

Reviewer 2 Responses

We sincerely thank Reviewer 2 for taking the time to thoroughly review our manuscript and for providing detailed and constructive comments. We greatly appreciate the thoughtful suggestions, particularly those highlighting areas that required further clarification or were previously overlooked. These comments have been invaluable in improving the clarity, rigor, and overall quality of the manuscript.

In the following, the reviewer's comments are presented in **blue**, followed by our corresponding responses outlining the revisions and clarifications made.

Major Comments

1. Lines 15–25: The manuscript states that this study presents the “first integration” of UAV LiDAR structure and intensity with multispectral and thermal data for AGB estimation. This claim appears quite strong and may need to be positioned more carefully relative to existing studies on UAV multi-sensor crop monitoring.

We agree that the phrasing of “first integration” is too strong. Our intention was not to claim that such multi-sensor approaches have not been explored, but rather to highlight the specific combination and systematic evaluation of multiple LiDAR-derived features, particularly the inclusion of intensity and multi-layer gap fraction alongside crop height, in combination with multispectral and thermal data. While studies exist that combine subsets of these sensors, fewer have explicitly evaluated this full set of LiDAR-derived structural and signal-based features together within a unified framework across an entire growing season.

We acknowledge that our original wording overemphasized novelty, partly out of concern that the specific contributions of the study might be overlooked. However, we agree that this resulted in overly strong and non-standard phrasing. Therefore, we plan to revise the manuscript to remove such statements and to adopt more precise, neutral scientific language throughout.

In the revised version, we will position this work as an extension of existing multi-sensor studies, emphasizing the systematic comparison and integration of underused LiDAR features rather than claiming a first-ever integration. These changes will be applied in the Abstract, Introduction, and relevant sections of the manuscript.

2. Results and Discussion sections: LiDAR alone already achieves strong performance, while the fusion model improves the RMSE, I am wondering whether this improvement is worth the additional complexity of combining LiDAR, MS, and TIR sensors in terms of cost, data acquisition, and processing, and above all, modelling and computing costs.

We agree that the additional complexity associated with multi-sensor fusion must be justified in terms of practical benefit, which was not sufficiently emphasized in the original manuscript. As the reviewer correctly notes, LiDAR-only models already achieved strong performance in our study, indicating that LiDAR-derived features capture a large portion of AGB variability. This is partly due to the complementary nature of LiDAR metrics, which include canopy height, canopy density (e.g., gap fraction), and signal intensity. Together, these features provide a multi-dimensional representation of the canopy, where height and density describe structural properties, and intensity provides an empirical signal related to canopy structure and, indirectly, vegetation condition.

While the inclusion of multispectral (MS) and thermal infrared (TIR) data led to further improvements in RMSE, we agree that these gains are relatively modest and may not always justify the added costs and complexity associated with additional sensors, data acquisition, processing workflows, and modelling. In response, we will revise the Discussion to explicitly frame multi-sensor fusion as a trade-off between accuracy and operational complexity. In particular, we will emphasize that, if a single sensor must be selected for biomass estimation, LiDAR provides the most comprehensive information in this dataset.

At the same time, we will clarify that MS and TIR sensors can still play important complementary roles in broader agricultural monitoring. For example, TIR data are commonly used for evapotranspiration and water stress analysis, while MS data provide insights into canopy reflectance and pigment-related dynamics. When these sensors are already being collected for other purposes, their integration into biomass estimation may offer additional, albeit incremental, improvements and contribute to a more complete understanding of crop condition, yield potential, and stress responses.

We therefore position multi-sensor fusion not as a universally necessary approach, but as a context-dependent strategy. In research or precision agriculture contexts where multiple crop processes are monitored simultaneously, the integration of MS and TIR data may provide added value beyond biomass estimation alone. These points will be more clearly emphasized in the revised Discussion.

3. TIR DATA: Collecting TIR data can be challenging because sudden changes in wind or cloud cover may cause significant temperature variations even within a single campaign, and this becomes even more critical when collecting time-series data. I did not find a clear explanation of how the TIR data were collected or how the authors ensured that measurements were obtained under relatively comparable or stable weather conditions.

We agree that UAV-based thermal measurements are sensitive to environmental conditions and sensor instabilities, including microbolometer drift, wind effects, and cloud variability.

In the original manuscript, the thermal data acquisition was described too briefly. We now plan to expand this section to clarify that all flights were conducted near solar noon under predominantly clear-sky conditions. To improve radiometric stability, an external heated shutter was used to perform regular non-uniformity corrections (NUC) and reduce microbolometer drift during each mission. Flight parameters were selected to minimize variability, including relatively low flight speeds ($\sim 6 \text{ m s}^{-1}$) to reduce wind and directional effects. Radiometric parameters (air temperature, humidity, emissivity, and object distance) were consistently defined prior to each flight, and in-field thermal reference targets were used for calibration and quality control. A constant emissivity value was applied, as measurements were restricted to dense winter wheat canopies with minimal soil influence.

We acknowledge that some environmental variability is unavoidable and have added this clarification to the Discussion to encourage cautious interpretation of TIR-derived results. These details are now included in the revised Methods and Discussion sections.

4. Machine Learning Method: Multi-sensor ensemble approaches can improve model performance, but they also introduce additional complexity in data collection, preprocessing, and computational requirements. Given these considerations, I wonder whether a simpler setup, for example, using LiDAR alone combined with more advanced machine learning approaches (transformers), could provide comparable performance. A brief discussion of this point in the discussion, as a possible future direction, might be useful.

We agree that more advanced machine learning approaches, such as transformer-based models, may improve predictive performance by better capturing feature interactions and temporal dynamics.

In this study, our objective was to evaluate sensor contributions rather than optimize model architecture. Therefore, we used a consistent ANN framework to ensure that differences in performance were driven by the input data rather than the model.

We also note that transformer-based approaches typically require substantially larger training datasets to avoid overfitting. Given the limited number of samples in this study, we considered simpler models more appropriate.

We will add a brief discussion highlighting advanced models as a promising direction for future work, particularly for building on the insights identified here using LiDAR-only or reduced-input datasets, provided that sufficiently large training datasets are available.

Minor Comments

Line 122: I would suggest removing this information from here and citing the figure at the end of the sentence (Fig. 1).

We will revise the sentence accordingly by removing the information and placing the figure reference at the end of the sentence (Fig. 1).

Line 115: Make sure that the figure caption stays together with the Figure.

We will revise the manuscript formatting to ensure that the figure caption remains together with the corresponding figure.

Lines 138: Why LiDAR flight height 50 m and MS 100 m? Maybe a short explanation here would help?

We thank the reviewer for pointing this out. We agree that the difference in flight altitudes was not sufficiently explained in the original manuscript and will clarify this in the data acquisition section. In general, with our particular sensor, the LiDAR data are collected at a lower altitude (e.g., 50 m) to ensure sufficient point density and signal strength for reliable estimation of structural metrics such as gap fraction and crop height, given the specifications of the sensor. In contrast, multispectral (MS) and thermal (TIR) data are typically acquired at higher altitudes (e.g., ~100 m) to improve spatial coverage and operational efficiency while maintaining adequate spatial resolution for canopy reflectance analysis.

However, upon re-examination of the dataset, we identified an error in the manuscript: the passive sensors (MS and TIR) were also acquired at 50 m in this experiment, rather than 100 m as originally stated. This deviation from our typical acquisition strategy was due to the relatively small size of the experimental field. In most applications, we operate over larger areas and therefore fly passive sensors at higher altitudes to increase efficiency while maintaining sufficiently high spatial resolution. We will correct this inconsistency in the revised manuscript.

Finally, we note that newer LiDAR systems may support higher-altitude acquisitions while maintaining sufficient point density; however, the flight parameters used in this study were selected to match the performance characteristics of the specific sensor employed

Lines 226: CWSI: Please explain this abbreviation.

We will now define CWSI as *Crop Water Stress Index* at its first occurrence in the manuscript. In addition, we plan to expand the description to provide more context on how CWSI is used and its formulation, along with supporting references to relevant literature.

Lines 239–241: " All sensor features were normalized 240 to a 0–1 range" : I suggest removing this sentence, as it does not seem to be directly related to the topic discussed in this paragraph. belong here, and you mention two times in the following paragraph that the input data is normalized, which is sufficient.

We agree that this statement was redundant and not optimally placed. The sentence will be removed from this paragraph, and the description of data normalization will be retained in the subsequent paragraph where it is more appropriately discussed.

Terminology (e.g., Line 247): The terminology occasionally switches between UAV data, input data, and sensor features, and predictor rasters. More consistent terminology could improve clarity.

We agree that the lack on consistent terminology in the original manuscript could reduce clarity. We plan to revise the text throughout to improve clarity and consistency, using more standardized terms to distinguish between raw UAV-derived data, processed sensor features, and model input variables.

Figure 6: The resolution of the plots could be improved, as the text is difficult to read. In the last column, green dots overlap with the text labels, which further reduces readability. It might help to bring the text in front of the plotted points.

We will try to improve the overall resolution and visual quality of the figures to enhance readability. In addition, we plan to reorganize the figure content to improve clarity. The scatter plots, which provide detailed insight into feature behavior and bias across the growing season, will be moved to the Appendix and referenced in the main text. This will reduce visual complexity in the main figures, allowing the bar plots to more clearly present the key results.

The overlap between data points and text labels occurs only in a limited number of cases and does not significantly hinder interpretation. However, if deemed necessary, we will adjust the plotting order to ensure that text labels are rendered above the data points.

Line 438-439: The text suggests that LiDAR may be more beneficial than other data types?

We agree that the current phrasing may imply that LiDAR is universally more beneficial, which was not our intention. Our aim was to highlight a potential advantage of LiDAR-derived structural features in terms of model transferability, as they may be less sensitive to variations in illumination conditions and background reflectance compared to passive multispectral data. However, we acknowledge that this is not definitively established and depends on several factors, including crop type, canopy structure (row and crop spacing), sensor configuration, and flight parameters.

We will revise the manuscript to soften this statement and clarify that this represents a possible advantage that warrants further investigation, rather than a confirmed outcome. We will also emphasize that multispectral and thermal data provide complementary information, particularly for capturing physiological processes, and remain important depending on the application.