Dr Christopher Tomsett
School of Geography and Environmental Science
University of Southampton
Southampton
SO17 1BJ

26th June 2023

ESURF-2021-102: Exploring the 4D scales of eco-geomorphic interactions along a river corridor using repeat UAV Laser Scanning (UAV-LS), multispectral imagery, and a functional traits framework.

Response to Comments from Reviewers

Dear Lina,

We thank you and the reviewer for your constructive and helpful comments on our submitted manuscript. Below we outline how we have addressed the reviewers overall feedback and the individual comments. For clarity, we summarise the review comments in ***bold and italics*** and list our response below each one:

Reviewer 1:

The revised manuscript is a tighter, more focused contribution to the understanding of the coupled evolution of river-corridor vegetation and geomorphology. The document is now exclusively focused on the two-year period when detailed high-resolution surveys were performed, as well as the ~1 km reach where they were obtained, thus providing an opportunity to examine vegetation-sediment feedbacks in real time.

Unfortunately, however, this revised focus brings into sharp clarity several underlying weaknesses. The manuscript describes itself as an "exploratory analysis" on Line 889, a term that succinctly confirms that an intriguing sequence of analytical steps was performed, but that the limitations and uncertainties of the chosen method(s) were poorly characterized, and as such the outcomes are challenging to interpret and use to motivate further work.

First, there remains no field validation of remote sensing metrics or guild assignments. The authors summarize in the Introduction the importance of direct measurement of guild characteristics (frontal area, flexibility, etc.), as well as how different individuals from the same species may have different growth forms at different sites, yet in the end they rely on remote sensing algorithms for these determinations, without any field validation at this site for these individuals. Classification results (which require a substantial amount

of manual attention and parameter choices) are compared to previous findings (TRY database) for these species at other locations; this is better than nothing, but (as noted above) remains an imperfect comparison. Without any field validation, all trait assessment should be considered preliminary or suggestive, rather than quantitative and authoritative. As such, it is difficult to assess whether the primary negative results in the paper (that there was no consistent pattern of erosion or deposition within different functional groups) results from lack of signal, or lack of ability to detect an actual signal.

The initial scope of our field campaigns was not to target traits based methods specifically, but vegetation interactions more generally, it is regretful that no specific traits based data was collected in the field for validation. This is something that with hindsight we would certainly do should we attempt to replicate the research elsewhere, and as such is commented on in section 6 of the paper.

As the reviewer correctly points out, the extraction of these traits is predominantly determined by remote sensing methodologies as explained within our methods. However, to add some additional estimation of the error associated with the extracted values (beyond the previous revisions which compared them to wider literature) an assessment of both the UAV-LS and QSM methods has been undertaken and included in the manuscript.

The former compares the DBH and tree height values of trees that are within both the TLS scans and the two closest UAV-LS scans temporally before and after this survey date. This enables us to establish how well the UAV-LS methods performed in comparison to TLS which is widely regarded as the benchmark in high accuracy remote sensing.

In addition, to assess the outcomes of the QSM methods, manual measurements of tree height and DBH for all four remotely sensed groups were performed on the raw TLS and UAV-LS point clouds in CloudCompare, and these were then compared to the resultant values from the QSM methods.

These datasets are then used to assess bias or variation occurring during the trait extraction process (new Figure 4). Overall, this analysis allowed us to quantify the uncertainty in the vegetation metric extraction methods and provided some additional confidence that the methods used are reconstructing vegetation well. We hope that in the absence of manually measured field validation data, the use of TLS data and this additional analysis provides some further measure of the quality of the plant characterisation techniques for the reader,

Second, numerical (Delft3D) modeling was introduced to sharpen the focus on the reach-scale implications of the various vegetation characteristics, but there was apparently no field validation of the numerical model. The surveyed flood extents during Storm Dennis provided the minimum initial/boundary conditions to run the model; there were apparently no separate measurement(s) of water depth, discharge, or velocity within the model domain that could be used to validate the numerical representations of functional group drag (Lines 461-485). Without validation, the modeling "results" (e.g., Figure 11, Table 3) become preliminary or suggestive and should not be treated as robust outcomes.

The introduction of a numerical model was made in the last set of revisions in an attempt to address the comments of reviewers in relation to interpretation of the results, and of course no in field measurements had been taken during the flood conditions during the field campaign. It was only ever supposed to be indicative, to explore the ways in which depth dependency may be influencing morphological change. We agree that the way this exploratory analysis was presneted perhaps overstated the output from the model and we have rewritten the discussion around this to tone down the interpretations. In this vein, the annual approach and analysis has been removed from the manuscript as this confuses the simple core message that we are trying to convey.

We retain an example of morphological change over one winter based on the high flow output, but have moved the presentation of this from the results section to the discussion, to help emphasise that this is exploratory analysis undertaken for indicative hypothetical flows based loosely on observed data. Taken together, these changes shift the focus from this section being treated as 'results' to them being considered 'preliminary' and 'suggestive' as recommended by the reviewer.

Third, the article title ("Exploring the 4D scales of eco-geomorphic interactions…) remains imperfectly matched to its contents. Although the 3D point cloud data are used for vegetation assessment, the land cover assessments (Table 2, Figure 6) remain 2D, with only one vegetation type assigned to each horizontal pixel; understory and overstory plants are not distinguished. Without 3D spatial analysis, this is not a 4D analysis. In addition, the analysis is only completed at one spatial scale (individual plant --> convex hulls of unspecified size but presumably a few meters across --> pixels at reach scale), so multiple scales are not explored. As the authors point out in the Introduction (Lines 35-49), scaling from individual plants to >1 km reaches remains a challenge, but not enough information is provided to conclusively claim that they succeeded here. The current analysis apparently did not even include a training (calibration) and a test (validation) data set (Lines 419-420) to assess efforts.

We have changed the title of the paper to better represent the content, accepting that the '4D' assertion is not fully justified.

We do believe that vegetation analysis has been undertaken across scales as identified in the reviewer comment, occurring at both the individual plant and reach scale during different parts of the analysis. However, in line with the suggestions we have also reduced the focus on scales explicitly, only retaining reference to this when explaining the methods (in response to the 5th comment below).

With regard to the training and calibration dataset, the decision not to split this data was based on the number of samples collected, and the fact that increasing this significantly was not an option due to the time intensive nature of manual selection and analysis. To have split a few samples off would have left too small a dataset to undertake meaningful statistical analysis. At the same time, it would have also reduced the number of samples to train the model on. Likewise, in the lines that follow 419-420, the method of random forest classification undertakes out-of-bag accuracy, the method by which subsamples are tested for each decision tree. This can be used to infer a primary level of model accuracy, and is one used widely in the literature. We have supplemented this with a test dataset from high resolution ortho-imagery and field notes to check predicted accuracy, which is detailed in sections 3.2.5. and presented in 4.2.4. and Figure 8.

Fourth, for a manuscript that focuses on remote sensing, it remains extremely odd that the Introduction does not summarize lessons learned and knowledge gaps from remote sensing of riparian vegetation.

In line with changes to the introduction and background sections that have aimed to streamline the paper, more emphasis has been placed on including some of the wider remote sensing of vegetation literature, and showing how a knowledge gap exists in relation to using these methods to identify traits at greater spatial scales. Given that the manuscript is already lengthy, we have refrained from including an extensive review of the literature, instead pointing to some important review papers in addition to the changes outlined above.

***Fifth, even though the remaining contribution of this paper is its innovative methods, numerous key analytical details remain unspecified. Although it is nice that the flight dates are clearly laid out (Table 1), it appears that results were combined into only two vegetation maps (Line 405 and Figure 6). This combination should be clear from the beginning (i.e., Table 1 and Lines 248ff). In addition, justify why wintertime data were combined with subsequent summer data, when it appears from Figure 1c that many high flows occur during late winter/spring so therefore winter vegetation may not be comparable to vegetation the following summer. Some of the classifications rely on TLS data, which were only collected once. It is unspecified/unclear when/where several of the steps (point cloud segmentation, trait metric extraction) were performed, and whether this occurred for repeat scans or not. (For example, were all 24 individual tree segmentations performed from data collected on the same date, or multiple different dates?) It is unfortunate that the authors wrote in their response to comments that they consider this level of detail difficult to provide, since it remains impossible to understand, duplicate, or apply these methods without understanding what was done here*.**

The methodological approach we used was complicated and it is clear that the overall workflow was not communicated clearly enough within the manuscript. Due to the (not always linear) methodological workflow undertaken to go from the initial field data to final extracted metrics, it is challenging to convey how each section links together, especially when in addition the processing is sometimes performed on different datasets at different scales.

We have attempted to improve the clarity around the methods in three ways in the revised manuscript:

1) Introduction of a new workflow figure to help readers understand which surveys were used for each different section of the methods (new Figure 2), and also to show how each of these processes work together to progress from raw data to the final extracted values. The flow diagram also includes reference to numbered sections and figures in the manuscript to help the reader navigate through when these processes occur.

2) We have carefully worked through the manuscript and introduced explicit use of the terms plant, group and reach scale, to make it clear whether the analysis is undertaken on a plant level basis, or across the reach. This also helps when referencing methods in the discussion. Plant and reach scale data have also been included in the flow diagram referenced above to help better direct the reader.

3) The classification approach that we used requires both leaf on and leaf off data, and is justified due to the analysis of reach scale characteristics of each group. This meant that we necessarily needed to combine winter and following summer data. We have commented on this point in both the methods (to justify the decision) and future work section (to discuss it as a limitation). We have also tried to better present and explain the overlap in the timings of different surveys, and how they feed into data collection and analysis. Specifically, in the methods section, we now refer to the month and year of each survey and, where required, point the reader to the survey (Table 1) to which we are referring.

Sixth, the authors' response to comments repeatedly mentions the importance of conciseness, yet the authors' attention to this important consideration seems inconsistent. There are numerous places where excess information is included (e.g., overly authoritative discussion of bank erosion on Lines 66-74, overly loquacious discussion of speculative findings on Lines 649-729, overly optimistic Discussion on Lines 730ff), while the authors omit key information that describes what they did or justifies why it makes analytical sense.

In revising the manuscript we have taken on board the reviewers comment that parts of the text are extraneous and have made extensive edits (mainly through reduction and removal of text) to address this. We hope that the changes outlined above to the methods section, and edits to the results and discussion sections, have helped to improve the manuscript by making it clearer what was undertaken when, why it was done, and the resultant outputs. We have also reduced the length of the introduction and discussion sections in order to compensate for the addition of the extra methodological detail requested. This has led to an overall reduction in paper length, despite an increase in the length of the methods.

A few remaining specific comments:

Line 232: The site description states there are "two distinct reaches," which are not shown on the map, not obvious from the map, and apparently not discussed anywhere else in the manuscript. Elsewhere, the word "reach" seems to be used consistently to refer to the entire study area. It is recommended that different terms be used to refer to the upper and lower portions of the study reach.

This was an oversight in the transition from the original manuscript which contained the long term analysis of the study site, where the upper and lower reaches were referred to more explicitly. As suggested, the term reach is altered to refer to the entire study area in line with the rest of the manuscript, and the upper and lower portions of the reach referred to as sections.

Line 581: Unclear how the "over classification of shrubs" was assessed, since there was no independent validation of the ortho-imagery. In addition, unclear where this overclassification occurs on the landscape (Figure 6).

Attempts to clarify this statement have been made in reference to the spatial pattern whereby areas classed as trees consistently have surrounding classification outputs shown as being shrubs. This is especially the case for isolated trees which have small classification of shrubs, and are probably the result of image segmentation cutting out lower portions of the tree structure more so than a small area of shrubs being present. However, it is acknowledged that use of the terminology ‘over-classification’, rather than a more qualitative description as is now presented, did not well present the narrative that we were trying to convey.

Line 613: It is claimed that “single branching herbs” were misclassified as “branching herbs.” This is a confusing statement. Be consistent with functional group names throughout the text, and perhaps avoid two names that are this similar.

We acknowledge the confusion caused here which is due to incorrect inclusion of the term branching in single stemmed herbs. The naming of these two groups has been checked throughout the manuscript to avoid any inclusion of the term branching when referring to single stemmed herb groups. The figures have also been updated to increase clarity in reference to each of the individual groups.

Finally, the manuscript is plagued by numerous punctuation mistakes (e.g., Lines 173, 221, 288) as well as spelling issues (e.g., Lines 239, 584, 649, 666, 955, 960, 961). In addition, it is very difficult to read the text on some axis labels (e.g., Figures 8 and 9).

The specific mistakes in the text highlighted above have been corrected, as well as several others found whilst reviewing the manuscript. Text in figures 8 and 9 (now 9 and 10), has also been increased in size for histogram subsets, and for figure 9 (now 10) we have also attempted to reduce the amount of information present in the figure by removing repetition of phrases such as net volume change and the volume change axis labels to only appear once.

We hope that the editors and reviewers are happy that we have fully addressed all comments and are happy to clarify any responses and changes as required. Thank you for considering our manuscript for publication in Earth Surface Dynamics.

Dr Christopher Tomsett
School of Geography and Environmental Science
University of Southampton