Response Letter

Dear Dr. Stocker,

Thank you for your time and effort in dealing with our manuscript "*Assessing forest biomass and productivity with data-driven vegetation indices: insights from 900,000 simulated forest stands*". We appreciate the opportunity to submit a revised version of our manuscript for re-evaluation and would like to thank the you and the reviewers for your constructive comments and helpful suggestions.

We have thoroughly considered and responded to all comments and addressed them by changing the manuscript accordingly. In particular,

- We have rewritten the entire introduction, focusing on improving the clarity of the paper and its motivation.
- We have made the paper's limitations more transparent by adding a dedicated "Limitations" section to the discussion and by stressing the "in-silicio" nature of our work at several places throughout the manuscript.
- We have added a section focusing on recommendations for validating and applying our results based on field data.

In addition, we have made several smaller changes to improve the accessibility of the manuscript and fix minor issues.

Below we respond to each of your and the reviewers' points in greater detail.

We hope that our responses and improvements of the manuscript are suitable to dispel remaining concerns and qualify our manuscript for publication in Biogeosciences.

Yours sincerely,

Samuel Fischer on behalf of the authors

Dear Dr. Fischer,

Thank you very much for submitting your manuscript "Assessing forest properties with data-driven vegetation indices: insights from 900,000 forest stands" to Biogeosciences. The reviewers and I appreciate the effort you have invested in this study and recognize its potential contribution.

Both reviewers, especially reviewer 1, however, have raised important concerns that require substantial revision before we can move forward. At this stage, the decision is to invite Major Revisions.

In particular, two points raised by reviewer 1 seem to be critical and I advise you to address the concerns raised more fully than proposed in your reply.

1. The discussion should address the aspect that all investigations were made *in silico*. In this model world, the link between LAI and GPP and NPP may be different from how these quantities relate in real forests. It should be discussed to what extent the findings presented here are subject to this caveat. This point was made also by Reviewer 2 (”Discussion: I believe this analysis would benefit immensely from (even a very small) application to real-world data.”).

- Thank you, this is an important point. Forest models have a long and very successful history in forest research (see e.g. Shugart et al., 2018), and Formind and its ability to model carbon fluxes has been extensively validated at multiple sites both in Germany and internationally. For example, validation studies using Eddy-Covariance data were conducted by Rödig et al. (2017) for a site in Thuringia, Germany, and Holtmann et al., (2021) for the Hohes Holz site. We have added corresponding notes to the main text (lines 70f, 106).
- Nonetheless, all models have limitations, and we have extended the discussion of limitations in the discussion section and introduced a headline for this subsection to make it easier to find (lines 345ff).
- It will be very important to validate our findings with respect to the inference of forest attributes with field data. Unfortunately, the small-scale field data available to us were not suited to conduct the analysis, as autocorrelation in noise and parameter uncertainty often dominate on small scales. Acquiring and analyzing larger-scale data is beyond of what we are able to do within this project. Therefore, we have stressed the theoretical / preparational character of our study in the discussion and dedicated a section to making concrete suggestion regarding future steps considering field data (section 4.4 in the revised manuscript; lines 364ff).

2. The transferability of results to (perhaps structurally the same) forests growing under different environments (not under conditions found at Hohes Holz) is not clear. The paper deals with this point in terms of investigating “uncertainty” in the environment.

However, this is also about *variations* in the environment, in view of typical applications of remote sensing-based assessments which typically aim at mapping a quantity of interest across space, that is, across environments. This point made also by Reviewer 1 (”Methods and Discussion: A lot of effort …”)

- We agree and are thankful for the word “variation”. This term describes more accurately what we considered in this study: situations in which certain parameters are heterogeneously distributed within the dataset. The term “uncertainty”, which we used in our initial submission, was misleading and suggested that we were conducting a sensitivity analysis or similar. We have therefore exchanged the term “uncertainty” with “variation” throughout the entire manuscript. This has also helped us explaining the motivation behind this analysis (lines 55ff).

Now wearing a reviewer's hat, I found three more points critical and ask you to address them in your revisions:
3. I found this a somewhat misleading statement: "we analysed 900000 forest stands". Please clarify in title, abstract, discussion, and conclusion that your study analysed synthetic data.

- We agree and have adjusted the title, abstract, discussion, and conclusion accordingly. Now the title reads “Assessing forest biomass and productivity with data-driven vegetation indices: insights from 900,000 simulated forest stands”.

4. An explicit view to time scales and the potential influence of short-term variations in the environment should be made in the description and interpretation of the target variables. LAI and biomass can be safely assumed that they do not change much over the course of a growing season, but NPP and GPP may vary strongly. Are results valid under short term variations in light, temperature, water availability, atmospheric humidity or are they representative for growing-season aggregates?

- This is a good point. Our analysis refers to yearly totals only, which are of particular interest when considering carbon balances and forest health. We have clarified this by adding the word “yearly” to GPP and NPP throughout the entire manuscript.

5. Are perturbed traits in the radiative transfer model affecting C dynamics in the forest model (FORMIND)?

- No, the traits used in the radiative transfer model are not linked to the carbon dynamics in the forest model. This would be a very interesting idea for a follow-up study. We have clarified this point in the “Limitations” section (lines 350ff).

We ask that you carefully address each of the reviewers’ points in your revision and include a clear, point-by-point response letter. If you disagree with a comment, please explain your reasoning clearly. Highlighting the changes in the revised manuscript will also be very helpful for the review process.

Please aim to resubmit your revised manuscript.

We value your contribution and look forward to receiving your revised work.

With best regards,

Beni Stocker

Response to RC1

This manuscript examines the effectiveness of existing vegetation indices and novel hyperspectral indices for estimating forest properties: biomass, leaf area index (LAI), gross primary production (GPP) and net primary production (NPP). The methods combine individual-based forest and radiative transfer models to simulate many forest stands, their estimated biomass, LAI, GPP, NPP and 250 canopy wavelengths. Results suggest that unique wavelength pairs often offer stronger estimates of NPP, GPP and biomass than common vegetation indices. A data driven approach also offers strong correlations.

In general, I found the applied methodology interesting and compelling. The manuscript offers several valuable insights, for example, Table 2 could serve as a nice reference when considering existing vegetation indices for estimating forest properties. This said, in my opinion, the manuscript also offers significant room for improvement, particularly in its focus and clarity. I apologize in advance for my many comments below, but I hope they will aid the authors in their revisions. Overarching comments are followed by line-by-line comments.

- Dear Colin Bloom, we would like to reiterate our sincere thanks for your thorough and thoughtful review. We appreciate your comments and your efforts to help us improve our manuscript and addressed the issues you raised in the revised manuscript, e.g., by rewriting the introduction. Below we respond to your individual comments and explain how we have adjusted the manuscript. Overall, we put a particular emphasis on presenting the paper's focus more clearly.

Introduction: A systematic and data driven identification of important reflectance wavelengths for forest properties is absolutely warranted. Few conventional vegetation indices have been systematically developed to target important forest parameters like NPP or GPP. This is a compelling argument for conducting this analysis. Hyperspectral data and data driven indices derived from synthetic samples are only one way to investigate the problem. Much of the introduction, however, is taken up by a methodological background (Lines 33 to 74), which attempts to set the stage for methodological novelty. I find this argument less compelling. As mentioned, Henniger et al. (2023) has already applied the novel combined methodology used in the analysis. Tweaking the approach is just an application of the methods. In my opinion, the methodological background could be substantially condensed and placed at the start of the methods. In its place, a clearer introduction to the overarching problem and research question would greatly improve the clarity of the work. Some relevant questions which I think need to be addressed therein: What are NPP, GPP, and biomass and why are they important? Why do we need to monitor these parameters? How have vegetation indices traditionally been developed? Have vegetation indices been

produced for NPP or GPP specifically? What platforms exist (or will exist soon) to monitor changes? I believe more complete answers to these questions will make it abundantly clear why a new systematic approach to vegetation indices is needed.

- Thank you. We have rewritten the introduction completely and revised the story line, putting a strong emphasis on the motivating underlying problem of estimating forest properties from remote sensing data. We also have motivated more clearly why considering the variability in physiological and environmental parameters is needed. At the same time, we have strongly reduced the background on methods.
- It is correct that the general ideas of (1) generating forest stands randomly using a Monte Carlo–type approach and (2) coupling a radiative transfer model with a dynamic forest model are inspired by earlier work. However, we would like to emphasize that our contribution goes beyond merely "tweaking" existing approaches. First, these two approaches had not previously been combined; second, our forest stand generation technique employs a different algorithm than that of Henniger et al. (2023b) and exhibits different—and more rigorously specified—statistical properties. We have therefore formulated the establishment of an analysis pipeline as one major aim of this study.

Methods and Results: Understanding the manuscript in its entirety took several reads and more effort than I think is necessary. While I appreciate and understand the sensitivity testing of the radiative transfer model, this is, in my opinion, secondary to the overall intent of the main manuscript. Moving this portion of the analysis into the Supplement would make it much easier to read the manuscript through and to understand the methods and results.

- It is important to us to make the manuscript as accessible as possible, so thank you for this valuable feedback. Reading your comment, we believe that the motivation for this analysis has not been laid out sufficiently clearly. In our direct response to your comment (on-line), we have already provided an explanation for the analysis. For better reference, we repeat that response further down below.
- We have addressed the issue in two ways:
    1. We have put an additional emphasis on explaining the purpose of the analysis already in the introduction.
    2. We have changed our wording from the term "uncertainty" (often associated with parameter uncertainty and therefore parameter sensitivity) to "variation", which expresses more accurately what we do: showing the potential of data-driven indices and their optimal design (wavelength choice) changes if the considered dataset is heterogeneous in certain aspects. This situation is expected in real-world applications. Hence, addressing heterogeneity appropriately is key to using data-driven vegetation indices in practice.

- *Original response*: The typical purpose of a sensitivity analysis is to understand how strongly a model result changes if the input parameters change. The result is some range of confidence: even if we do not know the input parameter perfectly, the model results will remain valid to a specified degree. However, this is not what we do here. In our analysis, we provide advice on how **the model itself** (not its output) should change in the presence of uncertainty.
  For example, under dry conditions, bare soil reflects the sunlight brightly, under wet conditions less so. Therefore, we may use a different vegetation index if we know the soil is dry than if we are uncertain. The same applies to the other analyzed factors: if we are certain of the chlorophyll or carotenoid content of leaves, we may use different wavelengths than if we are uncertain; if we know we are looking at a dense forest, we may use different wavelengths than when looking at a sparse forest or a forest whose density is variable.
  The wavelength-R^2 map we show in Fig. 2 corresponds to the one case in which all parameters except the forest structure are certain. This is arguably a special case. Hence, to guide the appropriate selection of wavelengths for other applications, we would need to construct a similar figure for the respective expected uncertainty regime, which we do not know a priori. Therefore, we chose to provide a comprehensive analysis of many uncertainty scenarios and summarized each wavelength-R^2 map in one bar of Fig. 4 or one circle in Fig. 5, showing which areas of the wavelength-R^2 yield good vegetation indices.

Methods and Discussion: A lot of effort was made to evaluate different forest structures and sensitivity test the influence of different model parameters. However, it seems like the models are all based on the environmental conditions at one site in Germany (e.g., line 114). Does this also warrant a sensitivity test? At a minimum, it requires justification. Is this site representative of a broad range of sites? Is the model going to be overfit to the conditions at Hohes Holz? If not, why not?

- Thank you, this is a good point. An earlier analysis showed that the ForestFactory approach is able to cover the range of forests found in Germany (see Fig. 8 in Bohn and Huth, 2017), so that the generated dataset is expected to be representative for a broad range of forests. The bandwidth of considered forests is also depicted in Fig. S1 (supplement) and includes very sparse forests with biomasses below 50t ODM/ha as well as very dense forests with biomasses exceeding 600t ODM/ha on small scales.
- Already in our initial submission, we had randomized the weather. However, we noted that we did not describe this properly in the manuscript. We have addressed this issue in the revision.
- The set of climate sequences we used was taken from the Hohes Holz site and overlaps with the climate in many parts of central Europe. Nonetheless, there are regions (e.g. montane areas) where the climate differs significantly. We have added this issue to the “Limitations” subsection of the discussion section.

Methods and Discussion: It seems that all wavelengths are an average of the plot, but I cannot find this mentioned. This needs to be discussed. Higher and higher spatial resolution multi- and hyperspectral data are commonplace. Are the model results still relevant on different scales? EnMAP, for example, has 30 m resolution. What is the influence of variability across the plot?

- We considered forest patches of 20m x 20m (see line 103 of the original mansucript), which is the natural scale of the FORMIND model and the same order of magnitude as the EnMAP data. An analysis across spatial scales would certainly be interesting and relevant, but we see particularly the EnMAP mission as our main motivating application. Hence, we believe it is fair to defer a scale analysis to future studies if more coarse-grain results are desired.

Methods and Discussion: Could a simple correlation analysis between parameters and single wavelengths provide some additional insight into appropriate wavelengths before looking at pairs?

- We are uncertain which parameters in particular you refer to. A correlation analysis between the parameters of the radiative transfer model and single wavelengths have been carried out in several studies already, see e.g. Mousivand et al. (2014), Verrelst et al. (2015), and Prikaziuk and van der Tol (2019).

Discussion: I believe this analysis would benefit immensely from (even a very small) application to real-world data. I realize that it is an additional effort which the authors may consider beyond the scope of the analysis, but it is relevant to the level of trust that can be placed in the primarily theoretical findings. Could several derived indices simply be applied to data from Hohes Holz to estimate biomass and compare with survey data? At a bare minimum, some concrete recommendations for next steps are required in the discussion. What models should we try with real-world data? For example, a short table of high performing wavelength combinations which match common satellite data could be very helpful. A huge amount of detailed effort has gone into the analysis, so it seems a bit of a shame for the discussion to end on high-level findings about general regions of the light spectrum that are good for GPP and NPP.

- Thank you for these suggestions. A direct application of our approach to ground data would be very desirable. The field data available to us for the Hohes Holz site were not suited to conduct the analysis, as they were too small in scale. At small scales, noise and parameter uncertainty with spatial autocorrelation can become dominant, making it difficult to validate indices designed for larger scales.
  Analyzing other data, e.g. from EnMAP, would require an exact mapping of to ground-based data, appropriate pre-processing such as error correction of the ground data, and a thorough statistical analysis. We hence considered this out of scope of this study. Therefore, we have stressed the theoretical / preparational character of our study in the discussion and dedicated a section of the discussion to making

concrete suggestion regarding future steps considering field data (see section 4.4 in the revised manuscript). There, we refer to the specific bands of the MODIS, Sentinel, and Landsat satellites that are most promising to be used for estimating the considered forest characteristics according to our results.

Line-by-line Comments:

Line 5: "with two wavelengths (400 nm- 2400 nm)" suggests only two wavelengths are evaluated. I think this should read something like: "using wavelengths between 400 nm and 2400 nm"

- We revised this sentence in the abstract.

Line 9: "provide highly accurate estimates" of what?

- Forest attributes. Fixed.

Line 9-11: This sentence is generally unclear and 'estimable' (which means 'worthy of great respect') was unlikely the intended word choice.

- We have adjusted the wording and exchanged "estimable" with "inferable" in the entire manuscript.
- FYI, "estimable parameters" / "estimability" is an established concept in the stats literature.

Line 12: "did not primarily reduce the achievable accuracy" I am not sure what this means. The accuracy of what?

- The accuracy of the wavelength-based models for estimating forest properties. Fixed.

Line 21: "Here" implies that remote sensing is part of the analysis.

- The introduction has been rewritten, and this sentence has been removed.

Line 24: Just because hyperspectral data captures more data does not make it inherently more useful than traditional multi-spectral data. See my comment on the direction of the introduction above.

- The introduction has been rewritten, and this sentence has been removed.

Line 55: 2500 nm does not match the rest of the text.

- The introduction has been rewritten, and this sentence has been removed.

Line 133: This seems somewhat circular but also makes a good case for applying new vegetation indices to the actual site to see how it performs.

- The sentence refers to the parameterization of the mSCOPE model. An application to field data would be desirable but is beyond of what we could do in this study. Please refer to our response to the last main comment for further explanation.

Line 134 to 149: Moving this (and associated results) to the Supplement would help smooth out the main text.

- Thank you. Please refer to our response to your first comment on Methods and Results.

Line 151 to 152: How much was removed by this filtering?

- Less than 2%. We have added this to the text.

Line 161: sp. value

- Fixed.

Table 1: Different sensors capture different wavelengths, why were these chosen for classical indices? I think it deserves an explanation in the text.

- We chose wavelengths compatible with existing satellite missions (MODIS, Landsat) which typically cover larger bands rather than individual wavelengths. As such, the potential wavelengths are not unique. We have clarified this in the text.

Line 182: “in the absence of noise” Is this the uncertainty analysis? It is not clear.

- Yes. We have clarified this.

Line 195: Can this include a percent difference?

- Yes. We have added quantitative results to the revised manuscript.

Line 202: “estimable” again.

- Fixed. See our comment above.

Line 280: Yes, but it was not the original intent of most VIs to determine NPP or GPP. This should be clarified here and in the introduction.

- This is true, and we have addressed this in the revision.

Thank you for the opportunity to evaluate this work and I wish the authors the best of luck in their revisions.

- Thank you, your comments were very helpful to improve the manuscript! We hope that our changes to the manuscript could dispel your remaining concerns.

Colin Bloom

Response to RC2

This manuscript presents a comprehensive and methodologically innovative study that systematically explores the potential of data-driven vegetation indices (VIs) to estimate key forest properties (biomass, LAI, GPP, and NPP). By coupling an individual-based forest dynamics model (FORMIND) with a multilayer radiative transfer model (mSCOPE) and applying a Monte Carlo sampling strategy, the authors generate an exceptionally large and well-structured synthetic dataset. The analysis of all possible two-wavelength combinations across the 400–2400 nm range, combined with an explicit treatment of multiple uncertainty sources, represents a substantial advance over existing studies.

Overall, the manuscript is scientifically sound, clearly written, and highly relevant to the remote sensing and forest ecology communities, particularly in the context of emerging hyperspectral satellite missions such as EnMAP. The study offers valuable conceptual insights into wavelength selection, index design, and uncertainty robustness. I think the manuscript suitable for publication after some revisions, mainly aimed at clarifying applicability and ensuring reproducibility.

- We would like to repeat our thanks for your time and effort in reviewing our manuscript and all your helpful comments! We respond to your individual comments below and explain how we addressed them to improve the manuscript.

Major comments

1. The introduction could be strengthened by improving accessibility for a broad audience. A greater emphasis on the ecological motivation would be beneficial. The authors could clarify why forest parameters such as forest biomass, LAI, GPP, and NPP are critical variables and why their large-scale estimation remains challenging.

- Thank you. We have completely rewritten the introduction, putting a strong emphasis on the motivating underlying problem of estimating forest properties from remote sensing data. To make space for the additional information, we have reduced the background on methods and moved it to the methods section where applicable.

2. While the comprehensive coupling of forest model and radiative transfer model is impressive, the manuscript would be more convincing if the authors could demonstrate that FORMIND reasonably represents real forest conditions. For example, how do the observation-based estimates of forest properties at sites such as “Hohes Holze” (i.e., biomass, LAI, GPP, NPP) compare with the simulated ranges generated by FORMIND? Are all observed values captured within the model’s simulated distributions?

- Thank you; these are important questions. We have validated FORMIND at the “Hohes Holz“ site by initializing it with forest inventory data and comparing the

model predictions with values for GPP, NEP, and respiration derived from eddy covariance data at the site (Pohl et al., 2023). The validation results are presented in our supplement (S5.3). Further validation results on a daily time scale (but an earlier version of the parameterization) can be found at Holtmann et al. (2021) for the Hohes Holz site and at Rödig et al. (2017) for a different site in central Germany. We added this information to the main text (lines 105ff).

Minor comments:

Line 5: “two wavelengths (400 nm-2400 nm)”à”two wavelength (within 400 nm -2400 nm), to avoid the confusion.

- We changed this sentence in the revised manuscript.

Line 116: You may specify the meaning of the ODM and the correct unit for different variables.

- Done.

Lines 117-118: Consider briefly explaining the interpretation of DBH entropy values—for example, does a more negative value indicate lower heterogeneity?

- We have done that. Yes, a lower value indicates lower heterogeneity and / or stronger dominance of a single individual.

For Table 1: It would be helpful to add references or explanations regarding the rationale for selecting the wavelengths used in classical indices.

- We chose wavelengths compatible with existing satellite missions (e.g. MODIS, Landsat), which typically cover larger bands rather than individual wavelengths. As such, the potential wavelengths are not unique. We have clarified this in the revision.

Lines 179-180: what are the criteria to select the “separating thresholds” for biomass and DBH entropy?

- We chose the thresholds so that the groups had different qualitative properties, such as being sparse (low biomass / low saturation effects) or dense (high biomass / saturation effects). We have added this additional information to the text.

Lines 341-342: The study would be further strengthened if the developed hybrid model could be validated against sites with hyperspectral observations and observation-based estimates of forest properties in future work.

- We fully agree. Though outside the scope of this study, we have added more detailed recommendations for testing and developing new vegetation indices based on field data in section 4.4. This will facilitate follow-up studies. Please see also our response to the corresponding point of reviewer 1.

References

The references can be found in the reviewed main text. The only new references used in our response are the following:

Holtmann, A., Huth, A., Pohl, F., Rebmann, C., & Fischer, R. (2021). Carbon Sequestration in Mixed Deciduous Forests: The Influence of Tree Size and Species Composition Derived from Model Experiments. Forests, 12(6), 726.

Rödig, E., Huth, A., Bohn, F., Rebmann, C., and Cuntz, M. (2017). Estimating the fluxes of forests with an individual-based forest model. Forest Ecosystems, 4(1):4.

Shugart, Herman H, Wang, Bin, Fischer, Rico, Ma, Jianyong, Fang, Jing, Yan, Xiaodong, Huth, Andreas, and Armstrong, Amanda H, "Gap Models and Their Individual-Based Relatives in the Assessment of the Consequences of Global Change", *Environmental Research Letters* 13, 3 (2018), pp. 033001.