Response to Reviewer #2

In my view, after multiple rounds of revisions, the author has addressed all my concerns, significantly enhancing the quality of the paper. It now meets the publication standards, and I recommend accepting it for publication pending the journal's technical review.

Reply: We thank reviewer #2 for the positive feedback.

Response to Reviewer #1

I'm sorry for being so strict and may make the authors frustrate, but I hope this ensures the publication of high-quality data.
I can now accept the situation, given the challenges in obtaining independent training labels for the entire terrace. Although there is no independent accuracy assessment, a thorough discussion of the issue, including the potential caveats and limitations of using the product, would be necessary.

Reply: we agree that an independent evaluation is necessary and we have now tried to accommodate for that by defining an additional independent set of polygons. We used these polygons and another independent dataset (50 sampling points) with defined habitat types, to evaluate the accuracy of both, the central and the entire Lena Delta. As expected the accuracy is lower using the independent data points (~90% for the central Lena Delta and ~80% for the entire Lena Delta classification). We agree that this is a more realistic quantification. For the entire Lena Delta and notably for the smaller habitat types this might still be an overestimation. We discuss this in section "4.4 Classification accuracy and representativeness".

However, I would still being strict on this:
"Also, if the habitat class can be determined based on this expert knowledge, why not determine extra polygons (or pixels) for independent validation? Or alternatively, if you must split the exiting ESU and expert-identify polygons for training/validation, why not sample by polygons instead of pixels to eliminate the auto-correlation problem. Lastly, I still think a class-based accuracy matrix is needed. An overall accuracy metric is insufficient to suggest good classification from remote sensing images. Reply: This is what we have done. We have now added a class-based accuracy table (Table A1)." No, this Is not what you have done (you randomly sampled 50% pixels, according to the text, meaning pixels from the same polygon are both in training and validation set). Right now, the accuracy and precision are ridiculously high for each class (I would not say this is not possible, but rare) due to the reason you're training and validation all from the same polygon (autocorrelation). What I mean: if you have 216 polygons, you could use 70% of the polygons (all pixels in each polygon should use exclusively for training or validating) for training, and 30% for validation: no training and validation pixels would come from the same polygon. I hope this is clear.

Reply: As discussed above, we have now defined additional and independent polygons, based on expert knowledge. In addition, we found a dataset (Siewert et al. 2016) with independent sampling locations. Those were made for permafrost coring, but included vegetation description that could be linked to our habitat classes. We added text in '3.4.3 Central delta habitat classification' and provide details on accuracy/validation in FigureA4 and Table A1-2, Table S8, Figure S6.

Minor:
Line 275: what is BH?
Reply: We removed the sentence polygons defined by BH.

Figure A4: Adding scales.

Reply: done.

A confusion matrix should include both the user's accuracy and producer's accuracy for each class, as well as the number of samples of each class used for assessment. Refer to Dr. Robert Gilmore Pontius Jr work on "Best Practices for Classification Accuracy Metrics".

Reply: We added the statistical output and class based evaluation in Table A2, using the function confusionMatrix from the R Package caret, an established and frequently used method and output.