Authors Response to Reviewer's Comments

ECOSPHERE Manuscript ID (ECS23-0342) Tree regeneration in models of forest dynamics: a key priority for further research

Olalla Díaz-Yáñez*, Yannek Käber, Tim Anders, Friedrich Bohn, Kristin H. Braziunas, Josef Brůna, Rico Fischer, Samuel M. Fischer, Jessica Hetzer, Thomas Hickler, Christian Hochauer, Manfred J. Lexer, Heike Lischke, Mats Mahnken, Paola Mairota, Ján Merganič, Katarina Merganičová, Tobias Mette, Marco Mina$, Xavier Morin, Werner Rammer, Christopher P.O. Reyer, Simon Scheiter, Daniel Scherrer, Harald Bugmann

2023-09-24

Table of contents

1 Internal instructions

- underlined comments are comments that still need attention

2 Journal formatting (ToDo list)

- Respond to the review comments in the space provided. You can use this space to document any changes you make to the original manuscript. In order to expedite the processing of the resubmitted manuscript, please be as specific as possible in your response to the review comments. If you disagree with a point, please explain why. Please note that the field in ScholarOne does not retain type formats such as italics, boldface, or colors, so please format the responses accordingly. We suggest you upload a separate file for your "Author's Response to Reviewer Comments" document and use our template: https://www.esa.org/wp-content/uploads/2021/04/Author-Response-to-Reviewers-Template.docx.
- A version uploaded in the file list should be given the designation "Additional File for Review but NOT for publication".
- Please also include a "track changes" version of your file, along with a clean copy as your Main Document. This "track changes" version should be uploaded in the file list list and given the designation "Additional File for Review but NOT for publication".
- IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please replace files of the earlier version and delete any redundant files before completing the submission.
- To submit your revised manuscript, log into https://mc.manuscriptcentral.com/ecosphere and enter your Author Center. You will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision. Please DO NOT upload your revised manuscript as a new submission.
- DEADLINE: We encourage you to submit your revised manuscript within three weeks from today's date. The final due date for submitting your manuscript with the same manuscript number is 14-Dec-2023.
- If you are unable to meet the final deadline, you can request an extension by contacting the the journal's editorial staff (esajournals@esa.org). Please note that we can only consider extensions of a few weeks. If you cannot submit by the final due date, you must create a new submission for your manuscript. If creating a new submission, be sure

to note the previous manuscript ID (ECS23-0342) in the online submission form and include an "Author's Response to Reviewer Comments" document (using the template linked to earlier in this message) in the file list.

- Editorial staff note: In addition to revisions according to the review comments below, please make revisions to your paper to conform to the journal style at this stage, especially the tables, figures, and supporting information, anticipating the possibility that your paper may be accepted for publication.
- Please provide the main manuscript file composed in Word, taking care to use the Table function for the tables. However, if you created your manuscript in LaTex, please provide only the .tex file of the main manuscript, a .pdf file of the main manuscript (that is the exact same in content as the .tex file), and a .bib file.
- Ecosphere page style allows for five levels of headings. Please use different type styles (italics, boldface, boldface italics, etc.) consistently to distinguish between the levels of headings in the article and delete the section numbers in the headings.
- Literature references: Please check that each citation has complete information, and check that all references in text have a matching citation in the References section and vice versa.
- Table and figure references: Please be sure that all tables and all figures have in-text citations in the main body of the manuscript, and that tables and figures are called out in numerical order (i.e., Figure 1 is called out before Figure 2).
- Figures: Each figure should be provided in an individual, complete file (all panels in one image, identified by figure number in the file name), preferably as a high-resolution (300-600 dpi) image file or high-quality PDF. Preferred file formats include TIF, EPS, PDF, or AI at 600 ppi, while JPEG, PPT/PPTX, or DOC/DOCX are acceptable if the resolution is sufficient. Only one version of each figure should be provided. Please delete figure images from the main manuscript file. Please check for misspellings and be sure that all parts, including all lettering, symbols, labels, and scale numbers, are discernible, clear, and readable when the figures are presented at the maximum allowable size (18 cm wide by 22 cm tall) in the typeset and paged PDF version of the article. See additional instructions for preparation in the Manuscript Preparation Guide.
- Figure captions: Please be sure to include a list of figure captions in the main manuscript file and ensure that any abbreviation that appears in a figure is identified in the caption.
- Supporting Information: Each appendix must be assembled exactly as you intend it to appear online. (Appendices are posted "as is"; ESA and Wiley staff do not copy edit or assemble appendices.) Note also that the Ecosphere naming convention for appendices is of the style "Appendix S1," "Appendix S2," etc., and for tables and figures within appendices is "Table S1," "Table S2," etc. or "Fig. S1," "Fig. S2," etc., and requires that the numbering begins at "S1" in each appendix (and revise references to material in other appendices as necessary). Please be sure to eliminate line numbering and to

include the title of the journal, the title of the article, and all of the authors' names on the first page of each appendix. In addition, if there are literature references within an appendix, the appendix should include a list of literature cited.

- If the journal style is not followed at submission of the revision, we may return the manuscript to you for further revision before sending it along to the Subject-matter Editor. Closely following our manuscript preparation guide at this stage would expedite the production of your paper for publication, should it be accepted. You can download our guide here: https://www.esa.org/wp-content/uploads/2022/05/ESA-Manuscript-Preparation-Guide.pdf

3 Response to the Editor

3.1 General Comments

Dear Dr. Díaz-Yáñez:

Thank you very much for submitting your manuscript "Tree regeneration in models of forest dynamics: a key priority for further research" for review by Ecosphere. The reviewers and I appreciate the work you have accomplished. I am willing to consider a revised version for publication in the journal, assuming that you are able to modify the manuscript according to the reviewers' extensive recommendations. Your revisions should address the specific points made in the review comments. Please note that Reviewer #2 made extensive comments directly on the manuscript.
I should note that personally I can hardly disagree with your basic conclusion that models of forest dynamics need to do a good job of capturing regeneration and recruitment processes. I have devoted most of my research over the past 40 years to developing the field methods to allow incorporation of accurate seedling, sapling and canopy recruitment in the models SORTIE and SORTIE-ND, in systems ranging from the tropics to boreal forests. The specific ecological processes incorporated in those models depended on the goals of the modeling and the ecology of the system, and included processes such as masting cycles, effects of disturbance on seedbed substrate dynamics, seed and seedling predation by small mammals, browsing by ungulates, Janzen-Connell effects, and, of course, detailed analyses of shading and neighborhood competition. More recently, this includes developing statistical models of seedling recruitment from the US forest inventory network that have been directly incorporated in the model for forests of the entire eastern US. To my mind, the difficulties the models you reviewed encountered in capturing canopy recruitment speak more to the limitations of the approaches used in model development than in any inherent difficulties in capturing regeneration and recruitment in models of forest dynamics.

Sincerely,

Dr. Charles Canham Subject-matter Editor, Ecosphere

3.2 Response to editors comments

Dear Dr. Charles Canham,

I want to express my gratitude for taking the time to review our manuscript. Your insightful thoughts on our findings were very helpful.

Our main takeaway from this manuscript is that while it may be possible to calibrate all 15 models against the EuFoRIa data, it is important to ensure that this aligns with the study's goals. If the goal is to provide extrapolation into a future climate, calibrating against data under the current climate may not be reliable. This is why we believe it is valuable to test model performance without any changes or calibration as we have done in this study.

We have carefully reviewed and addressed each of the specific points made in the review comments in this revised version. We believe that the changes we made have significantly improved the manuscript. Once again, I want to thank you for the time and effort you dedicated to reviewing our work. We look forward to hearing from you again.

Best regards,

Olalla Díaz-Yáñez on behalf of all the co-authors

4 Response to reviewer 1

4.1 General Comments

The manuscript compares how regeneration and recruitment is simulated in 15 forest models, using a large network of forest reserves in Europe as the observed data. I want to start by saying there are many strengths of the paper. I was impressed with the comparison of 15 models, that the observed data was not used to parameterize or fine-tune the models, and the high quality of the writing. The introduction and discussion were easy and enjoyable to read, and the topic is critical and relevant. Some of the details about the models and methods were placed in the very substantial appendix, and I agree with their choices. However, the manuscript is still quite long with lots of multi-paneled figures. If the journal does not have a page limit, then this is not an issue. If there is a page limit, then some tough choices will have to be made for which figures to keep. Figures 2, 3, 6 and 7 are the most important (in my opinion).

I have a few major comments. One of which requires a different statistical analysis – which is why I do not consider this a 'minor revision'. However, there are no fatal flaws, and all of my comments are fixable. I hope they serve to improve the manuscript.

4.1.0.1 Our general comments to reviewer 1:

Dear Reviewer 1,

Thank you for taking the time to review our manuscript. We appreciate your thoughtful comments and have carefully considered each of them. We have made changes to the manuscript based on your feedback and believe that these changes have improved the overall quality of the work.

We understand your concern about the length of the manuscript and the number of figures included. We have made an effort to reduce the text where possible and have verified that we are following the journal guidelines. Despite our best efforts, however, it has been difficult to significantly reduce the length of the manuscript.

Please find below our responses to each of your comments and the concrete actions we have taken in response.

Thank you again for your valuable feedback.

4.2 Major comments

4.2.1 COMMENT 1

'regeneration' versus 'recruitment'. I personally have had this conversation multiple times over the years, so I understand that there is variation in the definitions and uses. However, the authors do define these terms in the methods (L167-171) and I completely agree with them. Regeneration is for flowers, seeds, and seedlings while recruitment is for reaching a certain size threshold. However, the entire manuscript refers to regeneration as "passing of trees across a specific diameter threshold ("ingrowth") (L142), either reaching a size of 7 or 10 cm DBH. This is recruitment. Since this is the response variable for the manuscript, I would suggest replacing 'regeneration' with 'recruitment' throughout the manuscript.

4.2.1.1 Response to comment 1:

We should discuss this point again I do not agree to simple change regeneration with recruitment. Specially without a detailed definition as we had in the previous version.

SUGGESTION FROM YANNEK:

Thank you for pointing this out. We also discussed terminology several times and we agree. We followed your suggestion replaced 'regeneration' with 'recruitment' where the difference is not clear. However, we believe that in some places regeneration is more accurate. Especially in many parts of the introduction and discussion. We argue that regeneration is often only

modeled via "recruitment", but with the aim to simulate the whole process of regeneration. In this context it makes more sense to use the general term (e.g., for discussion consequences.)

COMMENT from YK: I would argue that recruitment is evidence for successful regeneration. This allows us to use recruitment for discussions observation vs. simulation in both aspects: technical (recruitment) and general (regeneration).

4.2.2 COMMENT 2

The difference between 7 cm and 10 cm DBH in the observed data. From the methods, there is one dataset that has sites where recruitment was defined as 7 cm DBH, and then another where recruitment was somewhere between 7 – 10 cm DBH. It was not clear how these were treated and compared in the results. I assume that the models simulated all 200 sites, but then the recruitment results were compared like-to-like. So, if a site defined recruitment as DBH = 8.0 cm, then all 15 models would compare simulated recruitment at that site only at 8 cm. Since the data are boxplots, I am not able to tell. For example, Figure 2 – there should be 165 points for all of the 7 cm bars (observed and simulated) and 35 data points for the 10 cm bars (observed and simulated). If true, then please clarify. If not true, then please explain. Also, how was the data analysis performed when looking at the change in mortality and species diversity between 7 to 10 cm? I assume only the 165 stands were used. Please clarify.

4.2.2.1 Response to comment 2:

Thank you for bringing to our attention the issue with the way we communicated our diameter thresholds approach in the manuscript. The observed data had varying diameter thresholds for reporting tree recruitment, and, for the purposes of this study, we merged the recruited trees values into two thresholds of 7 and 10cm.

In total, we had 165 sites where the records showed the number of recruited trees at a 7cm threshold. For these sites, any recruited tree above 7cm was counted. We also had 35 sites where the minimum diameter threshold was 10cm, and for these sites, we could only account for trees recruited from 10cm diameter. We used the other 165 sites with 7cm diameter threshold in the 10cm database, but only accounting for newly established trees above 10cm. There was no like-to-like comparison for any other diameter threshold, only for 7 and 10cm.

The reviewer correctly pointed out that in Figure 2 for the 7cm threshold, there were 165 "dots" in the observed data, while for the 10cm diameter threshold, there were 200 "dots." However, simulated values were obtained for the 200 sites for both thresholds, as explained in the Simulation protocol section. Therefore, the simulated values in the figures are from the 200 sites.

When examining the change in mortality between 7 to 10cm or species diversity comparison between observed and simulated data (Figure 4), we could only use the 165 sites that contained

both pieces of information at the 7 and 10cm thresholds in the observed data. This is reported in the supplementary materials software, script "'Figures.R"' line 536 and L284-349 and now better explained in the manuscript.

To better clarify this aspect, we have made the following changes:

1) We have better explained the difference on the observed data in the section 3.2 Observed data with the following text:

"For our study, we defined two datasets where one included 165 sites with a diameter threshold of 7 cm or lower, and the other included another 35 sites with diameter thresholds from 10 cm. Therefore we had in total 200 sites with observations of newly established trees above 10 cm and 165 sites above 7cm. In the results section any result from observed data at 7 cm threshold is originated from the 165 sites."

2) We have added the following text in the caption of figure 4

"B) Examples for the three trends across models for regeneration at the 7 cm threshold (165 sites)."

4.2.3 COMMENT 3

The t-tests. The results in Table S2 and S3 (comparing at 7 cm and 10 cm DBH) should be at least by paired t-tests not two-sided t-tests (L319), where you are comparing the difference within the same stand at 7 and 10 cm. The results Table S5 and S6 should not be a t-test at all. Performing 15 individual t-tests is a classic Type-1 error situation (technically this applies to Tables S2 and S3 as well, a comment on that below). For Tables S5 and S6, I would suggest a general linear mixed effects model. You can take the difference between the observed, and the simulated output and make this the response variable (Y). Then, include model complexity, feedback, model type and scale as fixed explanatory variables. And include the stand ID as the random effect. This will more accurately answer your question about if any of these fixed explanatory variables can predict over or under representation of recruitment. For Tables S2 and S3, you could do the same thing (i.e., calculate the difference between the observed and the simulated output for each model) and make this your response variable. But the only fixed effect would be the actual model, and then you could do an ANOVA (again, stand ID as the random effect) and then a post-hoc test to see which models were or were not different from observed. If you do want to keep the 15 paired t-tests, then you need to account for Type I error rates and adjust your p-values accordingly.

4.2.3.1 Response to comment 3:

I am going to do this.

4.3 Minor comments:

4.3.1 COMMENT 4

The abstract is not clear if the models do a good or poor job in simulating regeneration (L59, that they match observed ranges, and then L66, considerable mismatch between simulations and observed data, but that they do capture tree regeneration). Please clarify the main message in the abstract.

4.3.1.1 Response to comment 4:

We have made some changes to the abstract based on feedback from both reviewers, including clarifying certain parts for better understanding. The models were able to capture some aspects of regeneration more effectively than others. One example, is the sentence highlighted by the reviewer that was refereing to the diversity of regeneration.

4.3.2 COMMENT 5

L123-138: this paragraph has a lot of "on the other hand" type of statements, and it does not lead nicely into the next paragraph about the current study. It was a bit of a surprise (i.e., I did not expect the main purpose of the study when I started to read the following paragraph). It seems like the main issue is that we just don't know how uncertain regeneration and recruitment methods are, because we lack the comparison across models and a comprehensive empirical recruitment data set that we can use to compare. And that is what your study is doing.

4.3.2.1 Response to comment 5:

Following the reviewers comments we have made several changes in this paragraph to improve its readability and flow if ideas with the main purpose of the study text.

4.3.3 COMMENT 6

Table 1. Please clarify what the letters are for (a, b, c and d in the table headers). For models that start from saplings or ingrowth, please add in the size as well (i.e., start from saplings with DBH = 1.0). Could also add in the classification scheme mentioned in the methods (L168, regeneration or recruitment) to Table 1.

4.3.3.1 Response to comment 6:

Following the reviewers comment we have added a new column classifying the models in recruit or regeneration models; and the letters a,b,c, d,e and f table are now better described. Regarding the suggestion on adding into this table the size of the saplings or ingrowth: there are models that start from saplings or ingrowth what without a constant diameter threshold as they are biomass based (e.g. LandClim), we find it complicated to add all this information in this table without making it too large and therefore we have decided not to included. Furthermore the models reports in the supplementary materials provide more details on how the models simulate regeneration.

ToDo All models check that the new column in the table is correct for your model.

4.3.4 COMMENT 7

L218 – do I understand this correctly, that regeneration rates are a mean of 56 trees per ha per decade? This seems really low (assuming that this is for >10 cm DBH).

4.3.4.1 Response to comment 7:

Yes, this is the mean value across all observations and sites, there are two main reasons why this number might seem low, first due to the zero inflated nature of the data, and second because these are forest reserves, corresponding with denser forests in which regeneration would be expected to be lower. This value is in agreement with a further observed data from reserved (Käber, Y., et al (2023). Sheltered or suppressed? Tree regeneration in unmanaged European forests. Journal of Ecology, 00, 1–15. https://doi.org/10.1111/1365-2745.14181)

4.3.5 COMMENT 8

L264 – 274: how will this be accounted for later, that not all models simulated the same set of species?

4.3.5.1 Response to comment 8:

For those models that simulated more than the eleven species, we created a different category "others" were we lumped those extra recruited species. This were counted when we calculated the diversity index and for total recruitment values. Following the reviewer comments we have added this information in the simulation protocol section with the following text:

"Two models included their standard set of species for the simulations, which is much larger (i.e., ForClim 1, ForClim 11, TreeMig), the species simulated beyond the eleven species were lumped in a category"others" "

4.3.6 COMMENT 9

L340: "regeneration basal area of that species" is mentioned twice, please delete the second one.

4.3.6.1 Response to comment 9:

Following the reviewers comment we have modified this text to improve it readability with:

"Furthermore, regarding the share in the regeneration basal area per species ($\bar{R}_{BA_{share_i}}$), we categorized this as zero when the regeneration basal area of that species ($\bar{r}_{BA_{t,i}}$) was zero, but also when both the total regeneration basal area ($\sum_{i=1}^{s} \bar{r}_{BA_{t,i}}$) and the regeneration basal area of that species ($\bar{r}_{BA_{t,i}}$) equaled zero."

4.3.7 COMMENT 10

L364: so please remove the 4C model from the figure as it is not comparable. Do you need to also remove other models that don't simulate the same number of species too?

4.3.7.1 Response to comment 10:

Even thought it is not directly comparable, we still believe is interesting for the reader that we share the findings related to the 4C model and those that simulate more species that were lumped in the category others.

4.3.8 COMMENT 11

Table S2 and S3: What are the 3 columns for p-values? Please report just the one you used. No need to have a column to state 't-test' – just move this to the table caption. Then add in a column for df, the test statistic (t) and the effect size (was it higher or lower and by how much?). NOTE, this comment may become irrelevant if you change the statistical analyses.

4.3.8.1 Response to comment 11:

To be addressed in combination with COMMENT 3

4.3.9 COMMENT 12

Figure S1 and S2 – I can't tell the difference between observed and simulated. Are they different colours? The legend should be larger. What is the black line for?

4.3.9.1 Response to comment 12:

Following the reviewers comment we have increased the legend size, and now the difference between observed and simulated is visible. We have also explained in the caption the meaning of the black line:

"Mean Shannon index per site at recruitment and at the stand level both for observed and simulated values for the diameter threshold of 7cm. The black line marks where the mean Shannon index per site at recruitment and at the stand level would be equal"

4.3.10 COMMENT 13

Figure 4 (and Figure S3) are interesting in that there are no patterns with overstory diversity, and that this was not explained by feedback with seed production. What happens if you colour the points, by those from models with the feedback? Do you address in the discussion why this feedback makes no difference to recruitment?

4.3.10.1 Response to comment 13:

We agree with the reviewer that it was interesting to find that the models with feedback did not captured better the species diversity of the regeneration compared with those without, with the exeption of ForCEEPS. The Figure 4 shows three representative patterns observed across models using 6 of them as an example, and the Figure S3 provides the details for each model. Models with feedback are defined in table 1 and are ForCEEPS(f) , PICUS, SIBYLA, xComp, iLand, LandClim, Landis-II, TreeMig, LPJ-GUESS, and aDGVM2. We have addressed this point in the discussion section with the following text:

"Furthermore, based on the simulation results, there is no evidence that models with feedback from the canopy (in terms of species composition of regeneration) captured better the species diversity of the regeneration than those without feedback, with the exception of ForCEEPS that significantly improved regeneration species diversity with the model version that included feedback (ForCEEPS(f)). The similar performance of models with and without feedback is likely because the models put more weight on the regeneration niche arising from abiotic and biotic filters, than from the habitat niche of the adult trees [@grubb1977]."

4.3.11 COMMENT 14

All figures in the Appendix, please make the X and Y axis labels larger and centered. I actually thought labels were missing initially and it took a while to find them.

4.3.11.1 Response to comment 14:

Following the reviewers recommendation we have changed the axis labels in the appendix figures.

4.3.12 COMMENT 15

L400 – 402: where is this conclusion about mixed species or monocultures coming from? I don't see that in the figures or the results.

4.3.12.1 Response to comment 15:

This paragraph is addressing the comparison with the theoretical Reineke value which is usually calculated for even-aged, single species stands and assumes a fixed relationship between the number of stems and the quadratic mean diameter in fully stocked pure stands.

4.3.13 COMMENT 16

Figure 6 is excellent.

4.3.13.1 Response to comment 16:

Thank you! We also think is a good summary figure.

4.3.14 COMMENT 17

Table S4: please clean up. No need to report 7 places after the decimal point. Please clarify in the table caption what the slope represents (maybe mention Figure S5 too, as I assume it is illustrating the slopes). Another suggestion, would be to make the slopes solid lines if significant, or dashed lines if not significant. Not sure if that would help, but if it does then it could eliminate the need for this table.

4.3.14.1 Response to comment 17:

Clean together with comment 2

4.3.15 COMMENT 18

Not sure both Figure 7 and Figure S7 are necessary (I admit I spent a few minutes going back and forth between them trying to understand how they differed). I would just keep Figure 7 and note that the scales differ.

4.3.15.1 Response to comment 18:

Following the reviewer comment we have removed figure S7.

4.3.16 COMMENT 19

L547-548: predicting stand diversity, did I miss this? I don't believe Figure 4 shows this, as if you put all data points on the same panel it would just be a random scatter of points that covers the entire space (not necessarily closer to the 1:1 line).

4.3.16.1 Response to comment 19:

Figure 4 shows three examples of the patterns observed across models (overestimation intermediate and underestimation) using one exemplary model of each category , the actual patterns per model can be found in Figure S3. In the results we showed that most models captured reasonably well or overpredicted species diversity at the stand level, as most of the models are under the patterns presented in plot 1 and 2 from A on Figure 4. We have decided to keep the referenced statement link to this findings in the discussion as it was.

4.3.17 COMMENT 20

L575: that most models did not deviate exceedingly from observations seems in contrast to what the previous section was stating. What is this based on?

4.3.17.1 Response to comment 20:

In this statement we were referring only to regeneration levels. Although we agree with the reviewer that we did state that most of the models overestimate regeneration, here our point is that even-thought they did overestimate regeneration they did not deviate exceedingly, specially under the consideration that these models have never been confronted with this type of data.

4.3.18 COMMENT 21

L584: What does this mean, about degrees of freedom for modelling regeneration?

4.3.18.1 Response to comment 21:

With this statement we were referring to the multiple potential combinations of approaches of modeling tree regeneration. Based on the reviewers comment we have modified this for clarity.

4.3.19 COMMENT 22

L624: Missing the end of the sentence.

4.3.19.1 Response to comment 22:

Thank you for noticing this. We have now corrected this mistake.

4.3.20 COMMENT 23

L647: Completely agree that observed data is a snapshot of a stochastic and dynamic system and care should always be taken with comparing to simulation data, however I disagree about the following stamen (L648). The models should be able to capture broader patterns along climatic or environmental gradients.

4.3.20.1 Response to comment 23:

We agree with the reviewer that ideally models should be able to capture broader patterns. With this sentence we just wan to explain that the stochasticity of this process may have partly explained the results on this regard.

4.3.21 COMMENT 24

L655: But the landscape models generally did about the same (i.e., not consistently worse) then the rest, correct?

4.3.21.1 Response to comment 24:

The reviewer is correct that landscape level models were not significantly performing worst than others. Here we just wanted to discuss that maybe another spatial consideration will have allowed landscape level models to better represent regeneration processes as they are designed for that.

4.3.22 COMMENT 25

L663: But in general, the models were more sensitive to water balance and the observed data did not have any real pattern. Why would this coarser representation of water balance in the model, cause it to be more sensitive?

4.3.22.1 Response to comment 25:

We agree with the reviewer, a coarse representation of water balance would not make it more sensitive to water deficits. And this was exactly our point in this part of the discussion where we wrote: "This may at least partially explain the unsatisfactory performance of many models along the drought axis (i.e., climatic water balance)."

4.3.23 COMMENT 26

L675: what would a 'comprehensive regeneration dataset" look like? There are large data sets for seed production (MASTIF, MASTREE, etc.) and seedling data from national forest inventory programs. Does EuFoRIa (L677) have everything you need? Is it just a matter of expanding the spatial and temporal extent? Please explain what sort of data that the models would need (imagine an ideal world).

4.3.23.1 Response to comment 26:

Following the reviewers comment we have expanded this aspect in the following paragraph XXXX

Both reviewers asked to be more specific on the data we are missing, what vairables would you say without being too detailed? also linked to comment 59 from reviewer 2

5 Response to reviewer 2

5.1 General Comments

The paper evaluates whether forest dynamics models are successful at capturing the essence of regeneration processes, by comparing outputs from 15 simulation models with tree recruitment data obtained from 300 mature forests spread across central/eastern Europe. Specifically, the models predicted how many trees of different species recruited each year (i.e. grew larger than 7-cm or 10-cm stem diameter) and indices derived from these numbers were compared with analogous indices derived from repeat-measured forest inventory plots. Most of the simulation models performed poorly for most of the indices examined. In particular, they tended to overestimate recruitment rates.

It is valuable to demonstrate that leading forest dynamics models are unable to characterise regeneration processes very well. I also thought the "blind trial" approach was clever. Huge effort must have gone into running 15 models under various starting conditions and synthesising the outputs into a coherent paper. Choices have been made about which details to include in the main body of the text vs supplementary information and I thought a good balance was struck, although I have made several suggestions on the pdf.

5.1.0.1 Our general comments to reviewer 2:

Dear Reviewer 2,

We would like to express our sincere gratitude for all the detailed comments they provided on our manuscript. We have carefully considered all of their feedback and made the necessary changes to improve the manuscript. Thanks to your comments, we have improved our manuscript.

Thank you again for your valuable feedback.

5.2 Major comments

5.2.1 COMMENT 1

Where the paper falls short, in my opinion, is in providing a clear path forward. It shows that the models are too simplistic to simulate regeneration processes accurately but provides few insights into how we can improve their performance. The paper reports that creating more complex models is not valuable in and of itself, but that's not to say that intelligent improvement of specific subcomponents may hold the key. For instance, deer browsing and competition with herb layer plants may well contribute to lower recruitment being observed than simulated, but how can we demonstrate that, and if the inadequate representation of these processes in the models is the problem, what should be done to improve the models?

The review by Hanbury-Brown et al. (2022) on "future forests within Earth system models: regeneration processes critical to prediction" provides a good summary of current knowledge and future directions, including thoughts on 1. reproductive allocation and seed production; 2. dispersal; 3. seed survival, germination, and resprouting; and 4. seedling survival and growth. I would encourage the authors to frame their discussion in a similar way. Advanced statistical approaches that bring together simulation models and Approximate Bayesian Computation to estimate parameters of recruitment submodels may hold the key (e.g. https://onlinelibrary.wiley.com/doi/full/10.1111/ecog.04824). To me, a key question is how we can make better use of field measurements and statistical analyses to refine forest dynamics models, i.e. not only parameterise them but also to identify which submodels need refinement and the function that need inclusion.

5.2.1.1 Response to comment 1:

I think we have proposed some ideas of how to do this better in the discussion. What else do you think we should add here?

5.2.2 COMMENT 2

I felt the readability of the paper would be improved by introducing a "road mapping" paragraph at the end of the introduction or start of the methods section that sets out the approach in broad terms.

5.2.2.1 Response to comment 2:

Following the reviewer comment we have added a "road map" at the end of the introduction.

5.2.3 COMMENT 3

The words "ingrowth", "regeneration" and "recruitment" are used interchangeably to describe the recruitment of new trees into the 7-cm diameter class. I would encourage the authors to use the term "ingrowth" or "recruitment" to this to describe this process, and keep "regeneration" as an overarching term describing the entire process of seed production, dispersal, early establishment, and onward growth.

5.2.3.1 Response to comment 3:

Decide final terminology

5.2.4 COMMENT 4

The article is written entirely by Europeans, unless I'm mistaken, and very much focused on the central/eastern European forestry literature. I would encourage a deeper dive into the relevant North American and Asian literature when introducing and discussing their work.

5.2.4.1 Response to comment 4:

The reviewer is correct that most of the people involved in this manuscript have a central European focus in their current work, and so did the study set up. We have included some relevant global or with a focus outside Europe references when relevant. e would be happy to consider some other relevant suggestions.

5.3 Minor comments:

These comments have been extracted from the reviewer 2 comments on the PDF document of the manuscript.

5.3.1 COMMENT 5

L51-52 "However, an assessment of their ability to accurately represent tree regeneration is lacking." Not true.

5.3.1.1 Response to comment 5:

Based on the reviewers comments we have modified this sentence to: "However, there is a need to evaluate their capacity to accurately represent tree regeneration"

5.3.2 COMMENT 6

L56 "The results are evaluated against comprehensive data from unmanaged forests." comprehensive?

5.3.2.1 Response to comment 6:

Following the reviewers comment we have changed the word "comprehensive" to "extensive" to express that we used a wide-range data that considers many important elements of regeneration.

5.3.3 COMMENT 7

L56 “Models simulating higher species diversity at the stand level do not feature higher regeneration diversity.” feature?

5.3.3.1 Response to comment 7:

Following the reviewers comments we have changed “feature” with “present”.

5.3.4 COMMENT 8

L76-78 “A wide range of models of forest dynamics were developed over the past decades considering the impacts of climate”

5.3.4.1 Response to comment 8:

Following the reviewers comment, we have changed the word “were” with “have been”.

5.3.5 COMMENT 9

L79 K. Vanclay and Skovsgaard 1997;

5.3.5.1 Response to comment 9:

Following the reviewers comment, we have corrected the reference style.

5.3.6 COMMENT 10

L83-84 “Which is a clear research gap in the context of climate-induced forest disturbances and forest resilience.” incomplete sentence

5.3.6.1 Response to comment 10:

Following the reviewers comment, we have changed this sentence to: “This presents an important research gap, particularly in the context of climate-induced forest disturbances and forest resilience.”

5.3.7 COMMENT 11

L85-86 “Tree regeneration arises from multiple processes such as pollination, fruit maturation, seed production, dispersal, germination, juvenile growth and survival” including

5.3.7.1 Response to comment 11:

Following the reviewers comment, we have changed “such as” with “including”

5.3.8 COMMENT 12

L85-86 “Tree regeneration arises from multiple processes such as pollination, fruit maturation, seed production, dispersal, germination, juvenile growth and survival” including

5.3.8.1 Response to comment 12:

Following the reviewers comment, we have changed “such as” with “including”

5.3.9 COMMENT 13

L90-92 “Currently, tree regeneration processes in dynamic forest models are handled in a multitude of ways (König et al. 2022; Bugmann and Seidl 2022): from 1) entirely ignoring it (as done in classical forest growth models, e.g., Pretzsch et al. 2002),” it.

5.3.9.1 Response to comment 13:

Following the reviewer’s comment we have changed “it” with “them”

5.3.10 COMMENT 14

L96-98 Particularly given your conclusion that the models are poor at predicting regeneration processes, should we now return to statistical parameterisation afresh and find better ways of extrapolation/interpolation across space/time?

5.3.10.1 Response to comment 14:

It is indeed interesting to account with the increasing data vailablity to better understand tree regeneration processes. However, we believe that the main take away it is not that it would be impossible to calibrate all 15 models against e.g. the EuFoRIa data to better capture regeneration, but that the focus should be in the study goals. And when the goal is to provide extrapolation into a future climate, then calibration against data under current climate may be unreliable. With the approach we have followed in this study where we have asked the models to simulate forest dynamics and capture regeneration without statistical predetermination we can better assess their behavior and understand how their current model structures and inclusion of exclusion of certain ecological processes impact the outputs we are seeing.

Following this and other comments from the editor, we have tried to capture this better in the discussion/conclussions section. Should we add something on this?

5.3.11 COMMENT 15

L100-101 "Overall, models are needed to…(3) identify the most important processes that are shaping ecological patterns." doesn't that depend on what goes into the model?

5.3.11.1 Response to comment 15:

We are not complete sure of the reviewers point here. If the reviewer is referring to the fact that the most important processes will be those included in the model, it is true that you can only asses those processes included or how including or not another processes impact regeneration. This was also the idea with considering a large pool of models, that we could assess how their structural complexity, or the inclusion/exclusion of a certain process affects the results we are seeing. This also goes very much in line with our research recommendations presented in the discussion where we proposed that improvement of the regeneration modules is implemented as additional features that can be traced back, as done here for the variants of ForClim and ForCEEPS, and that model complexity and structure must always be connected with modeling objectives.

5.3.12 COMMENT 16

L101-106 "Given the current strategies that are used in models of forest dynamics to represent tree regeneration, model behavior often is prone to problems, such as very high levels of tree regeneration that necessitate excess mortality at early stages of tree life to simulate correct stand structure and composition." ref? tens of thousands of seeds...

5.3.12.1 Response to comment 16:

Following the reviewer's comment we have added a reference

ToDo find a reference! Any suggestions?

5.3.13 COMMENT 17

L112-114 "A related issue is the excessive reduction of species diversity due to positive feedback effects, such that eventually just single-species stands remain." ref

5.3.13.1 Response to comment 17:

Following the reviewer's comment we have added a reference

ToDo find a reference!Any suggestions?

5.3.14 COMMENT 18

L127 I suggest making better use of the hanbury-brown 2022 review.

5.3.14.1 Response to comment 18:

We currently use the exciting Hanbury-brown 2022 review in several places across our manuscripts, following the reviewer's comment here and the general comment 2 in Section 5.2.2 we have also get inspired by their structure by creating a road map to help the reader to navigate the manuscript content.

5.3.15 COMMENT 19

L145-147 "Due to the large variability in tree regeneration patterns in nature and the large number of factors driving this process including some that are not incorporated explicitly in most models, such as deer browsing—"

5.3.15.1 Response to comment 19:

We believe the reviewer highlighted the words "models, such as deer browsing" because they think this is not correct. From our sample pool only 4 models incorporated explicitly deer browsing.

5.3.16

5.3.17 COMMENT 20

L150 “regeneration niche of multiple species”

L154 “traits”

5.3.17.1 Response to comment 20:

We are unsure why the reviewer highlighted those words

5.3.18 COMMENT 21

L185 “(mean regeneration formulation complexity across all processes >0, Table 1).” Meaning

5.3.18.1 Response to comment 21:

This statement relates to Bugmann and Seidl (2022) where complexity values were provided. The exact details of how we calculated this mean regeneration formulation complexity value are available in the software suplementary materials to this paper: “‘figure.R“‘ line 706-750. This value represents the mean complexity of regeneration formulations per model.

5.3.19 COMMENT 22

L186-188 “ForClim variant 1 (Bugmann et al. 1996) is based on a recruitment module that adheres closely to the concept introduced by Botkin, Janak and Wallis (1972)” which is…

5.3.19.1 Response to comment 22:

Is there a way to explain this shortly??

5.3.20 COMMENT 23

L197-198 “Regeneration data covering a wide range of environmental conditions are hard to obtain, and this is one of the reasons why most models of forest dynamics have never been confronted with a dataset” not sure that’s true. Some of the large forest inventories include small seedlings and saplings in nested plots, which are a step in the right direction.

5.3.20.1 Response to comment 22:

We agree with the reviewer that permanent plots such as those present in National Forest Inventories could potentially be a good source of information to obtain regeneration data. However these often are collected from managed sites, do not capture ingrowth below 10cm, and are hard to harmonized between different inventory strategies.

5.3.21 COMMENT 24

L200-202 "The observations used here are derived from a novel and unprecedented network of sites in forest reserves that represent the range of environmental gradients in temperature and precipitation in Central Europe as compiled in the framework of the EuFoRIa network (EuFoRIa 2019)" I don't feel "novel and unprecedented" adds anything to this sentence. There are plenty of networks outside Europe providing this sort of data

5.3.21.1 Response to comment 24:

We agree with the reviewer's comment that there are other forest inventory data networks outside Central Europe even on forest reserves that could provide this sort of data. However, the EuFoRIa network is indeed novel and unprecedented network in Central Europe.

5.3.22 COMMENT 26

L208-209 "We selected 200 sites from this network as the benchmarking dataset for the simulation to be representative of the environmental variation contained in the data." COOL

5.3.22.1 Response to comment 26:

We agree :)

5.3.23 COMMENT 27

L211-212 "exposition (i.e.,slope and aspect)." unfamiliar term to me

5.3.23.1 Response to comment 27:

Following the reviewer's comment we have changed "and exposition (i.e.,slope and aspect)." with ", slope and aspect."

5.3.24 COMMENT 28

L212-213 “Regeneration thresholds for these sites differed between diameters of 0 and 10 cm” minimum size threshold? and stem diameters

5.3.24.1 Response to comment 28:

Following the reviewer’s comment we have changed this sentence to: “Regeneration minimum size thresholds for these sites differed between stem diameters of 0 and 10 cm”

5.3.25 COMMENT 29

L215 “another 35 sites with diameter thresholds between 7 and 10 cm.” OK, so we’re talking about v large trees here

5.3.25.1 Response to comment 29:

The minimum threshold that could be standarized across reserves was of at least 7cm.

COMMENT 30

L215 “another 35 sites with diameter thresholds between 7 and 10 cm.” OK, so we’re talking about v large trees here

L218 “featured 30,900 newly established trees” so they are trees….

5.3.25.2 Response to comment 30:

The minimum threshold that could be standardized across reserves was of at least 7cm.

5.3.26 COMMENT 31

L215 “on this unique dataset” not so unique

5.3.26.1 Response to comment 31:

Please, see our reply on comment 25 in here **?@sec-R2C25r**

5.3.27 COMMENT 32

L221 “nformation, cf. Käber et al. (2023).” in review and not available

5.3.27.1 Response to comment 32:

This paper have recently being published and the reference has been updated accordingly.

5.3.28 COMMENT 33

L223 “The overarching goal of the experiments was to assess the tree regeneration as it arises” simulation experiments and remove “the”

5.3.28.1 Response to comment 33:

Changed as suggested

5.3.29 COMMENT 34

L225 “tree regeneration as the passing of a breast height diameter threshold of 7 or 10 cm, respectively” I suggest “tree recruitment” not “tree regeneration” here. respectively??

5.3.29.1

5.3.29.2 Response to comment 34:

We have removed “respectively”.

We have re-assessed the terms used across this manuscript XXX

5.3.30

COMMENT 35

L 227-228 “simulations, providing input variables” provided with

5.3.30.1 Response to comment 35:

We have removed “providing”.

5.3.31 COMMENT 36

L229 “Neither were further site information (except for the data specified below) nor any data” These were “blind trials”: modelling group were not provided with site information ...

5.3.31.1 Response to comment 36:

The reviewer is correct that we did not provide e.g. site coordinates (with the exception of iLand that needed this for a model specific soil input, this is described in the manuscript), but we did provide topography, climate and soil site specific variables.

5.3.32 COMMENT 37

L230-231 “That is, the models were run in”blind flight” mode.” Remove

5.3.32.1 Response to comment 37:

Changed as suggested.

5.3.33 COMMENT 38

L237-238 “Soil quality data were provided as continuous values between 1 and 5” ?hydrological quality?

5.3.33.1 Response to comment 38:

The soil quality value is based on a random forest that was trained to predict expert knowledge based soil quality on a scale from 1 to 5 from the SoilGrids250 data set and the WISE data set. The most influential variables in the random forest were total depth to bedrock and sand content. Thus the variable is a good predictor for plant available water storage capacity. This information has been provided in the protocol of this study.

5.3.34

COMMENT 39

L247 “The exact length of the simulation was also decided by the modeling teams” seems odd not to constrain that.

5.3.34.1 Response to comment 39:

We understand the reviewers concern on not constraining the length of the simulations, but we decided to leave it to the assessment of the individual modelers, as we wanted to obtain simulation that were run from from bare ground in the absence of management to a simulated equilibrium (“potential natural vegetation”) with current climate, and each modeler know best how to obtain this in each model. This information is currently available in the manuscript and in the protocol of this study.

5.3.35 COMMENT 40

L250 “The simulations were run in the absence of management to a simulated equilibrium” The absence of disturbance and running to equilibrium set quite specific conditions; suggest this is emphasised more in abstract / intro / discussion.

5.3.35.1 Response to comment 40:

We agree with the reviewer that the simulations set up regarding absence of management and to a simulated equilibrium are quite specific. We found challenging to incorporate this in the introduction as we see this a more of a methodological specification, and in the abstract due to space constraints as we though is more interesting to focus in the results and meaning of our findings than in methodological aspects. However we currently have a lengthy discussion on the equilibrium assumption and also about the importance of considering disturbances, which we think it would be a great further research direction to better understand regeneration from models of forest dynamics.

5.3.36 COMMENT 41

L252-253 “This entails the assumption that (1) the observations from the forest reserves reflect no traces of forest management,” That’s seldom the case in Europe where the legacies of management are evident centuries on

5.3.36.1 Response to comment 41:

The reviewer is correct and management legacies might be evident for long periods of time. The data were collected in forest reserves where no management has taken place for long periods of time which makes our assumption of an equilibrium between forest properties and environmental drivers reasonable. This is the best data we have and we have address its limitations, as this one, in the discussion because we are also aware that is not ideal.

5.3.37 COMMENT 42

L275 “Each of the models reported the regeneration number by sampling 200 times in a 10-year interval” regeneration number = number of “recruits” entering the 7-cm size class?

5.3.37.1 Response to comment 42:

Following the reviewer’s comment we have changed “regeneration number” with “number of trees crossing the two diameter thresholds”

5.3.38 COMMENT 43

L295 "(2) regeneration species diversity" species diversity of recruitment

L295 "(3) regeneration mortality" mortality of recruits

5.3.38.1 Response to comment 43:

Change if we change regeneration with recruit

5.3.39

5.3.40 COMMENT 44

L296 "ingrowth gradients along the regeneration niches." ?? not clear to me. How is regeneration niche defined??

5.3.40.1 Response to comment 44:

We defined the regeneration niche as the passing of a diameter threshold of 7 and 10 cm and we focused both in the width of the regeneration niche (i.e., in environmental space) as well as the intensity of the regeneration process (i.e., the number of ingrowth trees per area and per unit of time) were of interest. This is currently defined in the material and methods section and in the protocol of this study.

5.3.41 COMMENT 45

Equations 1-5. not convinced these formulae need to be shown as the Shannon index is very well known. Consider converting the Shannon index to the effective number of species, by taking the negative of the SI and exponentiating it. This is gaining ground in ecology as it's more intuitive to interpret than SI (see paper by Jost).

L1046 Such converting to effective no of species using Jost formula

5.3.41.1 Response to comment 45:

We agree with the reviewer that the Shannon index is very well known index and its equation as well. However we think is still important that we define how we did calculate the Shannon index based on the available data we had as we had some internal discussion on how to do this and the meaning of this index in this context, we though the reader will benefit from a more detailed explanation. In any case the exact details of how we did calculate this index based on the simulation outputs are also available in the software supplementary materials, in the script figures.R L42-78. We also thank the reviewer for the suggestion of using the effective number of species instead of the Shannon index, it is indeed a good alternative, however we have decided to leave it with the SI.

5.3.42 COMMENT 46

L308 "Mortality in tree regeneration was assessed based on the ratio of regeneration between the 7 and 10" of recent recruits. Now I understand the 7 and 10 cm thresholds mentioned earlier; please clarify earlier mention.

5.3.42.1 Response to comment 46:

Following the reviewer's comment on the lack of claritiy of the selected thresholds explanation we have changed at the beginning of the section simulation protocol the first sentence where we explain that we considered tree regeneration as the passing of a breast height diameter threshold of 7 and 10 cm. We have also reported later on in this section (but before this line) that each of the models reported the number of trees crossing the two diameter thresholds.

5.3.43

5.3.44 COMMENT 47

L311-313 "The Reineke self-thinning rule is usually calculated for even-aged, single species stands and assumes a fixed relationship between the number of stems and the quadratic mean diameter in fully stocked pure stands." more to the point, it's always (I think) applied to all stems in the stand, not just regenerating trees, as it's fundamental concept is that entire canopies are space filling . Why would we expect it to apply to a narrow cohort? I think not.

equation 5 Suggest writing N7 / N10

Given my comment above, is there a strong justification for the -1.605 exponent? Or just keep it at one?

5.3.44.1 Response to comment 47:

Following the reviewer recomendation we have changed the equation formulation to better represent the Reineke relationship. The reviewer is correct that the Reineke is applied to all stems in the stand, in our case the N value is only changing due to the available regeneration at the two diameter thresholds as the adult trees number is the same. By comparing our expected relationship of recruited trees at 7 and 10cm with the Reineke value we are just comparing to a theoretical value in fully stocked stands. We have use the coefficient -1.605 because Reineke attributed a general validity to this allometric coefficient for fully stocked, even-aged forest stands, regardless of tree species and site.

5.3.45 COMMENT 49

L316-317 "model complexity as defined by Bugmann and Seidl" say a little more…

5.3.45.1 Response to comment 49:

Any suggestion of how can I add something else without making this much longer!

5.3.46 COMMENT 50

L317 "model type (empirical or process based)," if statistical analyses are used to estimate the coefficients of complex functions that underpin a process based model, is that model empirical or process based?

5.3.46.1 Response to comment 50:

The reviewer's comment is a very valid one as it is true that many processes in process based models have been defined based on empirical approaches. In this study we have categorized the models in process based and empirical based on their general approaches to define forest dynamics. This categorization is available in table 1. It is also possible to find further details on each of the models in the supplementary materials.

5.3.47 COMMENT 51

L496 "confronted with a unique dataset from unmanaged"

L498 "The EuFoRIa data (Käber et al. 2023) are exceptional, particularly"

I think there are plenty of other datasets that could be used and have better information on smaller trees than EuFoRIa, so suggest a more nuanced sales pitch here.

5.3.47.1 Response to comment 51:

To our knowledge this is the only network of forest reserves in Europe covering a large environmental gradient and from which a standardization effort has been made toward understanding tree regeneration across different countries. We would be happy to learn of any other network of this characteristics in Europe and potentially use it in follow up studies.

5.3.48

5.3.49 COMMENT 52

L525-527 "There are multiple constraints to the regeneration niche of tree species (Price et al. 2001), and therefore the absence of regeneration is likely to be common (Fortin and DeBlois 2007), even over larger areas" YES - good point. This should be featured more prominent I think

5.3.49.1 Response to comment 52:

Thank you for this remark, we agree with the reviewer that this is an important point and specially how different models are capturing or missing to represent the absence of regeneration. We have now captured this aspect also in the introduction in the sentence: "Due to the large variability in tree regeneration patterns in nature and the large number of factors driving and constraining this process" on top of the current whole paragraph on this issue present in the discussion.

5.3.50 COMMENT 53

L530-533 "This substantial difference may be due to the fact that the simulation results were drawn from equilibrium forests, whereas in reality many of the forest reserves are recovering from past management activities and have become denser over the past decades (e.g., Heiri et al. 2009), leading to less regeneration than in an equilibrium situation." But also --- browsing, establishment sites, competition with herb layer all reduce recruitment

5.3.50.1 Response to comment 53:

Following the reviewer's comment we have included this point by adding into this paragraph: "Another possible reason for this difference may be the exclusion of factors like deer browsing or failure to accurately measure establishment filters, such as those to site limitants or competition with herb layers."

5.3.51 COMMENT 54

L552 “there is no evidence that models with feedback from the canopy captured better the species” could you explain what you mean by “feedback” here?

5.3.51.1 Response to comment 54:

Following the reviewers comment we have modified this sentence to clarifu what do we mean here with feedback: ” Furthermore, based on the simulation results, there is no evidence that models with feedback from the canopy (in terms of species composition of regeneration) captured better the species diversity of the regeneration than those without feedback, ”

This had also been defined in the material and methods where we stated that “i.e., where the level and species composition of regeneration is influenced by the existing adult tree community via the production of seeds, seedlings or saplings, from models that do not contain such a feedback.”

5.3.52

5.3.53 COMMENT 55

L586-587 “Our study showed that increasing complexity in the regenerationmodules is not linkedwith a higher accuracy of the projections of regeneration levels,” Surely more complex models WOULD be better if they captured the relevant processes! Arguing for simple vs complex models is a distraction here... we need models that are as simple as possible but not too simple..

5.3.53.1 Response to comment 55:

We agree with the reviewers statement that we need the “right amount” of complexity which considers the most relevant processes. however there are limited studies looking at this aspect in models of forest dynamics and even more limited looking at this aspect in the regeneration related processes in models of forest dynamics. We still think it is important to discuss this issue in the context of our findings even though this was not the solely focus of our research.

5.3.54

5.3.55 COMMENT 56

L596-598 “Competition for light as a strong filter for tree regeneration has been widely documented (Messier et al. 1999; Collet and Chenost 2006; Berdanier and Clark 2016), but the

models examined here did not reproduce this expectation.", because ... remind us of the evidence you are putting forward

5.3.55.1 Response to comment 56:

Following the reviewers comment we have added the following sentence in this part of the text:

"(i.e., decreasing regeneration levels with increasing stand basal area)"

5.3.56 COMMENT 57

L603-606 "This made it impossible to evaluate the regeneration for the extremes of the stand density ranges in some models. For example, regeneration levels at low stand densities are relevant to assess how well forests are recovering e.g. after gap creation due to disturbance" Agreed ... and this paper isn't testing performance of models in predicting regeneration in large gaps. It's possible that the models do a better job at this, but it's not something you tested. Suggest making that point e.g. in the abstract.

5.3.56.1 Response to comment 57:

Should we add here:

"Using 15 models built to capture long-term forest dynamics at the stand, landscape, and global levels, we simulate tree regeneration at 200 sites representing large environmental gradients across Central Europe WITHOUT DISTURBANCES CONSIDERATION" ? or something like this?

5.3.57

5.3.58 COMMENT 58

L656 "Yet, the global models should not be at a disadvantage due to the limited spatial consideration" Don't quite follow your point here. what does "at a disadvantage" mean?

5.3.58.1 Response to comment 58:

In this sentence we are discussing how the sampling strategy presented in the study protocol might have represented a disadvantage for landscape level models due to the limited spatial scale, but how this might not be the same for Global level models as they usually lack dispersal between cells.

5.3.59 COMMENT 59

L676-677 “Therefore, we further recommend that more effort should be invested into collecting harmonized datasets on tree regeneration.” regeneration in what sense? Seedlings/ ground layer/ …. What is really needed?

5.3.59.1 Response to comment 58:

Linked to comment 26 from reviewer 1

5.3.60 COMMENT 60

L685-687 “However, this will require an entirely different set of observed data, and potentially not all models of forest dynamics would be able to assess the relationship of these aspects on tree regeneration, e.g. due to the lack of disturbance or appropriate management modules.” more akin to data already available in national inventories I think

5.3.60.1 Response to comment 60:

We agree with the reviewer that a more systematic kind of inventory could be useful to detect regeneration patterns under management or disturbances impacts.

5.3.61

5.3.62 COMMENT 61

L696 “Exercises like the one presented here, where the models are operated in”blind flight” mode.” Definition of blind flight “To do something based on guesswork, intuition, or without any help or instructions”….is that what you mean?

5.3.62.1 Response to comment 60:

With blind flight we refer to doing the simulations without having important information about the expected regeneration values.

5.3.63 COMMENT 62

Table 1 caption: A useful table. I'd encourage the authors to write a lengthier table heading, which briefly summarises the methods section, so that readers can know what "start from", "runtime for sampling" mean without having to delve into the text. Also I'd like to see an explanation of "formulation complexity" in the text and here. The supplementary information gives more details about model assumptions, and I do wonder whether those could be included here; in particular, seed production and dispersal is clearly important for regeneration, so could columns be introduced to compare assumptions of model with respect to these. Also deer browsing, herb-layers and nurse logs all have an influence of regeneration, but not incorporated into these models; I suggest the table heading makes that clear.

5.3.63.1 Response to comment 62:

Following the reviewers comment we have expanded the caption of table 1 with the following text:

"Models included, their characteristics and simulation strategies used in this study. Where Scale refers to the models scale, *Type* to the models type, *Population structure* to the model population approach, *Feedback* to the inclusion or not of adult tree feedback to the level and species composition of regeneration, *Approach* to the model regeneration module approach, *Start from* to the stage from which trees are recruited, *Complexity,* to *the* mean regeneration formulation as estimated in (Bugmann & Seidl, 2022), *Species* to which species from the original eleven species were simulated , *Runtime (spin up) [years]* to the number of years used per model in their simulations for the spin up period and regeneration sampling; *Climate data type (use of years)* to the type of data used in the simulations and how it was used, and Reference provides the link to the publication with more model details. Further information can be found on the individual model reports available in the supplementary materials."

5.3.64 COMMENT 63

L831 Add publisher

L939 Correct doi

5.3.64.1 Response to comment 63:

Changed as suggested.

5.3.65

COMMENT 64

L1041 suggest sticking with “ingrowth” (as used in the figure) or “recruitment”

5.3.65.1 Response to comment 65:

To follow up after terminology discussion

5.3.66 COMMENT 65

L1050 Overperdicted

5.3.66.1 Response to comment 65:

Changed as suggested.

5.3.67

COMMENT 66

L1056 no mortality of recruits as they grow from 7 to 10 cm diameter.

5.3.67.1 Response to comment 66:

We use the difference of observed trees recruited between 7 and 10cm as a proxy for mortality.

5.3.68

COMMENT 67

L1077 Explain the fitted curves

5.3.68.1 Response to comment 67:

We are not sure if we are correctly understanding the reviewers point here. In figure 8 we do not show any fitted curves.

6 Other changes made in the manuscript