1. We thank the reviewer for the insightful comment and valuable assessment. In response, we have expanded the Introduction and Discussion sections to include more nuanced discussions of treeline ecology. Additionally, we revised L85–92 to more clearly articulate the study's hypotheses. As Reviewer #2 correctly pointed out, we would like to emphasize that this paper is primarily of a methodological nature, and the point pattern analysis serves as an initial application to demonstrate the approach's potential.

Response: The additional material in the introduction and discussion is an improvement, though I would not say that it is nuanced given that it is not at all specific to the study area where you were piloting these methods. While I understand that this paper is a methods paper, you chose to try to demonstrate the applicability of the methods with a case study analysis. However, the analysis itself is lacking in direction. It is unclear how/why you chose the size categories you chose, or why they would be valid irrespective of species or any other factor. Is there a dendrochronological study that suggests that these size breakpoints represent pulses of seedlings establishing in the ecotone during periods of favorable weather? Do you have any information on the species composition of these size classes or how the distribution of trees by size class changes by study site? You need to justify/explain your size classes in your methods or in an appendix or supplementary information.

You later mention facilitation and competition, but there is no system-specific context. The reader is left with the vague idea that it might be useful, but without much concrete evidence to demonstrate this. I think that the analysis actually weakens your case that the method is useful, because it is lacking in context and direction or hypothesis. It comes across as being disconnected from the system you're studying.

You also mention in the introduction (L86-88 and L103) and conclusions (L499) that the analysis allows for ecological inference on treeline processes, but you have no data with which to verify your conclusions that competition is the mechanism behind the patterns you observed. You must clarify that your spatial analysis may be useful as a hypothesis generation tool *only*, which is still a useful contribution. That is, you can use these pattern analyses to generate hypotheses about processes, which may then be tested in the field. You must then discuss your specific hypotheses—in the plural—regarding the case study results you present (your point-pattern analysis).

2. thanks for the thoughtful comment. In the Discussion section L380-384 we meant to highlight a limit of the deep learning model which we deemed right to emphasize. We agree with the reviewer that the lack of detection of smaller individuals right in proximity of bigger ones might affect the PPA results. However, as it is also visible in

Figure 3, the detection of small trees was possible and achieved, even when in proximity of big trees. Attached below are a boxplot and violin plot displaying the distribution of omission errors of small trees in function of the distance to medium and tall trees (distance computed as closest distance between the Ground Truth data and the border of the medium/tall crown). As the graphs show, omission errors are uniformly distributed when related to medium trees. When related to tall trees there is a concentration of missed detection on small scales (average value ~9 m), however, these scales greatly surpass the distance over which a small canopy might be obscured or hidden by a bigger one (and also distance values greater that those we might expect to occur between clonal stems, as mentioned in your next comment below) Concerning the discussion of the observed patterns found, L465-469 & L470-484 already discuss possible reasons of the found patterns but we rephrased and implemented chapters 4.1 and 4.3 for better clarity.

Response: Lines 381-384 specifically state "In addition to being inherently more challenging to detect in the imagery due to their diminished size, smaller trees often present altered lighting conditions due to being partially obscured or completely concealed by taller ones (Beloiu et al. 2023; Dietenberger et al. 2023; Hamraz et al. 2017), leading to missed detections (i.e., false negatives). The problem is exacerbated in dense clusters (Vauhkonen et al. 2012), common in most of our study sites."

Your figure is potentially convincing that the small crowns are not literally being obscured by larger crowns if the imagery was all taken exactly on-nadir as described in the methods. But, if this is the case, your statement in lines 381-384 is directly contradicted by your own evidence here. So, which do you believe is true? You need to include this figure and discuss the impact of missing the greatest number of small crowns within 5-15 (or even 5-10, looking at the violin plots) meters of tall trees. How would your analysis change if the missed small crowns were included?

I can see, for example, that for most sites, the largest difference between the observed g12(r) and the expected confidence interval is in the range of the distribution of missed small crowns (~ 2-28 m, esp. 5-10 m). It is reasonable to assume that the spatial bias in these missed crowns with respect to taller trees would make the conclusion that the statistically significant relationship between the two size classes is biologically meaningful—with small trees typically found further from large trees, irrespective of species or any other factor—tenuous at best.

The fastest way to resolve this problem is to discuss the very real possibility that the finding is an artifact of missed detections, but that it may be a hypothesis worth investigating in the

future. To be perfectly transparent, you should also discuss this as a limitation of your method. As it stands, you're encouraging readers to draw conclusions from these findings and also encouraging them to replicate your work or use your dataset in the way you've suggested. It is not clear to me that this is defensible.

If you insist on keeping this case study analysis on how the data could be used and you wish to make strong conclusions about the relationship between small and large trees, I would like to see a re-run of the analysis on the gps points of the groundtruthed crowns as opposed to the predicted crowns. Then, you can compare the results and show that the results are not impacted by the bias in missed crowns with proximity to large trees. If you do this, you should still discuss the biological and ecological meaning of your findings in terms of multiple working hypotheses.

3. Thank you for your comment. Regarding the potentially missing small trees, we hope the adjustments described in the previous response address your concerns satisfactorily. As for the definition of individual trees, since our study is focused on the detection of tree canopies and builds upon that, we defined individual trees as "*individual tree crowns clearly separable from the other adjacent crowns* " (definition implemented in chapter 2.2 of the revised manuscript). In a purely remote sensing-based analysis like the one employed in this study, it is not possible to determine whether multiple tree crowns belong to the same individual (e.g., clonal stems), which justifies the definition used.

Response: The definition is fine, but this is a limitation you need to discuss. You're assuming that you have clusters of even-aged trees due to competition or facilitation, but this could be because the clusters are actually clusters of clonal stems, which would explain why they are all in the same size class. The same would apply to clusters of stems due to caches made by birds or rodents. You need to state something about the biology of the species you were mapping. You may not be able to say which canopies were which species, but you can at least discuss hypotheses with some actual system-specific context. You should use this context to revise lines 466-475. If this is only a methods paper and you're not prepared to think about your hypotheses with respect to the actual system you are studying, you shouldn't include a system-specific case study to demonstrate the usefulness of your remote sensing methods. Instead, you should focus on the remote sensing methods in this paper and put the application in a separate paper where you can devote more time to thinking through the processes you're speculating on.

Be careful – you treat your size classes as biologically meaningful, but you haven't presented a case for this. They are likely a mix of different species, and growth rates can vary widely in a heterogeneous treeline environment, so they are likely different cohorts

(ages) as well. It therefore doesn't make much sense to discuss "inter-size class competition" or "intra-size class facilitation"—it's not the same as talking about inter- or intra-species competition or facilitation. Furthermore, competition and facilitation are two possibilities, but you're very likely lumping together multiple processes that could be driving these patterns. How, for example, could seed dispersal influence these spatial patterns? If the trees are wind-dispersed, you would expect dispersal to be more abundant downwind of seed sources, and depending on interactions between topography and wind, as well as the typical dispersal distance, you could end up with small seeds flying further from parent trees, followed by strong selective pressure for seedlings to survive to your 50 cm or 130 cm threshold in protected microsites.

You need to discuss multiple hypotheses for the patterns you're seeing, and then you can promote your method as a way to generate hypotheses for processes by analyzing patterns at broader spatial scales. You absolutely cannot claim to draw conclusions about processes using these methods.

It's unclear what the height thresholds are for small, medium, and large, and whether you use the same thresholds for assessing your model performance as for your cluster analysis.

4. we agree with the reviewer that Elliot's work is indeed an example of a similar study that is comparing treeline on an impressively large geographical extent. However, L421-422 *"a study investigating such patterns over large extents across multiple sites simultaneously is unprecedented."* aimed at emphasizing the overall extent of the study areas (90 hectares), not the geographic range. We aimed at highlighting that our study sites are highly representative of the studied landscapes and no previous analysis was performed with such a level of detail on such large extents. We agree with the reviewer and with reviewer#2 that the sentence can be misleading. We hence rephrased as follows: *Several recent studies have highlighted how tree spatial patterns vary along an elevational gradient within the treeline ecotone (Garbarino et al., 2020; Jia et al., 2022; Wang et al., 2021). Other works have investigated tree recruitment at different sites at broad spatial scales (Nicoud et al., 2025), and others investigated spatial patterns on multiple sites in the Pyrenees (Birre et al., 2023). However, to the best of our knowledge, there are no previous studies that have simultaneously investigated the patterns of multiple treelines at the same level of spatial extent (90 ha) and resolution (5cm) as presented in this work*.

Response: The revision is mostly acceptable, but if you insist on keeping this statement, you should clarify that this is a remote sensing study vs. a field study. The spatial resolution

doesn't apply to the studies by Elliot and Sindewald, which were in the field. Again, Elliot's spatial extent surpasses yours. I'd like to note that this statement isn't really the strongest part of your work. You're presenting a useful remote sensing method that can support spatial analyses at large geographic extents and small spatial resolutions. Whether or not you were the first to do this exact version of this type of spatial analysis at treeline with this exact spatial resolution and extent isn't as important.

5. Thank for the thoughtful comment. We modified chapter 2.3: *Our methodology consisted of the following steps: i) cropping the RGB orthomosaic of each study site into adjacent tiles of 512 x 512 pixels; ii) systematically selecting 10 tiles per each study site to create the reference dataset; iii) semi-automatic classification of tree crowns; iv) hyperparameter tuning and model calibration using a dataset randomly split into training, validation, and testing subsets; v) performance evaluation; vi) validation of model transferability through spatial cross-validation*. And chapter 2.3.1 was modified and implemented to clarify the aspect: *To generate the training, validation and test datasets, the reference dataset of 100 tiles (512 x 512) was split into 70 % of images for training, 20 % for validation, and 10 % for testing. [⋯]. The model trained in this way was used to perform predictions on the rest of the tiles to generate tree maps. However, this type of dataset partitioning does not guarantee model transferability since images from all sites are included in each phase of training, validation, and testing. Hence, we performed a spatial cross validation from the beginning to evaluate model generalizability. A k-fold spatial cross-validation was performed using training and validation datasets partitioned according to their geographic distribution. The dataset was partitioned into ten folds based on study sites. In each iteration, images from nine sites were used for training, while the remaining site's images were reserved exclusively for testing. This procedure was repeated across ten iterations, such that each site served as the test set once, thereby ensuring a leave-one-site-out cross-validation scheme*. We also changed *Figure 2* for better clarity, adding an additional panel displaying the model transferability testing as a separate phase to the rest of the workflow. Regarding the use of only 3% of our data for model training, this was not merely a test—this limited subset was indeed the full extent of the training data used, and the reported results reflect the model's performance based solely on that amount.

Response: I think I understand better what you did. I was thinking in terms of methods used to train, validate, and test machine learning models during model development including a hyperparameter experiment. If I understand correctly, you did not perform a hyperparameter experiment because you were using a model off-the-shelf. I think you should clarify this fact and describe these two validation methods as two different

experiments. In the first experiment, you tested how well the model would work when trained on only 3% of the data. In the second experiment, you tested how well the model would perform on a geographically independent dataset when trained on other geographically independent datasets. Your work is therefore comprised of two experiments testing the usefulness of this off-the-shelf model for treeline studies in terms of accuracy, as well as a case study applying the model to treeline ecology.

Follow-up comments:

1. You need to clarify your groundtruth sampling methods: What height threshold did you use for sampling trees? 50 cm? (Did you ignore all trees smaller than 50 cm? Or did you have a different threshold?)
2. You did record the species of each groundtruthed tree, so what is the species composition of your size classes by study site? What is the joint distribution of species and size class at each study site? Please include this in an appendix or in supplemental information and use this to consider how you're interpreting your PPA findings.
    a. Based on the species distribution by size class, do you see a shift in species composition between size classes at a given site? Or are the classes mixed by species? (Could there be a shift in the species establishing at the site over time due to climate change? Or, perhaps, could different species recruit with greater frequency to different microsites? Could the pulses of recruitment of different species be related to climatic trends?)
    b. If there's no shift in species composition, consider dispersal mechanism and growth patterns. Could you be seeing smaller trees further from the large trees because of high winds at sites dispersing wind-dispersed seeds further away? Could seeds be getting trapped in terrain features that happen to be at a distance from the larger trees? Could animals be caching the seeds in open areas for easy retrieval in the spring because snow melts sooner? Could each cluster of tree crowns actually be clonal stems of the same individual, each with a separate crown? This last would explain homogeneity in size (and species, if applicable).
3. You need to clarify in your PPA methods whether the size classes defined in lines 247-248 were the same used in your PPA. (You didn't, for example, use the 50 cm breakpoint, correct?) You need to explain these size class breakpoints – why did you choose them? Do they mean the same thing across all study sites given that the study sites have different conditions and sometimes different species? Why did you

ignore the medium trees when you had higher detection accuracy for those? Did you also test those and not find a significant relationship? If that's the case, you should report these results as well. Was the bivariate analysis presented only looking at the relationship between small and tall trees? (Clarify this in the caption of Figure 6.)

4. In line 476, you state that your findings with respect to small and tall trees suggest inter-class competition and are "in line with previous studies findings (Carrer et al. 2013)". Please elaborate on what Carrer et al. found and how this field study supports your conclusion for treeline ecosystems in the Italian Alps.

Final copyedits:

1. When you discuss "chapters", you actually mean "sections". The term chapter isn't generally used in this way and applies to sections of a book.
2. Figure 4 should be revised so that you can distinguish the categories when a paper is printed in black-and-white. Please use different symbols for the mean of each class (rather than diamonds for all) to achieve this.
3. In line 352 of the Figure 5 caption, change "black continue line" to "solid black line".