Response to Reviewers

We sincerely thank both reviewers for their thoughtful, detailed, and constructive comments. We greatly appreciate the time and effort devoted to evaluating our manuscript and believe that the suggestions have substantially improved its scientific quality, clarity, and overall presentation. The critical feedback prompted us to carefully re-evaluate many aspects of the study, resulting in extensive revisions throughout the manuscript.

In response to the reviewers' comments, the manuscript has undergone substantial revision, with the Abstract, Introduction, Results, Discussion, and several Methods sections extensively rewritten to provide a clearer and more balanced presentation of the study. Novelty claims and overly strong language have been moderated throughout, and the discussion of LiDAR intensity has been substantially revised to more clearly explain that the normalized intensity measurements used in this study do not represent radiometrically calibrated reflectance. The interpretation of LiDAR intensity has also been tempered to emphasize its role as a complementary wavelength-dependent signal influenced by both canopy structure and vegetation surface properties, while avoiding direct attribution to specific biophysical plant properties without additional radiometric calibration.

To further strengthen the interpretation of LiDAR intensity, additional destructive measurements of plant area index (PAI), leaf area index (LAI), green leaf area index (GLAI), and green canopy fraction have been incorporated, together with comparisons to established canopy profile metrics includingn3DPI. The modelling framework has also been substantially expanded through the addition of repeated stratified 10-fold cross-validation, providing a more robust evaluation alongside the original 70/30 holdout analysis.

Additional analyses have been included to evaluate the influence of scan angle and point density on gap fraction and biomass estimation, while the temporal sensor evaluation has been reorganized and substantially rewritten to improve clarity and interpretation. The nitrogen treatment comparison has been repositioned to emphasize the practical value of spatially continuous UAV observations, and several new figures and appendices have been added to improve transparency, reproducibility, and understanding of the proposed methodology.

Finally, the manuscript has been revised throughout to more clearly emphasize that the artificial neural network (ANN) is employed as a consistent modelling framework for systematically comparing UAV sensor features and feature combinations rather than for developing a universally optimal biomass prediction model. Although several numerical results changed slightly following the additional analyses, the overall conclusions and scientific message of the study remain unchanged while being supported by a substantially stronger methodological and statistical foundation.

In the following pages, each reviewer comment is reproduced in **blue**, followed by our response and a description of the corresponding revisions. **Line numbers refer to the revised manuscript.**

Reviewer 1

Responses in Blue

1. The manuscript repeatedly claims to be the first to integrate LiDAR CH, multi-layer GF, LiDAR intensity, MS, and TIR for biomass estimation. This is not accurate. Multi-sensor fusion combining LiDAR and MS has been done, including in wheat. LiDAR + TIR exists in related cereal studies. The unique contribution here is the specific configuration and systematic comparison, not the existence of the fusion itself. The novelty claim must be rewritten. As written, it will not pass expert scrutiny.

   - Novelty claims have been revised throughout the manuscript. The study is now presented as a systematic evaluation of LiDAR features individually and in combination with multispectral and thermal infrared observations within a common ANN framework, rather than the first example of multi-sensor fusion for biomass estimation.

2. The paper makes strong claims about LiDAR intensity encoding physiological and structural canopy traits. These claims are not sufficiently supported.

   - Claims regarding LiDAR intensity have been moderated throughout the manuscript.
   - The manuscript now explicitly clarifies that normalized LiDAR intensity is not a radiometrically calibrated reflectance measurement.
   - Lines 512-544 Additional comparisons with destructively measured PAI, LAI, GLAI, and green fraction of LAI to provide further evidence that normalized LiDAR intensity provides complementary information beyond purely structural metrics.

3. **The ANN modelling framework lacks statistical rigor**

The modelling approach is a weak point of the paper.

- **Only 86 destructive samples** are available, which is extremely limited for training multi-feature ANN models.
- A simple 70/30 split is insufficient for robust validation; no cross-validation, no repeated sampling, and no uncertainty estimates are provided.
- There is no demonstration that spatial autocorrelation was controlled. The model may be learning subplot-level patterns, not generalizable relationships.
- Feature dimensionality is high relative to sample size, yet model selection is described loosely and without formal procedure.

As it stands, the modelling is not statistically robust enough to support the strong performance claims made throughout the manuscript.

- The manuscript now clearly emphasizes that the ANN is used as a consistent framework for comparing sensor features, not for developing an optimized biomass prediction model.

- Lines 355–495 Repeated stratified 10-fold cross-validation with RMSE, MAE, R^2, and uncertainty estimates has been added throughout the Results (Lines 355–495).

- In the Study area section and Section 4.8 Model Evaluation now provides additional details regarding model selection, validation strategy, and the spatial and temporal distribution of training/testing samples to reduce potential spatial autocorrelation.

4. **Multi-layer GF method needs deeper justification**

The multi-layer GF analysis is a valuable idea, but the implementation and interpretation need more discipline.

- The segmentation thresholds (10/20/30 cm) appear arbitrary.
- No empirical or theoretical justification is given for why these scales represent meaningful canopy stratification.
- The comparison to 3DPI and classical GF methods is underdeveloped.
- The model's sensitivity to point density and scan geometry is not addressed.

The method shows potential, but the manuscript overstates its generality and does not provide enough evidence for the proposed optimal configuration beyond this specific dataset.

- Lines 229 – 233: The rationale for the selected layer depths and grid resolutions has been expanded
- Lines 355 – 372 Comparisons with classical GF and 3DPI have been further developed
- Lines 345-352; Appendix B Figures B3 and B4 Additional analyses evaluating scan angle and point density effects have been added

5. **Sensor dominance over time is overstated**

The temporal analysis suggests shifts in sensor utility across growth stages. While the general trends are plausible, the manuscript repeatedly makes categorical statements (e.g., "MS dominates during senescence," "CH dominates early") without rigorous statistical backing.

These conclusions need to be presented as observations from this dataset, not generalizable statements about sensor behavior

- The temporal evaluation has been substantially revised. Interpretations are now presented as observations from this dataset, growth-stage comparisons have replaced individual datebased discussions, and cross-validation results are incorporated to provide a more balanced assessment (line 495).

6. **Nitrogen treatment analysis draws conclusions not supported by data**

   Figure 12 shows expected spatial smoothing from UAV-based predictions compared to subplot-level destructive samples. This does not prove that UAV "captures management effects more effectively." It only shows that UAV sampling is spatially denser.

   The manuscript conflates spatial resolution advantages with biological sensitivity. This needs correction.

   - The nitrogen treatment analysis has been moved from the Results to the Discussion (Lines 610–620) and is now used solely to illustrate the practical value of spatially complete UAV biomass estimates, rather than enhanced biological sensitivity.

Other Critical Points

7. The abstract uses promotional language ("breakthrough," "promising tool") that is not supported by the analysis.

   - The Abstract has been completely revised with promotional language and overly strong claims removed and replaced with more balanced interpretations.

8. Figures are overly dense, especially Figures 7–9; interpretation is difficult.

   - Detailed figures have been moved to Appendix B while retaining the main findings in the primary manuscript to improve readability and transparency.

9. No error bars or confidence intervals are provided anywhere, this weakens all conclusions.

   - Repeated stratified 10-fold cross-validation has been added and all corresponding summary figures now include error bars representing variability among folds.

10. The discussion section extrapolates beyond the evidence, particularly regarding the physiological meaning of intensity and the claimed operational advantages.

    - Lines 573 – 586 The Discussion has been revised to moderate the interpretation of LiDAR intensity and operational implications. Additional evidence from the PAI, LAI, GLAI, and green fraction analyses has been incorporated to support the complementary role of normalized LiDAR intensity relative to structural metrics.

Reviewer 2

Responses in Blue

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.

    - Novelty claims have been revised throughout the manuscript. The study is now presented as a systematic evaluation of LiDAR-derived canopy height (CH), multi-layer gap fraction (GF), and normalized intensity (INT), individually and in combination with multispectral (MS) and thermal infrared (TIR) observations within a common ANN framework, rather than the first example of multi-sensor fusion for biomass estimation.

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.

    - This point has been expanded throughout the Results and Discussion. The revised manuscript now emphasizes that although multi-sensor fusion provides measurable improvements, these gains are often modest and should be considered relative to the additional sensor, acquisition, processing, and operational complexity (Sections 6.3 and 6.4; Lines 600 and 625–630).
    -

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.

    - Additional details describing the TIR acquisition protocol were added

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.
    - Revised as suggested.

Minor Comments

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

- Corrected in the revised manuscript.

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

- Corrected in the revised manuscript.

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

- Corrected the flight altitude for all to 50m

8. Lines 226: CWSI: Please explain this abbreviation.

- Added further explanation in the appendix

9. 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.

- The suggested sentence has been removed, and the normalization procedure is now described only once in the appropriate Methods section.

10. 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.

- Terminology has been revised throughout the manuscript to consistently distinguish between UAV observations, sensor features, and predictor variables.

11. 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.

- Figures moved

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

- The discussion has been revised to provide a more balanced interpretation.