Responses to Reviewer #1: RC1

Thank you for the valuable assessment and insightful comments to the manuscript. Below, you can find the answers and explanation to all points you raised. All comments were incorporated in the revised manuscript. On behalf of the authors,

Erik Carrieri

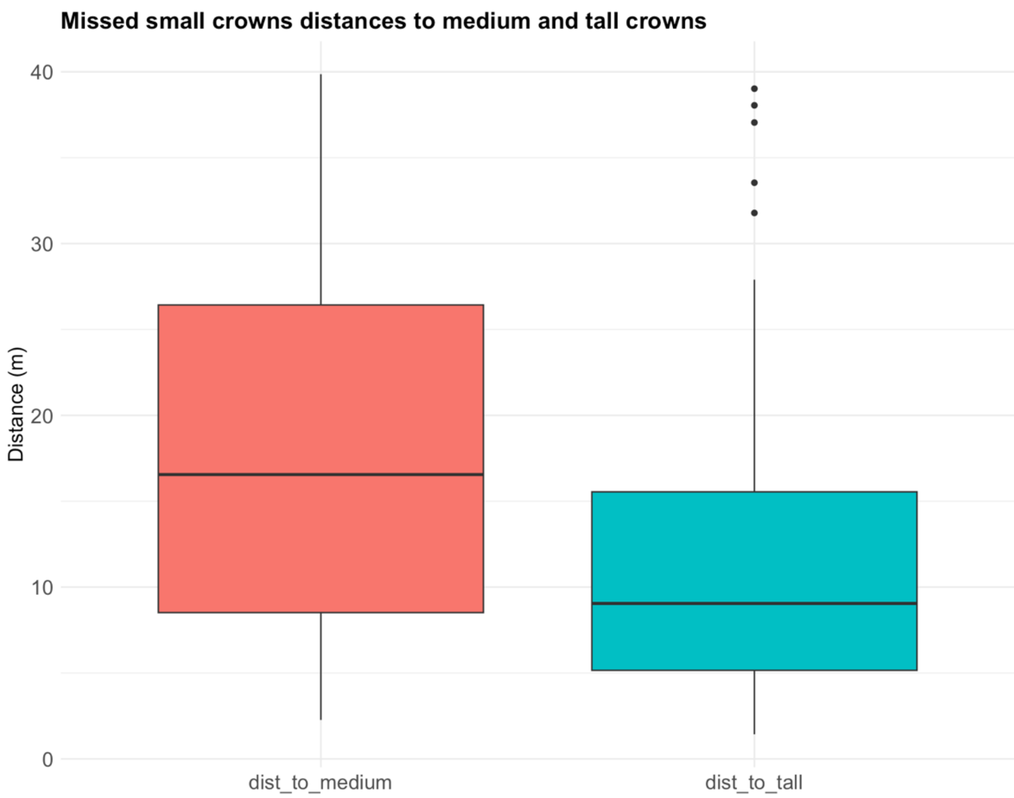
General comments:

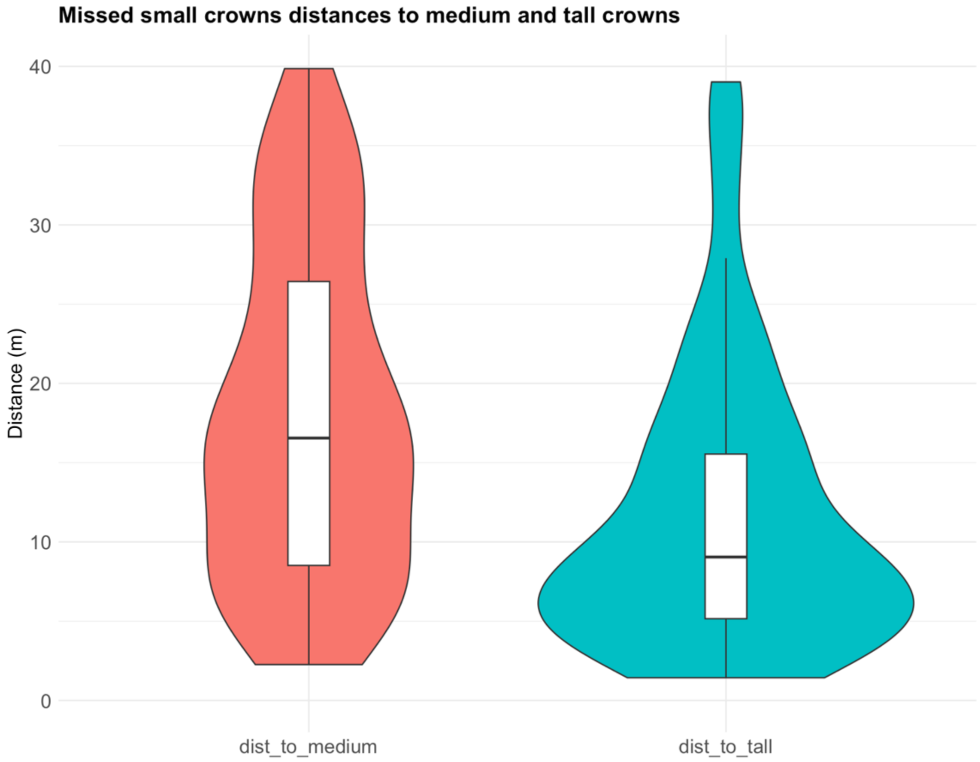
Reviewer#1: Your work is important and will likely be very impactful! You demonstrate the effectiveness of using UAV imagery in combination with pre-trained deep learning models for 1) detecting and delineating tree crowns in the ATE, and 2) estimating tree attributes (position and height). Your field dataset is impressive, covering a wide geographic area of the Italian Alps that is representative of heterogeneity at multiple scales and with respect to important climatological, biological, and topoedaphic variables.
Response: We thank the reviewer for the positive evaluation and the careful read.

However, your work could be elevated with more nuanced discussion of treeline ecology. You make some general statements in the introduction and conclusion about facilitation and competition, but you do not discuss important nuance related to 1) the degree of stress in the system (especially wind-stress), which can lead to a predominance of facilitative interactions; 2) whether any moisture or nutrient limitations are known to exist in your system that would lead to a predominance of competitive interactions; 3) the species composition of your sites and any important biological factors related to the species present, such as relative tolerance of known stressors at the adult and seedling stages, dispersal modality, and growth rates (or any evidence of age-size relationships in your treeline systems); and 4) anything known about the spatial patterns of variables related to the suitability of sites for colonization, such as distribution of soil characteristics, snowpack, or shelter. You need to discuss the relevant ecological literature in your introduction, use it to inform your hypotheses about the spatial analyses you conduct (which are also missing from the introduction), use it to justify the size classes you delineate in your spatial analysis methods, and finally discuss your results within this ecological context in your discussion.
Response: 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.

Reviewer#1: Secondly, you need to revisit the size classes you use for your point-pattern analyses from the standpoint of data reliability. In your discussion, you make it clear that your model’s detection of small trees is biased based on the proximity of those small trees to larger trees. You specifically state that you are more likely to miss the small trees that are closer to large trees. Yet you did a bivariate analysis that relied on accurate detection of small trees to see if they tended to occur close to, or far away from, large trees. Your finding that the small trees tended to be located further from large trees than expected from a relaxed-random distribution could simply reflect the bias in your dataset with respect to the detection of small trees. You absolutely must demonstrate that this is not the case to justify the conclusions you drew based on this analysis. Furthermore, as you discuss the findings of the spatial analyses based on your remote sensing dataset, you must frame these as hypotheses of process based on the observed patterns. There are multiple

possibilities that could explain the observed patterns, with competition and facilitation being among them. I listed many papers that you may find useful for adding this nuance.
Response: 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.





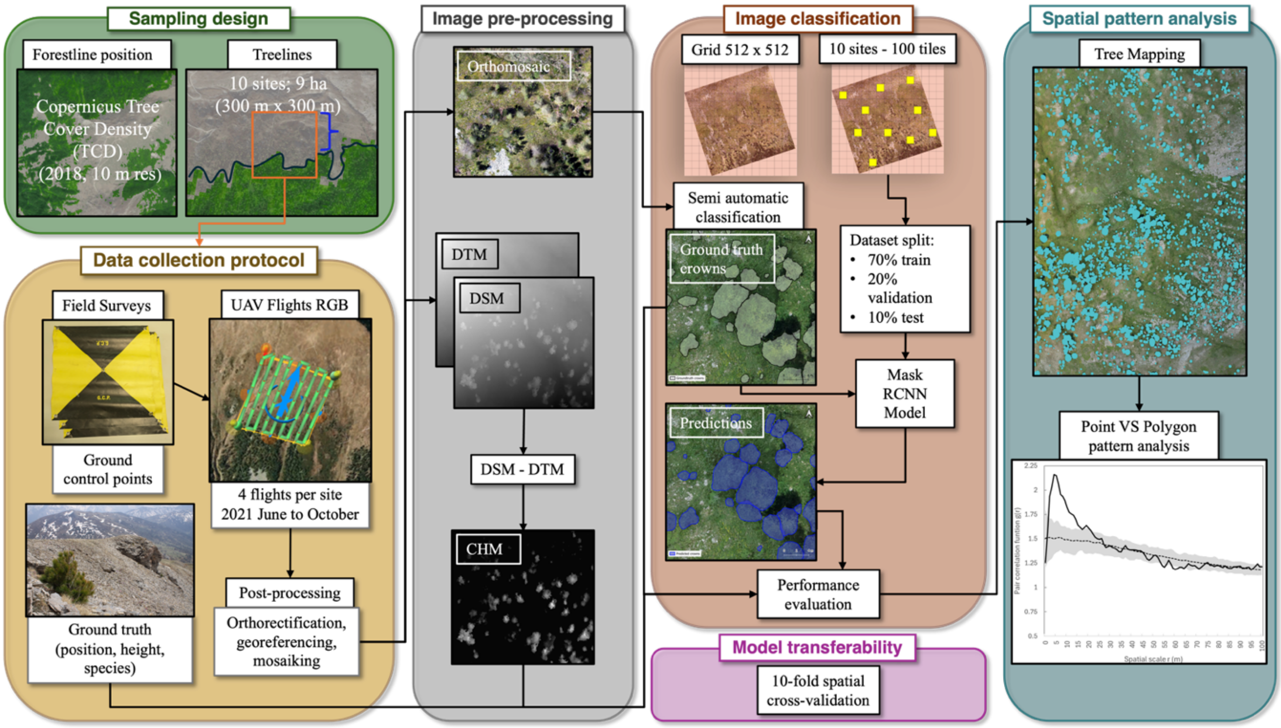
Reviewer#1: I want to emphasize that the above does not devalue your study overall, and I very much agree that UAV data fill an important gap in the translation of field data and small-scale processes and patterns at treeline to larger-scale patterns (and potentially processes). Your finding that trees tended to be clustered at scales less than 20 m is very interesting, and seems sound despite the potentially missing small trees. (The opposite result would be less justifiable, given that it could be due to missing trees.) This finding also makes sense based on what is known of processes at treeline (including, but not limited to, facilitation). However, previous spatial analyses of tree patterns in the ATE, done by Elliot et al. (2010), found spatial randomness. Sindewald et al. (2023 – dissertation publication embargoed until Sept. 2025) also found spatial randomness. I would ask that you elaborate on your definition of individual trees. Many conifer species reproduce clonally at treeline, and the stone pine is dispersed by the European nutcracker, which caches seeds. How did you determine which were individual trees and which were clusters of clonal stems? If your cluster analysis was based on discrete "canopies", but those canopies are actually different crowns of a single tree, it would explain why you saw such high levels of clustering at smaller scales as well as why your clusters tended to be of similar sizes.
Response: 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.

Reviewer#1: I would also like to note that Grant Elliot's work comparing spatial patterns across treelines predates your work and compares treelines spanning a greater geographic area (~600 km between the Medicine Bow range and the Sangre de Cristos range). My condolences, but you will

need to walk back your claims in your discussion. Instead, compare how your spatial analysis methods differ from Elliot's and, potentially, why your methods would be better as the standard to use for comparison across treelines.
Response: 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.*

Reviewer#1: Regarding your model evaluation, I think you need to add clarity to how you divided your data for training, validation, and testing. It sounded like you were using 70% of the data for training and 20% for validation, then later using those same data within your cross-validation. If this is the case, the results of your cross-validation of the model would not be a true test of model generalizability because the model would have already seen those data during training. Usually, researchers **either** divide their data into training, validation, and testing sets, **or** they do cross-validation and reserve a geographically distinct dataset for testing. I do not understand why you have done both. You can speculate that your model will generalize well based on the variability represented in your dataset, but *if I am correct* that the cross-validation folds included training data the model had previously seen, you cannot use that evaluation to draw conclusions about model generalizability. At the beginning of the section, you also state that you tested the effectiveness of training the model with only 3% of your data (given that the model was pre-trained, off-the-shelf), but unless I am mistaken, you do not report the results of this test. (It would be useful to know how this worked out!)
Response: 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.