Response to review

Assessing the effect of forest management on above-ground carbon stock by remote sensing

Dear editor and reviewers,

We would like to thank the editor and reviewers for their thorough review and constructive feedback. We greatly appreciate the recognition of the study's relevance and the importance of a multi-sensor approach for forest biomass carbon assessments. The comments have provided valuable insights, and we have carefully addressed the key points raised by referee 1 by elaborating on the remote sensing preprocessing and validation. Moreover, additional analyses to address possible extrapolation issues were performed as rightfully suggested by referee 2. We are confident that the implemented comments improved the quality of the manuscript, and answered to the specific comments point by point below.

Therefore, we have copied the comments in black, and we use blue text for the response. We quote text from the revised manuscript in green and we refer to line numbers in the clean revised version. Additionally, we have uploaded a supplementary version of the manuscript with highlighted track changes that indicate where the manuscript has changed. However, figure changes were not kept with track changes. Figure 2, 3 and 5 were adapted to implement comments R1.2 and R2.2 respectively. Figure 4 was split into figure 4 and figure 8 (added) (see R1.1) and figure 6 was moved from the appendix to the main text. Additionally, figure 7 and figure A1 were added (see R1.6 and general comment referee 2).

We sincerely hope the editor's and referees' comments have been addressed as intended and thank the editor for reconsidering the revised manuscript for publication.

Yours sincerely,

Sofie Van Winkel and co-authors

Editor comments

Two reviewers have reviewed the article. Both reviewers see the potential in the manuscript, particularly because it highlights the potential of multi-sensor approaches, which have been underutilized and will become increasingly important as the range of different types of data increases. The manuscript also addresses two key issues, namely the potential and need for biomass monitoring and aspects related to silvicultural practices and management. However, both the reviewers and I feel that there are several points that need to be further elaborated to ensure the robustness and replicability of the analyses and to verify some of the conclusions presented in the manuscript.

It is clear from your response that many of these points will be addressed in the revised manuscript, and therefore I suggest that you resubmit a version of the manuscript after major revisions.

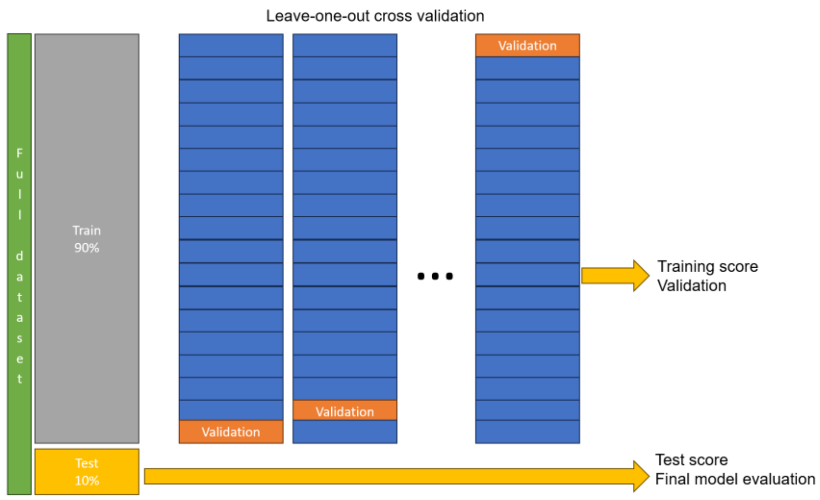
<u>Specific comments</u>

Lines 14-15 Please specify the unit in RMSE and MAE.

Agreed & adjusted to tons/ha.

It is not entirely clear whether the model evaluation (called in the manuscript LOOVC) is done iteratively or just once, keeping 90% of the data set for training and 10% for evaluation. If this is the case, it is not a leave-one-out cross-validation but a training evaluation split approach. However, I would suggest using iterative cross-validation with a k-fold (k = 10) approach.

LOOCV was used for validation at intermediary stages of the model training, on the 90% training data only (e.g. while testing for oversaturation, detection of small trees and the inclusion of Sentinel-1 indices and/or canopy height estimates). As indicated in figure 3 in the manuscript, and the figure below, the 10% test data was isolated until the end, where it was used for the final model evaluation (in this regard thus a training evaluation split approach). We clarified this in the methods section and the caption of figure 3.



Referee 1.

General Comments

The manuscript entitled "Assessing the effect of forest management on above-ground carbon stock by remote sensing" evaluates a multi-sensor approach for forest above-ground biomass/carbon prediction and its usability for assessing the difference in biomass carbon content between managed and unmanaged forests. Both of these issues are highly topical. The potential of multi-sensor approaches is still underutilized and will become increasingly important as the range of different types of data grows. The requirements for monitoring biomass in managed and unmanaged forests is increasing in importance due to the new European regulations aiming to increase close-to-nature silvicultural practices with a simultaneous increase in monitoring requirements.

The study site and the field reference dataset are unfortunately rather small and limited to one specific ecosystem. This limits the usability of the results of this study. Nevertheless, I think the findings of this study would be a valuable resource for other researchers interested in similar topics in other areas. The manuscript is very concise. Although I generally like short and concise manuscripts, there are several points that need to be elaborated more to ensure replicability of the analyses and to verify some of the conclusions presented in the manuscript. I have itemized the key points in the 'specific comments' below. In addition, I have some minor suggestions or comments (listed in the 'technical comments' section) for the authors to consider.

<u>Specific Comments</u>

R1.1. L8: For the clarity of the text, it might be good to consider keeping the two main topics (i.e. multi-sensor biomass prediction and assessment of managed vs. unmanaged forest) in the same order as they have been presented here throughout the manuscript (at least in the Results and Discussion sections). In my mind this would be the logical order and it would make it easier to follow the manuscript. Particularly Figure 4 is now a bit confusing since the reader does not yet know at that point what data/model has been used for the final remote sensing prediction.

Following the suggestions of the reviewer, the two research questions were now reversed in the abstract and introduction, following the structure of the paper.

L8-12: '1) assess the effect of forest management on carbon stock by comparing unconfounded pairs of managed and unmanaged forests in the National Park Brabantse Wouden (Flanders, Belgium), and 2) leverage the complementary strengths of optical, Light Detection and Ranging (LiDAR) and Synthetic Aperture Radar (SAR) remote sensing technologies to improve overall accuracy and scalability in carbon stock estimation.'

Figure 4 was split, and the figures representing the remote sensing predictions were moved to the end of section 3.2 (now figure 8).

R1.2.L12: Here and throughout the manuscript, I think it would be important to make it clearer that you did not use GEDI data as such, but a product that is based on a combination of GEDI data and Sentinel-2 imagery. This is significant because the Sentinel-2 data undoubtedly induces some saturation in the height predictions, potentially with varying effects in different areas. I would say already here something like "…derived from a combination of the Global Ecosystem Dynamics Investigation (GEDI) and Sentinel-2 data were used…".

The manuscript was checked to adjust the respective terminology to 'a canopy height product derived from GEDI and Sentinel-2' or 'the GEDI/Sentinel-2 (canopy height) product'.

R1.3.L59: This sentence is confusing. Spaceborne LiDAR (e.g. GEDI with 25-30 m footprint) does not provide information on higher [spatial?] resolution than passive optical sensors. Also, I would not say that spaceborne LiDAR provides detailed 3D profile of forest canopies, knowing the noisiness of the observations and other complications in deriving the height predictions for the footprints. The sentence would be true for airborne LiDAR, but I think it is an overstatement for spaceborne LiDAR and gives a wrong impression to those readers who are not familiar with GEDI data. I would rather say something like it is possible to derive predictions of the canopy profile for the 25 m footprints.

Agreed. This sentence was replaced by a more nuanced alternative:

L60-62: 'Light Detection and Ranging (LiDAR) technology is an active remote sensing method that enables predictions of the canopy profile, and from there biomass, within the sensor's footprint'

[There is some problem with the line numbers at pages 6-7.]

Regarding the Sentinel-2 data description, it would be important to elaborate further two points:

R1.4.1) What is the cloud probability product/layer used in the masking? To my understanding it is not part of the regular S2 L2A products, right? This product/layer should be somehow described, with reference if relevant.

The cloud probability band is part of the standard Sentinel-2 L2A output (https://documentation.dataspace.copernicus.eu/APIs/SentinelHub/Data/S2L2A.html) and openly available on Google Earth Engine. An appropriate reference and explanation was added:

L152-155: 'A cloud masking was performed (filtered with a 60% cloudy pixel percentage, masked with a 40% cloud probability threshold) and all 20-meter resolution bands were resampled to 10 m resolution. The cloud probability layer of Sentinel-2, openly available in Google Earth Engine, was used (Google Earth Engine, 2015).'

R1.5.2) The derivation of band values for plots is unclear to me. Did you take all pixels that had min 90% overlap with the plot and take their average? Or did you calculate weighted average, weighted by the area of plot overlap?

We calculated the weighted average, weighted by the area of plot overlap. This was clarified in the manuscript:

L155-157: 'Weighted average band values were then calculated for each selected plot, weighted by the percentage of plot overlap. A lower limit of 90% overlap between the pixel and the plot area was used.'

R1.6.L166: Regarding the GEDI/Sentinel-2 canopy height product, it would be very important to have some understanding on the accuracy of the product in the study area. Did you do any evaluation how well the product seems to work in the study area? Would FFI plots be available for this analysis? Or you could use your own plots. This is a crucial issue for the interpretation of the results. It is possible that there are significant saturation effects in the product in your study area due to the Sentinel-2 data.

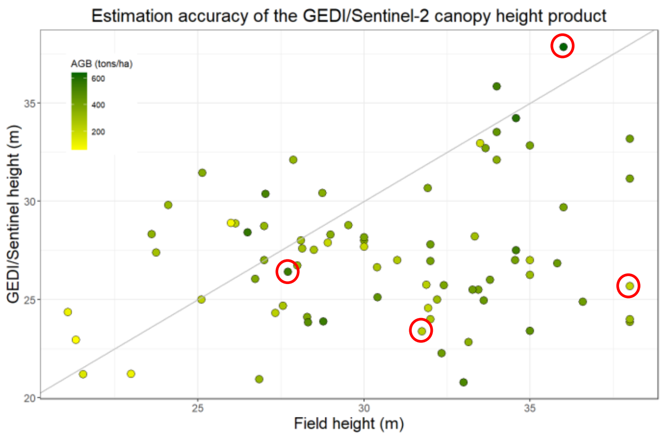


Figure 7 in the manuscript: Comparison of the canopy height estimates by the GEDI/Sentinel-2 product and the field measurements. Highlighted in red are some plots that refute the premised saturation effect.

We agree that this is an important point to consider. A comparative figure (Figure 7) was therefore added in the manuscript, showing a systematic underestimation of the canopy height by the GEDI/Sentinel-2 canopy height product (points below the 1:1 bisector). However, plots with a higher biomass are not systematically characterized by a larger underestimation (dots with a darker color can be close or far from the bisector of perfect prediction), so no saturation effect is detected. Some examples are highlighted in the figure below. It is therefore difficult to decide the origin of this underestimation, but we can see a similar underestimation with other remote sensing products in the area (e.g. CCI biomass). Nevertheless, the same canopy height differences between plots are generally shown by the GEDI/Sentinel-2 canopy height product and the field measurements. This additional analysis was added to the manuscript (L258-262).

R1.7.L167: The Sentinel-1 data needs to be described in much more detail for replicability of the study. Where did you get it from? Which data products you started the processing from? What kind of preprocessing steps were or had been done (by you or others)? How was the temporal compositing done? Details of the preprocessing are very important for radar datasets as they strongly affect the end product characteristics. If the text length is an issue, I would rather have the data preprocessing details here, rather than general description of C-band characteristics.

Additional preprocessing specifications were added to the methodology section, elaborating on the initial Sentinel-1 product and processing steps:

L178-183: *'Data were acquired in Google Earth Engine at level-1 Ground Range Detected (GRD) in Interferometric Wide Swath mode (10m resolution) with dual polarization (VV and VH) for both ascending and descending passes. Data was collected during the same time period as the Sentinel-2 data. Orthorectification, thermal noise removal, radiometric calibration and border noise removal were already conducted in this GRD product. A refined Lee speckle filter was then applied to the data for speckle reduction, and the temporal median was taken for every pixel. Mean plot values for VV and VH were finally calculated separately to serve as explanatory variables, similar to the Sentinel-2 processing.'*

R1.8.L194: Do I understand correctly that you used only eight plots to derive the error metrics presented in the results section (e.g. Table 4)? If so, this is a very small number of plots. Why did you not use e.g. 1/3 vs. 2/3 split (i.e. 33% for validation)? How where the eight plots selected? Do they include plots from all forest size classes? Did you test how sensitive the error metrics are to each of the plots (i.e. how the metrics change if you leave out one of the plots, one by one)?

This small selection (10% of the data) was chosen as a test dataset to maximize the training group. As noted before, the total dataset is rather small (78 plots), and maximizing this training dataset was thus necessary to maximize model performance. To avoid overfitting, cross-validation was applied at intermediary stages (Figure 3). The 8 plots in the test dataset were randomly selected, but included plots from all forest sizes, from 196 tons/ha to 581 tons/ha biomass. A maximal loss of 6.7% was noted when iteratively leaving out one of the test plots. The test metrics fluctuated around the final reported values and are not influenced majorly. An additional short explanation, including these important points, was added to the manuscript (L271-272).

R1.9.L232: It would be nice to see a bit more information on the results of the variable selection, to understand better how the variable selection progressed and how clearly the five chosen ones finally provided the best results (e.g. compared to using only spectral bands without indices).

Additional information on the feature selection was provided in the methods and results section of the revised manuscript, enhancing the reproducibility of the results.

L245-252: *'The Sentinel-2 variables selected through recursive feature elimination for the Generalized Additive Model (GAM) capture various photosynthetic and structural characteristics of vegetation. The first selection of features included several multicollinear features, such as B5 and B6, GNDVI and NDI45, STVI3 and STVI2. The first 5 features that were not highly correlated were selected. These included: B5, B12, GNDVI, STVI3, and MCARI. The inclusion of indices compared to only spectral bands enhanced the interpretation. The red-edge wavelengths were represented, which help detect vegetation density and type. The short-wave infrared wavelengths, along with GNDVI and MCARI, provide insights into photosynthetic capacity and chlorophyll absorption depth. Lastly, near- and mid-infrared bands were included in the stress-related vegetation index (STVI3).'*

R1.10.L241: I am very surprised that the addition of canopy height hardly improves the results. Usually, knowledge of canopy height is very beneficial for biomass estimation. This makes me wonder how accurate and useful the GEDI/S2 height predictions are in the study area. It would be important to validate the height product in the study area to better understand its effects and usability. On the other hand, the addition of S1 has a very clear and significant positive effect. Therefore, I would also like to see the S2+S1 results. They may be nearly on the same level as S2+GEDI/S2+S1. This affects also the discussion section (L285). It would be important to understand the role of the canopy product better so that it could also be clarified in the discussion.

Indeed, the model including S2 and S1 features results in very similar error metrics compared to the model with S2, S1 and GEDI/S2. Even though the GEDI/S-2 product did detect the correct canopy height trends, there was a systematic underestimation (see comment R1.6 and figure 7, added in the manuscript). This analysis and corresponding section in the discussion were added/adjusted (L257-260, L303-308).

Technical Corrections

R1.11.L32: Suggestion, perhaps change the last sentence into something like: "However, accurately capturing carbon stocks over large areas presents both technical and logistical challenges. In this context, remote sensing provides cost-efficient means for large scale monitoring of above-ground carbon in forests."

Agreed & adjusted

R1.12.L51: Suggestion: "This information can be extrapolated using…"

Agreed & adjusted

R1.13.L53: Comment: This is true for wall-to-wall mapping, which is why sampling has been traditionally used in forest inventories. However, the increasing requirements of spatially explicit information call for new approaches.

We agree with the stated comment, but the two written sentences at L53 address exactly that. We are therefore unsure what the comment is referring to, and consider it solved.

L53: *'While such in-situ methods achieve high accuracy at small extents, it becomes costly and labor-intensive when scaling to larger regions. Spaceborne remote sensing technologies have been widely adopted to expand the reach and efficiency of biomass estimation'*

R1.14.L56: Suggestion: Instead of -range, I would prefer to use either -area or -scale.

Agreed & adjusted

R1.15.L58: Comment: I tend to disagree with this to some extent. NDVI (and numerous other vegetation indices) also tell about the vigorousity of the vegetation, not necessarily about AGB. Also, there is great variation between tree species on the NDVI (for similar AGB).

We agree that the relationship between NDVI and AGB is indeed not straight-forward and the predictive success fluctuates between studies. We therefore find that NDVI is not the best vegetation index to indicate AGB. However, it is here only mentioned as a popular and best-known indicator of AGB. The section was nuanced to support this idea and relevant references were added:

L58-60: 'Vegetation indices, such as the Normalized Difference Vegetation Index (NDVI), indicate the photosynthetic activity and health of trees, and are often-used indicators of biomass (Askar et al., 2018; Laurin et al., 2018).'

R1.16.L70: Comment: When talking about Spaceborne LiDAR, they also suffer from "mixed observations" as the footprint is rather large.

We agree and we rephrased this to an overall limitation of remote sensing:

L77: *'Lastly, all three sensor types may suffer from mixed pixels when a single pixel captures multiple surface types and complicates accurate biomass estimation.'*

R1.17.L75: Suggestion: "..., because they do not provide wall-to-wall data."

Agreed & adjusted

R1.18.Figure 1: Suggestion: Change "3 plots" to "Three plots".

Agreed & adjusted

R1.19.L110: Comment: This sounds like a very short time period for me. Is this sufficient time to reach characteristics of unmanaged forest in Belgium?

In our study design, 20 years is the lower limit and many clusters also include unmanaged patches of 40 to 45 years of non-management. A time period of 20 years often includes already two or more thinning cycles in managed forests, which allows a noticeable difference in the field, which is indeed what we saw. This study shows that even though the difference is there, detection with remote sensing is more complicated. This may be completely different in forests that have been set aside for much longer, but they cannot be found in the study region. Therefore, the results of this study are not to be generalized to all unmanaged forests, including primary forests. The time horizon of 20-40 years is most relevant in the context of international climate goals (such as the Green Deal) and relates to the question: 'If we now leave this forest unmanaged, how will the climate mitigation potential evolve over the next decades?'. This was included in the discussion (L356-361).

R1.20.L113: Suggestion: IMFP needs to be written out somewhere. I did not find it anywhere in the manuscript.

This was indeed erased by mistake. IFMP stands for 'Informa Management Platform', which is now written out when first mentioned in L92.

R1.21.L119: Comment: Why were dead trees included? Depending on the case, this may badly confuse remote sensing based biomass/carbon mapping.

We agree that the inclusion of dead trees may badly influence the remote sensing model, as they are not detected by remote sensing and do not sequester carbon anymore. We saw that the dead trees included in the field dataset only composed 1% of the total biomass, and the adjusted model resulted in similar performance metrics. This was adjusted in the manuscript:

L113-114: *'The system boundaries were defined as standing above-ground biomass, because below-ground biomass or deadwood cannot be easily quantified by remote sensing.'*

R1.22.L132: Comment: I wonder if this is well established use of the term "two-entry tariffs". This is a totally new term for me. Please check.

This term refers to double-entry volume equations that are used to calculate stem volume from DBH and tree height (two input parameters). We replaced the term with 'double-entry volume equation', which may be more common (L125, 130, 132).

R1.23. [There is some problem with the line numbers at pages 6-7.]

It seems indeed that a problem occurred with the line numbering when converting the document to pdf due to an internal bug in Word. It was now solved.

R1.24.P6.fifth last line: Comment: And 8A is also not used. Perhaps it is not even available in the GEE data collection?

Band 8 and 8A are closely related and are both situated in the near infrared spectrum. Band 8A is a very narrow band, and B8 operates at a higher resolution. Only B8 was included to avoid data redundancy. This was added in the manuscript:

L150*: 'Band 8A was also left out given its close relation to B8, but with a lower resolution and narrower band width, to avoid data redundancy.'*

R1.25.L153: Suggestion: I would remove 'calibration', saying 'Once the field measurements were obtained...'

Agreed & adjusted

R1.26.L159: This sounds a bit simplistic in my mind. There are a lot of complications deriving forest canopy height profile from the GEDI data. I would rather formulate it somehow so that "prediction of canopy height profile for the 25 m footprint area can be derived from the GEDI observations".

Agreed & adjusted

R1.27.L214: Suggestion: Perhaps "Model application" instead of "Model extrapolation".

Agreed & adjusted

R1.28.Table 3: Comment: Might be more informative to use the plot size or the tree sizes, rather than A, B, C in the caption and the table itself?

The plot radius was added in the table caption as a reminder, and the tree sizes were mentioned with the DBH in the table:

Table 3: Overview of the plot characteristics in managed and unmanaged forests, measured in the field. The different levels (A,B,C) refer to the nesting levels as defined in Table 1. Plot radius level A= 2.5 m, level B= 9 m, level C= 18 m.

	Managed	Unmanaged
Nr of plots	39	39
Nr of trees	1348	884
Species richness	18	19
Mean DBH (cm)		
level A (<7)	2	3
level B (7-39)	16	21
level C (>39)	58	59
Mean height (m)		
level A	4	4
level B	14	18
level C	30	31
Mean density (stems/ha)		
level A	2717	580
level B	323	322
level C	89	104

R1.29.L237: Comment: Why so short text here and most of the text in the appendix? If not limited by manuscript length rules, I would not mind seeing more results on the modelling here in the main text.

As the results of the model optimization were not significant and do not contribute to the main study objectives, we chose to put these additional figures in the appendix. This was decided to keep the flow of the manuscript relevant to the subject. We therefore leave this to the editor's decision.

R1.30.L244: Comment: Again, perhaps application would be better than extrapolation?

Agreed & adjusted

R1.31.L290: Suggestion: You could expand the discussion on the benefits of longer L-band wavelength a bit e.g. by citing some more recent ESA CCI Biomass related references by Santoro et al., where they use combination of C and L band data.

Adjusted:

L317-325: *'C-band radars are more sensitive to detecting leaves and needles than trunks and branches, in contrast to P- and L-band SAR (Rüetschi et al., 2018). The shorter wavelength interacts more strongly with smaller vegetation elements with a higher water content. Possibilities to improve predictions even more may lie in further integration of C-band with L-band SAR, which can enhance the detection of texture features, vegetation diversity, and density (Laurin et al., 2018). For example, Santoro et al.(2021) successfully estimated AGB at a global scale from the ALOS satellite (L-band) and Envisat (C-band). They report the combination of different sensors with varying spatial resolution as an important challenge, and a source of systematic modeling errors at the regional level. This has however been a popular approach in the last years, with the CCI Biomass product as an important example. L-band SAR is generally more suited in high-biomass areas, as these longer wavelengths can penetrate deeper into the canopy than C-band wavelengths. Santoro et al. (2021) therefore implement a weighting scheme, where a different sensitivity of backscatter data from both sensors is applied depending on the growing stock volume. The successful implementation of multifrequency SAR data in remote sensing-based analyses of AGB is also shown in other studies, such as Huang et al. (2018) and Musthafa and Singh (2022).'*

R1.32.L291-L304: Comment: This is a good discussion paragraph. Indeed, the remote sensing predictions typically gravitate towards the average, causing underestimation of high volume forests and overestimation of low volume forests.

We thank the referee for the appreciative comment; this is indeed a trend we see recurring in many remote sensing-based biomass studies.

R1.33.L301: Suggestion: Should be overestimation, not underestimation, right?

Indeed, this was adjusted.

R1.34.L314: Suggestion: "as" should be replaced with "was".

Agreed & adjusted

R1.35.L332: Suggestion: Table A1, not Table 5.

Agreed & adjusted

Referee 2

General Comments

The study consists in building a model of forest aboveground biomass that combines data sources from optical, LiDAR, and Synthetic Aperture Radar (SAR) in order to estimate forest biomass and carbon stocks at a 10 x 10 m resolution over a small-sized study area (a national park in Flanders, Belgium). The forest biomass map produced was used to evaluate the effects of forest management on carbon stocks by comparing matched pairs of managed and unmanaged sites. The model used data from Sentinel-1, Sentinel-2, and a canopy height product (Lang et al., 2022, 2023) derived from the Global Ecosystem Dynamics Investigation (GEDI) mission as predictors in a generalized additive model (GAM).

Field assessments revealed higher carbon stocks in unmanaged stands compared to managed ones while the model-based estimations did not result in significant differences. The signal saturation and the need for additional training data are among the possible reasons for this discrepancy between direct but point-wise inventories and the predicted carbon stocks. The main reasons for this discrepancy are: i) the potential saturation of the remote-sensed products at high biomass stocks and ii) calibration issues of the GAM, iii) mean absolute errors larger than the difference between managed and unmanaged forest stocks.

The study brings two contributions: one is the use of remote-sensed data trained by field data in order to assess the forest aerial biomass over a given region, the second is to use this information in order to detect differences among forest management types.

There is a large investment of the community into using remote-sensed data to produce forest and/or volume biomass estimations, with a wealth of publications in this topic. The methods used here are not new, although the use of a GAM model is not as common as the use of a non-parametric model approach. Thus overall, the level of novelty is not very high on this aspect.

The use of the model to estimate the forest biomass over a given area to test for management effects is more novel. It has limitations but also has the merit of highlighting current challenges : i) the use of remote-sensed data to improve local estimations of forest biomass is not an easy task, ii) the gains are strongly limited by the amount of field data available and iii) the differences in forest biomass related to the management is sufficiently nuanced to be difficult to determine.

By making additional efforts into quantifying the impact of the reduced field-data sample size on the model development, and associated problems of extrapolation, the study would be much more conclusive.

Specific comments

The reduced sample size (39 per forest category, managed/unmanaged), probably too small for the calibration of the model, may be evoked as an overarching limitation to the study. Fitting complex models certainly requires a sufficiently large observational basis, unless the population studied is very homogeneous. Here, the population is probably not really homogeneous, as the range of carbon stock values seems quite large. While this does not invalidate the study, it certainly adds uncertainty to the results.

The assessment of predictions uncertainties would have brought some light and helped the diagnosis. In particular, splitting extrapolation issues from saturation issues would have been very helpful. Extrapolation issues highlight potential deficits in field sampling. The use of a convex hull to identify the proportion of extrapolated predictions (see ex. https://doi.org/10.1016/j.jag.2022.102939) would be one efficient way to respond to this. The question currently remains open, that is, the study is not very conclusive in this regard. That is a major concern.

To address these concerns, we explored an additional convex hull approach as suggested. This analysis is added in the appendix A2, and referred to in the methodology section (L228-229) and the discussion (L348-L354):

L392-L402: '*Extrapolation errors may arise as a consequence of the rather small training dataset (N=70) used for model training, inducing unreliable predictions. To address this concern, the feature space delineated by the training database is calculated as a convex hull, following the approach of Renaud et al. (2022). By assessing how many of the pixels for model application are situated in this space, the percentage of extrapolation can be assessed. However, due to the high number of features (7), and the resulting extremely narrow high-dimensional convex hull, a very large proportion of extrapolated pixels (98%) was identified. Nevertheless, the training dataset provided a comprehensive representation of the total value range for each 2D combination of variables separately (Fig. A1). Values that only slightly differ from the training data variable range fall strictly outside the hull, though they are in this case still believed to be in line with the training data. Even though labelled as 'extrapolated pixels', underestimated biomasses are in this case probably due to signal saturation rather than extrapolation errors. When allowing pixel values to deviate 5% of the hull range for each variable, the percentage of extrapolated pixels decreases to 28% (27% without extreme outliers), underpinning this statement.*'

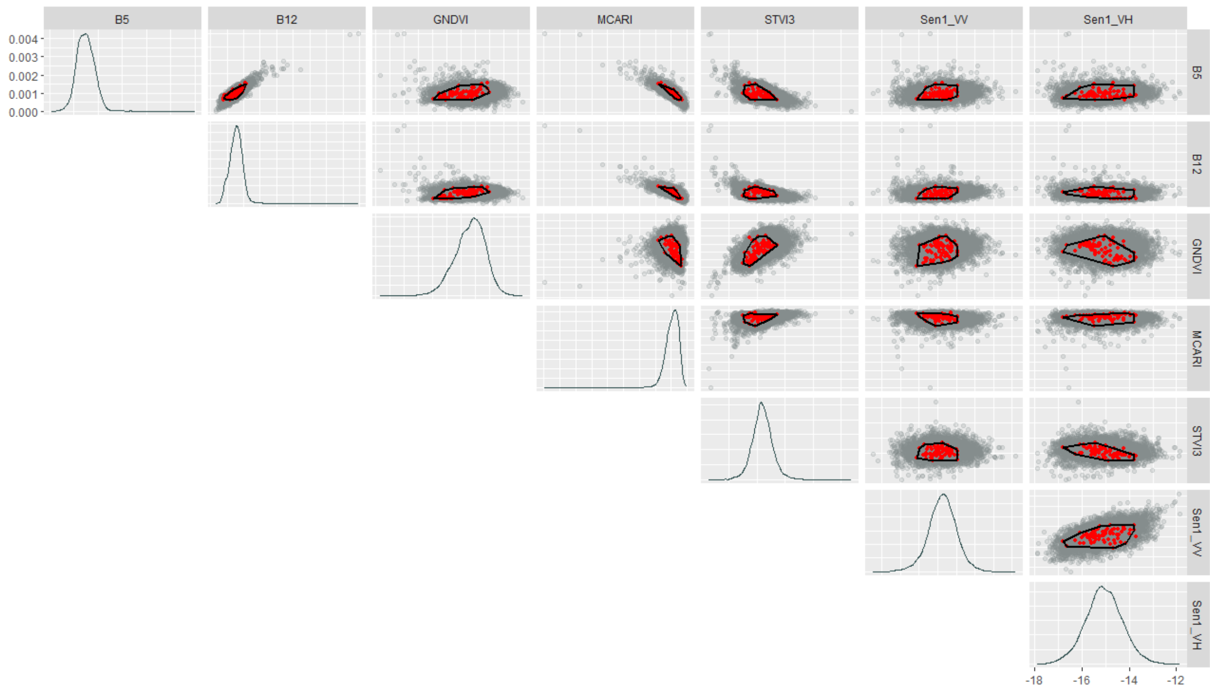


Figure A1: A 2-dimensional representation of the 7-dimensional convex hull (black) on the training data (red). A subset of 20.000 pixels used for model application are plotted in grey.

The way the estimations at the scale of the patches were made was not sufficiently well explained. It seems that the estimations of the biomass at patches level represent an average over the estimations made at the pixel level inside. They would therefore be purely model-based estimations. Besides the mean estimation error, the bias had to be estimated too. It seems to be quite substantial, as indicated by figure 5. The bias relates of course to the signal saturation. Signal saturation is indeed another known problem, for which the integration of multiple sensors was hoped to bring a solution here. The analyses on the model predictions do not allow to attribute these problems to a lack of data for fitting, or for a saturation issue per say. It therefore remains largely inconclusive.

The patch-level estimations were clarified in the methodology section of the manuscript (L222-225). Additionally, we quantified the bias at forest patch level, comparing field measured carbon stock with model-based estimations, which was found to be -0.83 tons/ha. As noted in the manuscript, this bias arises both due to a tendency towards overestimation at low biomass levels and underestimation at high biomass levels (oversaturation errors). However, these opposing effects largely balance out, resulting in an overall bias close to zero. This was added to the results section (L274-275).

Finally several uncertainty metrics were included in the results section, including the Mean Absolute Error, the Root Mean Square Error, the estimation bias and the standard error of the mean.

Technical comments

R2.1 This affirmation is false: L85: “The combined use of LiDAR, SAR and 85 passive optical remote sensing has, to our knowledge, not yet been investigated to assess above-ground biomass in temperate forests.”

It has been used, albeit with different modelling approaches:
https://essd.copernicus.org/articles/15/4927/2023/?utm_source=chatgpt.com

Agreed & adjusted (the sentence was removed)

R2.2 In Figure 3 it would be good to complement the “100%” Field dataset box with the actual number of data available (N = 78).

Agreed & adjusted

R2.3 The description of the GAM is wrong: L187: “A Generalized Additive Model (GAM) was chosen as a non-parametric extension of GLMs (generalized linear models)”

A Generalized Additive Model (GAM) is a semi-parametric, rather than a fully non-parametric extension, of Generalized Linear Models (GLMs). GAM models assume an underlying parametric structure for the response distribution.

Agreed & adjusted

References

Askar, Nuthammachot, N., Phairuang, W., Wicaksono, P., & Sayektiningsih, T. (2018). Estimating Aboveground Biomass on Private Forest Using Sentinel-2 Imagery. *Journal of Sensors*, *2018*(1), 6745629. https://doi.org/10.1155/2018/6745629

Google Earth Engine. (2015). *Sentinel-2: Cloud Probability* | *Earth Engine Data Catalog* [Dataset]. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY

Huang, X., Ziniti, B., Torbick, N., & Ducey, M. J. (2018). Assessment of Forest above Ground Biomass Estimation Using Multi-Temporal C-band Sentinel-1 and Polarimetric L-band PALSAR-2 Data. *Remote Sensing*, *10*(9), Article 9. https://doi.org/10.3390/rs10091424

Laurin, G. V., Balling, J., Corona, P., Mattioli, W., Papale, D., Puletti, N., Rizzo, M., Truckenbrodt, J., & Urban, M. (2018). Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *Journal of Applied Remote Sensing*, *12*(1), 016008. https://doi.org/10.1117/1.JRS.12.016008

Musthafa, M., & Singh, G. (2022). Improving Forest Above-Ground Biomass Retrieval Using Multi-Sensor L- and C- Band SAR Data and Multi-Temporal Spaceborne LiDAR Data. *Frontiers in Forests and Global Change*, *5*. https://doi.org/10.3389/ffgc.2022.822704

Renaud, J. P., Sagar, A., Barbillon, P., Bouriaud, O., Deleuze, C., & Vega, C. (2022). Characterizing the calibration domain of remote sensing models using convex hulls. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102939. https://doi.org/10.1016/j.jag.2022.102939

Rüetschi, M., Schaepman, M. E., & Small, D. (2018). Using Multitemporal Sentinel-1 C-band Backscatter to Monitor Phenology and Classify Deciduous and Coniferous Forests in Northern Switzerland. *Remote Sensing*, *10*(1), Article 1. https://doi.org/10.3390/rs10010055

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D. M. A., Avitabile, V., Araza, A., de Bruin, S., Herold, M., Quegan, S., Rodríguez-Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Moreno Martínez, Á., Cavlovic, J., Cazzolla Gatti, R., … Willcock, S. (2021). The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data*, *13*(8), 3927–3950. https://doi.org/10.5194/essd-13-3927-2021