season, using the BV-NEt algorithm (Weiss and Baret, 2016), implemented in the ESA SNAP software (https://step.esa.int/main/toolboxes/snap/). Canopy N was retrieved from Sentinel-2 data using the PyEOGPR package (Kovács, 2024) based on Gaussian Process Regression (De Clerck *et al.* 2024). Aboveground biomass was estimated from a Monteith type model implemented in the Tethys platform (https://tethys.farm/en/), exploiting the fraction of absorbed photosynthetic radiation (FAPAR) obtained from Sentinel-2. These biophysical variables were employed for the assimilation of DSSAT using the same method as used for the synthetic data experiment. Estimated grain yield and grain nitrogen were then compared with measured data collected in Rieti at crop harvest, by computing the RMSE and the coefficient of determination R^2.

Results and discussion

After the calibration, the dynamic variables showed, in general, a good fit with the synthetic observations, confirming the effectiveness of the calibration procedure.

Across all meteorological and soil scenarios, it was found that combining all observed dynamic variables (LAID, CWAD, and CNAD) led to an improved grain yield estimation, with a reduction of RMSE of approximately 10% compared to using only LAI observations (Table 1). However, for grain N, better results with a lower RMSE (about 5.5% lower than using only LAI) were observed when both LAI and canopy N content (CNAD) were used (Table 1).

With respect to the different soil and weather conditions, it was found that for a sandy soil the scenarios using only LAI values for calibration were the best choice for grain yield estimation,

Table 1. Mean RMSEs for estimation of final grain yield and grain N content after DSAAT calibration with observed seasonal variables, across all simulation scenarios, and their change relative to the case where only LAI observation were used, for each set of observed dynamic variables used (LAID, LAID+CNAD, CWAD, CWAD+CNAD, and LAID+CWAD+CNAD).

Biophysical variables used for assimilation	Grain yield		Grain N	
	RMSE (t/ha)	ΔRMSE (%)[a]	RMSE (%)	ΔRMSE (%)[a]
LAID	0.40	–	0.14	–
LAID+CNAD	0.38	−4.55	0.13	−5.47
CWAD	0.40	+1.54	0.14	−2.34
CWAD+CNAD	0.43	+7.71	0.15	+4.69
LAID+CWAD+CNAD	0.36	−10.36	0.14	+0.78

LAID, leaf area index; CWAD, crop aboveground biomass; CNAD, nitrogen canopy content.

[a] Relative change was calculated with respect to the RMSE of LAID treatment.

under all the different meteorological scenarios. However, in loam and clay soils scenarios, the use of three variables jointly for the assimilation, i.e. LAI, canopy N, and biomass, for the assimilation, resulted in better grain yield estimation than using only LAI. The only exception was the clay soil scenario under dry meteorological conditions, in which the calibration using two variables, i.e. biomass and canopy N, provided better results (Figure 2A). It is reasonable to expect a different calibration performance for different soil types, since, for example, for a sandy soil the crop would be mostly responsive to water stress rather than to nitrogen stress. The grain nitrogen estimation results varied across different scenarios. However, calibration with only LAI was generally sufficient in dry-sand, medium-sand, and medium-loam scenarios. In the other scenarios, LAI in combination with canopy N and biomass helped to improve the estimation results (Figure 2B).

These results indicate an advantage in using biomass and canopy N in addition to LAI during the assimilation process, for the estimation of yield and grain N. Whereas LAI is well-estimated from satellite multispectral data such as Sentinel-2, aboveground biomass and canopy N would require hyperspectral data (Verrelst *et al*., 2021; Wocher *et al*., 2022). However, verification of the synthetic data results with real data was not possible due to the unavailability of an adequate time series of hyperspectral satellite images for the study site; only two PRISMA images acquired during the wheat growth season of 2024 in the Rieti Plain. In fact, because PRISMA is a scientific precursor and not an operational mission, a repeated acquisition pattern is not possible. For this reason, Sentinel-2 data were used for the estimation of all biophysical variables, having a total of 15 clear sky satellite images available during the wheat growing season. The estimation of LAI from Sentinel-2 using the BV-N*Et al*gorithm was quite good when compared to direct ground measurements, with a R^2 of 0.82 and a RMSE of 0.96. The estimation of aboveground biomass using the model implemented in the Tethys platform was very good, with an R^2 of 0.98 and an RMSE of 96.7 g/m^2. The estimation of canopy N, using hybrid models with PROSAIL-PRO (De Clerck *et al.* 2024), compared with ground measurements, provided an R^2 of 0.48 and an RMSE of 52.7 kg N/m^2, generally underestimating at the beginning and end of the growth season and overestimating during the peak period.

The use of all these biophysical variables for the assimilation into the DSSAT crop model, using the same calibration method as illustrated above for the synthetic data study, revealed that, in most cases, the estimation of yield and grain N content was better in the open-loop simulations, i.e. in model runs without assimilation, than after the calibrations, in terms of RMSE (Table 2).

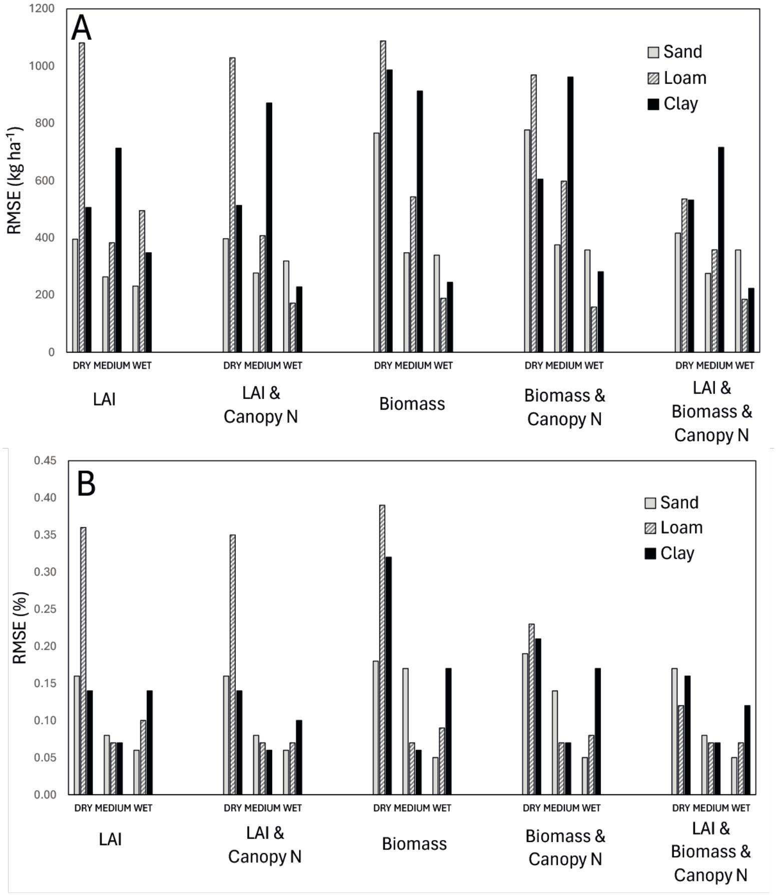

Figure 2. Results of the assimilation of different biophysical variables (LAI, aboveground biomass and canopy N, alone or in combination) in the DSSAT model using synthetic observations, for the estimation of grain yield (A) and grain nitrogen (B), under different meteorological (dry, medium and wet years) and soil texture (sand, loam, clay) scenarios.

Table 2. Mean RMSEs and their change, relative to the open-loop DSSAT model runs, for the estimation of final grain yield and grain N content for the Rieti 2024 ground campaigns, after DSAAT calibration (assimilation) with biophysical variables estimated from Sentinel-2.

Biophysical variables used for assimilation	Grain yield			Grain N content		
	R^2	RMSE (t/ha)	ΔRMSE (%)[a]	R^2	RMSE (%)	ΔRMSE (%)[a]
Open Loop Default	0.56	1.11	–	0.06	0.46	–
LAID	0.35	2.30	+1.19	0.19	0.79	+0.33
LAID+CNAD	0.17	2.42	+1.31	0.03	0.71	+0.25
CWAD	0.69	2.50	+1.39	0.06	0.68	+0.22
CWAD+CNAD	0.59	2.56	+1.45	0.003	0.71	+0.25
LAID+CWAD+CNAD	0.68	2.40	+1.29	0.02	0.69	+0.23

LAID, leaf area index; CWAD, crop aboveground biomass; CNAD, nitrogen canopy content.

[a] Relative change was calculated with respect to the RMSE of the Open Loop simulations.

However, the assimilation of aboveground biomass, alone or in combination with other biophysical variables, provided better R^2 values for the estimation of yield than the open loop simulations and also of those in which only LAI, or LAI and canopy N were used in the assimilation. When biomass was used in the assimilation, the R^2 was higher than for the open loop case, but the RMSE was also higher (i.e. worse), because the yield was considerably underestimated by the DSSAT model for these cases (Figure 3). The use of LAI, alone or in combination with canopy N, provided worse results than the open loop case (Figure 3B), both in terms of R^2 and RMSE (Table 2).

Concerning the estimation of grain N, the results were rather disappointing (Table 2) but seemed to confirm what was observed in the synthetic data experiment, i.e. that the inclusion of biomass and canopy N, in addition to LAI, does not seem to bring necessarily an advantage in terms of yield estimation accuracy.
In fact, the addition of biomass and canopy N to LAI in the assimilation process did provide an improvement in yield estimation in terms of R^2. However the calibrations employing biomass observations determined an underestimation of crop yield, which could depend on the modelled translocation of assimilates into the grain.

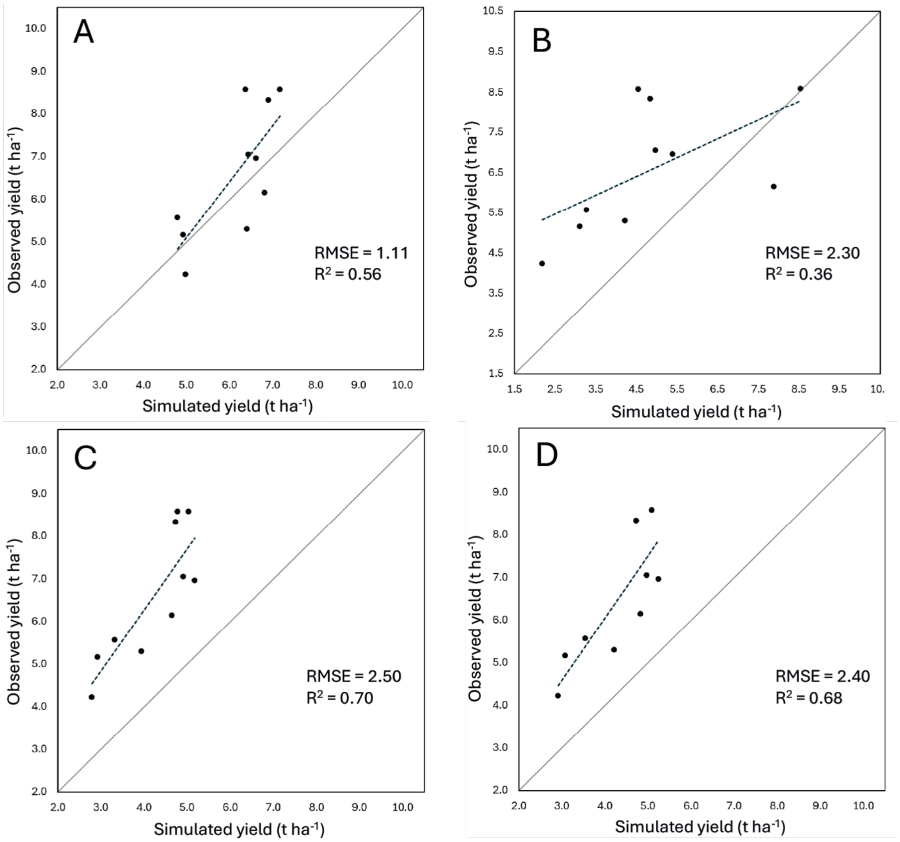

Figure 3. Grain yield estimation results from calibration of the DSSAT crop model using in-season biophysical variables estimated from Sentinel-2 for the Rieti 2024 field campaigns. Open loop simulations from DSSAT without calibration (A); assimilation of Sentinel-2 estimates of LAI (B), aboveground biomass (C); LAI, aboveground biomass and canopy nitrogen (D).

Conclusions

In conclusion, the study with synthetic data revealed that the calibration of the DSSAT model using joint observations of crop variables such as LAI, aboveground biomass and canopy N, which can be retrieved from remote sensing, particularly when using hyperspectral data, generally improves the estimation of wheat grain yield with the DSSAT model, as compared to the case when only LAI is employed. However, results are affected by the interaction between seasonal weather pattern and soil type. The tests carried out with real data, based on the assimilation of variables retrieved from Sentinel-2, given the unavailability of sufficient PRISMA data, generally confirmed the findings of the synthetic data study.
Overall, this work illustrates the potential of including biomass and canopy N content observations, which are variables typically retrieved from hyperspectral sensors such as PRISMA, to improve the yield estimation accuracy.

Acknowledgements

This research was partly funded by the Italian Space Agency ASI within the PRIS4VEG project Contract/Agreement No. N. 2022-5-U.0 CUP No. F43C2200005. We thank Dr. Samuel Buis (INRAE, France) for providing the R DSSAT wrapper and for the assistance with the CroptimizR package.

References

Alderman, P.D. (2021). DSSAT: A comprehensive R Interface for the DSSAT cropping systems model. R package version 0.0.6. Available online at https://CRAN.R-project.org/package=DSSAT (accessed March 2025).

Buis, S., Lecharpentier, P., Vezy, R., & Giner, M. (2023). CroptimizR: a package to estimate parameters of crop models. Available online at https://github.com/SticsRPacks/CroptimizR (accessed March 2025).

Celesti, M., Rast, M., Adams, J., Boccia, V., Gascon, F., Isola, C., & Nieke, J. (2022). The Copernicus Hyperspectral Imaging Mission for the Environment (CHIME): status and planning. In IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, pp. 5011–5014. https://doi.org/10.1109/IGARSS46834.2022.9883592

Curnel, Y., de Wit, A.J.W., Duveiller, G., & Defourny, P. (2011). Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. Agricultural and Forest Meteorology, 151(12), 1843–1855. https://doi.org/10.1016/j.agrformet.2011.08.002

de Clerck, E., Kovács, D., Berger, K., Schlerf, M., & Verrelst, J. (2024). Optimizing hybrid models for canopy nitrogen mapping from Sentinel-2 in Google Earth Engine. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 530–545. https://doi.org/10.1016/j.isprsjprs.2024.11.005

Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., …, & Jones, J.W. (2019). The DSSAT crop modeling ecosystem. In K.J. Boote (ed.), Advances in Crop Modeling for a Sustainable Agriculture. Burleigh Dodds, Cambridge, pp. 173–216.

Kovács, D. (2024). pyeogpr. Zenodo. https://doi.org/10.5281/zenodo.13373838.

Shaik, R. U., Periasamy, S., & Zeng, W. (2023). Potential assessment of PRISMA hyperspectral imagery for remote sensing applications. Remote Sensing, 15(5), 1378. https://doi.org/10.3390/rs15051378

Verrelst, J., Rivera-Caicedo, J.P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., …, & Berger, K. (2021). Mapping landscape canopy nitrogen content from space using PRISMA data. ISPRS Journal of Photogrammetry and Remote Sensing, 178, 382–395. https://doi.org/10.1016/j.isprsjprs.2021.06.017

Weiss, M., & Baret, F. (2016). S2ToolBox level 2 products. Version 1.1. Available online at step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf.ì (accessed March 2025).