The authors thank the referees for their thorough review, and the positive comments made on our work. The comments are addressed below one by one and we hope that the referee will find our answers satisfying. The changes are noted in blue for referee #1 and green for referee #2 in the revised version of the manuscript.

Comments of the anonymous Referee #1

Line 240: not clear what topology of the images means here. I would suggest removing "given the topology of our images" and keep the sentence as follows: "Therefore, image rectification was necessary…"

Done

Table 4: please, provide the details about the applied test of correlation (also in Line 433) and explain what the bold values mean (significant correlation?) in the caption.

The correlation test was applied on 8 different vectors, each represents on key parameter. It is described in lines 433 to 435 as: "we first tested the correlation between the factors identified in the previous section by calculating each one of the eight parameters for all detections as one vector and then calculating the correlation between each pair of parameters".

To clarify the table, we add "Values in bold show significant correlation." In caption of the table and "(the bold values)," in line 435.

The same for Figure 10 and the related text (e.g., Line 516)

Thank you for this comment. Here we used the length distributing of TPs and TP+FN as two vectors and calculate the correlation coefficient between them. Therefore, to clarify it, we add "The correlation coefficient between them the length distribution of TP as one vector and TP+FN as the other vector was calculated" to the line 516.

Line 523: virtual? I would simply say that missed pieces were estimated.

The word virtual is removed from the text. Then, for more clarification we add an additional sentence: "…, note that these pieces were imaginary pieces inferred from the wood length distribution and were not detected by the software".

Comments of the anonymous Referee #2

L119: What is an example of a larger dataset? Video over 50 minutes? 1 hour?

We add "(i.e., Video segments more than 1hr)." to the text.

Figure 1e: label within figure should be "15 Dec"?

Done

L153: In reading the paper, I was confused when the camera parameters are specifically input in the processing steps. I suggest pointing to the section where these different scales are utilized in the analysis.

It is correct that using this sentence at this stage is a bit confusing, so we remove it from this part.

L192: You mention noise in the image related to water waves, but is vegetation in the image also considered noise here? Please clarify.

We add "or surrounding vegetation" to line 404 after surface water waves.

L323-324: Do you have a sense if annotating for occurrence, objects, or pixels produces more accurate results? Which would you recommend for a first-time user with their own data? To get more wood pieces or more of the same piece over time?

It depends on the scope of the study, for example, if the statistics of wood pieces is important, more pieces result in smoother statistics. If we want to train software (e.g., what happens in AI) or if the dynamics of wood is important for us, it is necessary to trace the trajectory of wood by multiple detections of same pieces. Here, we only wanted to show that the annotation module able to records pieces in different modes.

L405: Change sentence to "Both of these conditions are noted in Figure 8a."

The darker patches are actually visible in all three figures (8, a, b, c) but it is highlighted only in one of them (8, a), and what is mentioned in the text is "Both of these conditions can be seen in **Error! Reference source not found.** which is highlighted in **Error! Reference source not found.**.a." which is correct.