

Response to Reviewer #2

Accepted as is.

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

Response to Reviewer #1

Thanks to the authors' revision, the manuscripts have been significantly improved and clarified. However, I still have some concern about the habitat classification of the central Lena delta, and big concern to the whole Lena Delta Habitat classification.

For the central delta habitat classification, the training/validation is based on pixels distributed in the 26 ESUs and additional polygons defined by the expert knowledge. Here: how many additional polygons have been defined, based on what expert knowledge or tools/approaches.

Reply: In the first revision, we added Figure A3 showing the ESUs and the defined polygons.

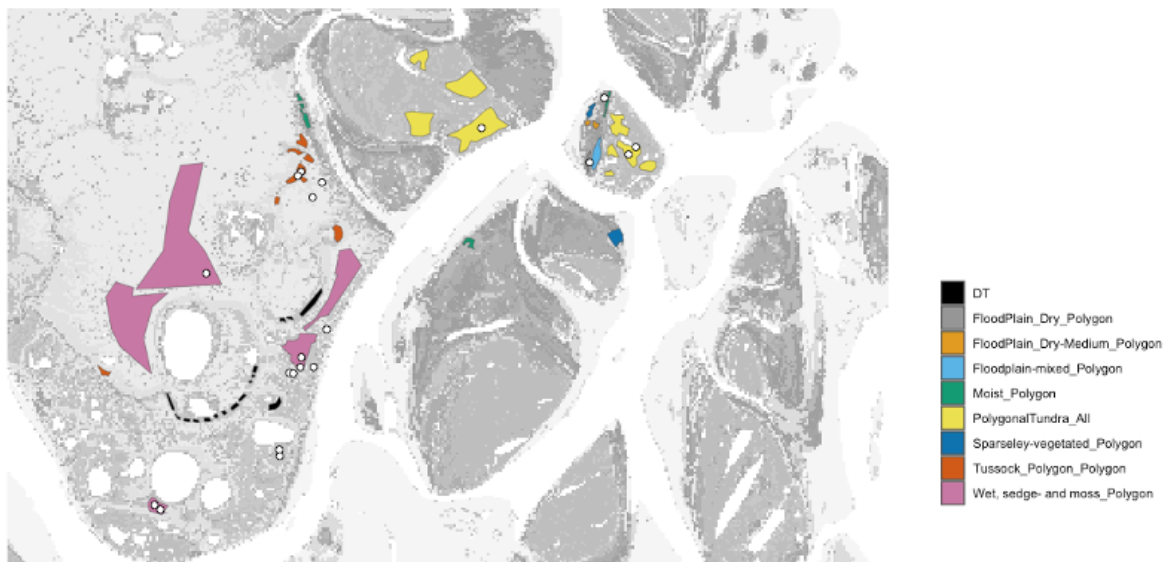


Figure A3: Subset of the central Lena Delta with 30 x 30 m ESUs (white points, dataset 1) and polygons defined by expert knowledge (published with dataset 4). Together the ESUs and polygons served areas to sample 8 626 training pixels for the central Lena Delta landcover/habitat classification (dataset 4, Landgraf et al. 2022a).

We added more information on expert knowledge and methods (L271-276).

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).

For the habitat classification of the whole Lena Delta, while I understand it is difficult to validate, this could not be the reason without independent validation. Something must be done. The authors mentioned expert knowledge here again, but still vague. Results were carefully checked with expert knowledge, but what is the result of this check, and how it is checked. Is it possible for experts to interpret manually from very high-resolution satellite images, or aerial photograph, or drone flies at some sites for accuracy assessment? I mean, there should be some quantitative validation metric beyond the central Lena Delta, especially when the central Lena Delta only represented 2% of the vegetated area of the whole Delta. Again, I would not accept an overall high accuracy when there is multiple classes.

I would not recommend publishing the datasets or paper if these critical concerns are not well solved.

Reply: We partly understand the frustration. We were fortunate that only one major habitat class, that covers large areas on the second terrace, was missing in the central lena delta. As you can see in the new supplementary figure S7, it is rather straightforward to select suitable training pixels. Creating additional evaluation would work well for such habitat classes that are less patchy and cover large areas, but even with our expert knowledge from years of work in the central delta, we would not dare to define pixels for the smaller patchier habitat types outside the area we know. However, we have included the confusion matrix results of the classifier for the training dataset (Table A1). We have also added three comparison figures in the supplement showing Sentinel-2 RGB and the results of the classification. We sincerely hope that you agree that within the terraces, the habitat structures are largely repetitive (driven by geomorphological features) and that the classification very precisely picks up these repetitive structures and differences between the terraces (new Figures S6-S8). We agree that a lack of formal and more spatially complete evaluation limits the trust in the habitat classifications, and more so in the patchy smaller habitat types rather than the larger distributed types. Besides the inclusion of the class-based accuracy, we added a few sentences regarding the limitations due to the lack of evaluation datasets across the Lena Delta (L586-590).

Minor: Spell S-2 as Sentiel-2

Reply: Done

Unaddressed comments from last review:

Dataset 1: The vegetation cover was recorded or measured at the center of each 30×30m plot with a ring of 50cm, and then scaled up to the whole plot. How is this done? And how floristic composition played a role in this process. Would the 30×30m plot include more vegetation species than the center 50cm-radius subplot?

Reply: (L195) we defined a 30 x 30 m square plot with a homogeneous or repetitive vegetation composition that was also representative of the wider land surface serving as an Elementary Sampling Unit (ESU).

We specifically selected homogeneous vegetation plots with the aim to enable upscaling.

Dataset 2: Again, I think more information about the scaling to the 30×30m plot is needed, and why it is reliable.

Reply: See reply above.