Reviewer #1:

Arctic wetland ecosystems are vulnerable to climate change, but there are still not a few datasets in the region to study the feedback between these ecosystems and climate change. This study provides a few useful datasets about the vegetation in the Lena Delta, which is a unique region. Particularly, the detailed habitat type map will potentially be the most valuable product provided by this study, which, unfortunately, raises me a few concerns about the quality.

The download links for these datasets provide data in tab-delimited text, I'd suggest also provide in other more readable format, such as excel-readable csv, and particularly for maps, in the format of geotiff or other commonly used image format, and a preview in map form (currently, the preview is not working).

- The datasets were uploaded based on the defined and required formats of the Data Publisher for Earth & Environmental Sciences PANGAEA. However, we understand that the format might not always be intuitive and also the downloading options and processes are sometimes confusing. We have therefore expanded the data availability section in the main text to better describe the file formats (e.g., tab-delimited text files for Shevtsova et al., 2021a, Shevtsova et al., 2021b, Runge et al., 2022, and the training data in Landgraf et al., 2022b, geoTIFF and shapefiles for the other datasets). Please see the revised text in Line 601ff.

  In addition, we added an example (with screenshots) on how to download the data in the supplementary material.

Line 155, 162-163. it is confusing here about homogeneous vegetation type. It is homogeneous at what level? It is supposed to only have one "habitat class" as defined in dataset 4 or what? Also, it is stated that in line 155 all 30×30m plots are homogeneous vegetation types, then later in lines 162-163, there are different methods applied to the plot if vegetation cover was homogenous and if vegetation cover is more diverse.

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?

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

Figure 2: Can you label the length of edges of these squares, so we may immediately understand which one is the main 30×30 m plot, and which are smaller-sized subplots? I could not tell which were 2×2 m plots and, which were 0.5×0.5 m plots.

- We understand the confusion and revised the text as well as the figure captions. Specifically, we provide a more detailed description of Dataset 1 and 2, and provide a revised Figure 2 according to your comments and to better highlight and describe what is meant by 'homogeneous' vegetation types. An additional Figure 3 was produced for Dataset 2, showing more examples of the different plot types and methods applied in the field (as well as the spatial scales for biomass sampling).

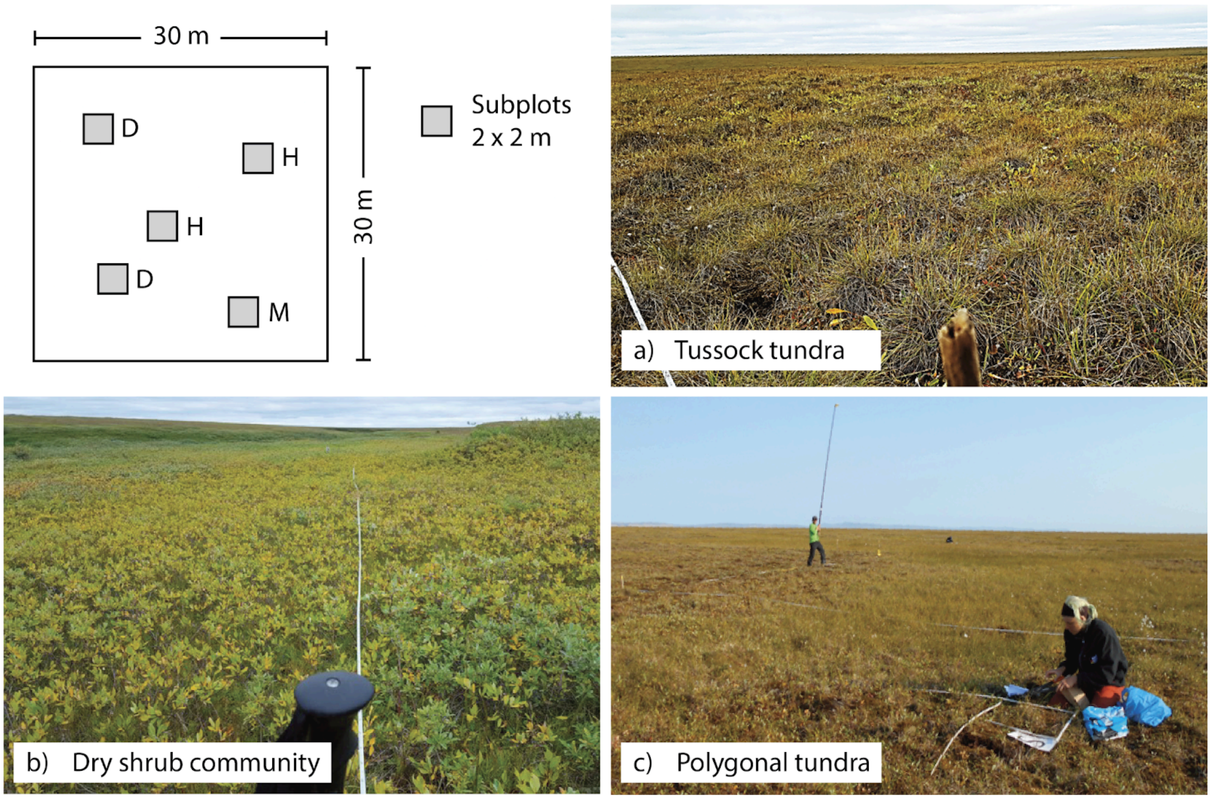

Figure 2. Vegetation plots (30 x 30 m) were established in different vegetation types across the central Lena Delta. For subplots (2 x 2 m), the projective vegetation cover was recorded and labelled according to vegetation and soil properties (H-Type: homogeneous, M-Type: moist, D-Type: dry). Figures illustrate example plots in a) tussock tundra (VP14), b) dry shrub communities (VP05), c) polygonal tundra (VP13). Photos: AWI.

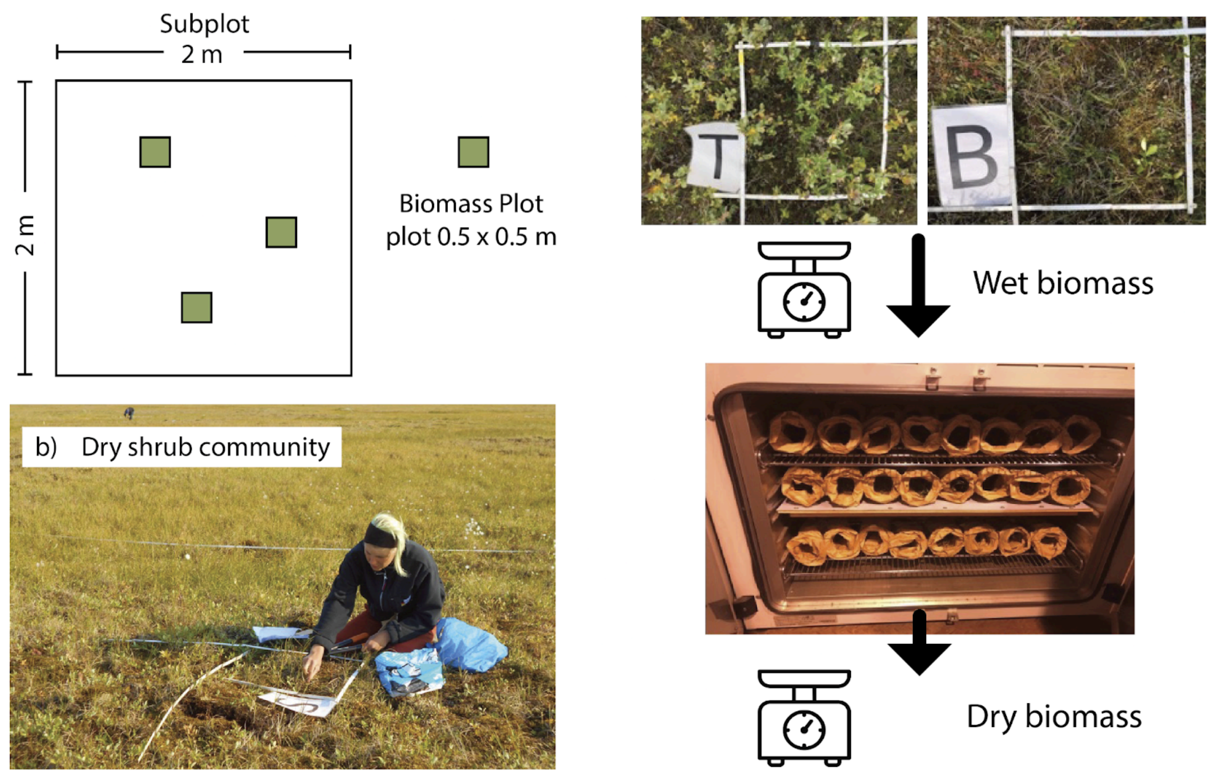

Figure 3. Biomass was sampled in subplots of 0.5 x 0.5 m (and 0.1 x 0.1 m for moss and lichens) distributed within the 2 x 2 m subplots described in Figure 2. Pictures illustrate biomass sampling in b) polygonal tundra c) high shrub communities d) shrub and sedge communities e) dry shrub communities. f) Collected plants were weighted (wet biomass), dried in an oven and again weighted (dry biomass). Fotos: AWI.

Dataset 3: Does each 30×30m plot only include one homogeneous land cover? Are these the same plots as datasets 1 & 2?

- Yes. With the revision the relationships are better described in the text. In a nutshell, the 30x30m plots only include homogeneous (e.g., dry tundra) and quasi-homogeneous (e.g. polygonal tundra) land cover types. We collected 28 hyperspectral field measurements, 15 of them within the same vegetation plots described in Dataset 1 and 2 (please see supplementary table S2 for an overview). Furthermore, we revisited plots later in August to capture senescence. In addition, we also conducted 10 field-spectroscopy measurements on non-vegetated areas such as sandy parts of the floodplain.

3.4.1: what is a S-2 based supervised classification?

3.4.2: what is S-205 2A Level 2A image? This section, without pre-information of the sensors, makes it hard to understand the different types/levels of images.

3.4.2: all spectral bands were resampled to 10m pixel resolution bands. Which bands are 10-m resolution, and what spectral bands does the sensor have?

- We added additional information on the Sentinel-2 satellite dataset in the text (L105 and copy pasted below), including information which bands have a 10 m and which 20 m pixel resolution. We also removed the too detailed term S-2 Level 2A (referring to the atmospherically corrected Bottom of Atmosphere (BOA) S-2 product level).

  "The availability of Sentinel-2 (S-2) Multispectral Instrument (MSI) data from two orbiting satellite missions since 2016 and 2017 provide high quality multispectral satellite data with a spatial resolution in the Visible and Near Infrared wavelength region of up to 10 m, and of 20 m in the red edge and the Short-wave infrared wavelength regions (Drusch et al., 2012, ESA 2015)."

3..4.3: what is the distribution of the 8,626 training pixels? Are they scattered in the classification area domain? If they are clustered or formed from polygons with the same class, the efficient number of training pixels will be much reduced. A map of the training/validation samples can help understand the situation.

- We understand the concern and the problem of autocorrelation with the method that was applied for the central Lena Delta classification. Not as a justification but to provide background, the already published dataset was developed as part of a bachelor thesis (Nele Landgraf). In the thesis, NL and BH tried to tune a classification model for a region where both (and also additional authors) have very good on-ground knowledge of vegetation types, communities and structures. With the final choice of training points the classification produced a highly accurate map separating the different habitats that are known from fieldwork and remote sensing images. We are aware of the methodological issues, however, given the qualitative assessment of accuracy based on expert knowledge, we considered this product valuable and useful as a training dataset for the entire Lena Delta (for which expert knowledge is very sparse and vegetation monitoring is lacking). In our revision, we added a map showing the training pixels and described the method (appendix), as well as a description of the qualitative nature of the evaluation procedure in the dataset discussion (L275) and a new specific section (L564).

What is ESUS?

- ESUs are Elementary Sampling Units that serve as spatial training and validation units representative for the land surface for quantitative and qualitative remote sensing operations. This process is in accordance with the Committee on Earth Observing Satellites Working Group on Calibration and Validation (Duncanson et al., 2021). Description and reference can be found in the text (L193).

Line 224: A detailed user's and producer's accuracy report as well as the overall classification accuracy is needed here to justify a good classification. The validation samples should be "independently" distributed from the training sample: i.e., they should not come from the same plot/polygon, because of the well-known auto-correlation problem.

Line 229: it seems you only have 26+69 training locations. Which is quite a small training sample. As I mentioned, counting the number of pixels in a continuously distributed polygon is misleading.

- For the training, S-2 pixels from the 30 x 30 m ESUs representing the 26 vegetation plots (Dataset 1, Shevtsova et al. 2021a), and additional training polygons defined by expert knowledge were used and are all published in the Landgraf et al 2022a data collection. To understand the class and feature performance we applied a cross-validation. As you correctly point out, the overall quantitative class accuracy of 96.78% is this high due to the autocorrelation of neighbouring pixels in the 9-pixel size ESUs (3 x 3 pixel size) and neighbouring pixels within the training polygons. The main quality assessment is not considered by the quantitative assessment but by the qualitative assessments that we kept during the process of the classification, interacting with colleagues who had been in the field in the Lena Delta for many years and are in the list of co-authors. The chosen classifier was able to distinguish all relevant classes and was even able to identify known patchy habitat spots. Therefore, we consider the central Lena Delta classification as reliable.

  See new Appendix.

3.5.2: how large is the area, e.g. at least how many tiles of sentinel-2 images are required to cover the whole study area?

- We added both information. The Lena Delta covers an area of 29873.7 km^2 and 15 S2 tiles overlap with the extent.

3.5.3: It is good you have 6,500 random points now for training the classifier, but still bad that they all come from the central Lena Delta. Some evidence or proof is needed to show that "reflectance" from other areas is like those in the central data when they are the same class, regarding the complex "Same Object, Different Spectrum" and "Different Objects, Same Spectrum" problem. What is the relative size of the central Lena Delta compared to the whole Lena Delta?

Using the training data from just the central Lena Delta to train a classifier to classify a large area is acceptable, but validation the result still using the location from central Lena Delta is not acceptable at all. This is because the accuracy of the classification of a class is dependent on its dominance, which varies by region. The training samples should be evenly scattered out to the whole classification area domain. A map of training/validation points is needed.

- We understand and fully agree that this would be the optimal procedure. However, there are simply no validation points from outside the central Lena Delta. Large parts are not accessible (except by helicopter) and have never been visited by humans. Therefore, the classification method we applied is, to our knowledge, the best way to derive a good quality habitat classification for the entire Delta. We added a discussion on how that lacking in-situ data from other parts of the delta is obviously a caveat of the classification and that quantitative assessments are therefore difficult and in this case largely impossible. We also added the relative area of the training dataset (central Lena Delta) in relation to the entire Delta (Central Lena Delta 644.9 km2 (Vegetation cover 55.2%), Lena Delta 29873.7 km2 (Vegetation cover 58.7%); Vegetated area of the central Lena Delta represented 2% of the vegetated area of the entire Lena Delta).

Line 294, can you provide more details about how the upscaling is conducted?

- Details were added and linked to Figure 5 showing the disturbance map (Dataset 6).

Line 299: here you have it named habitat map, but in the data link you call it land use land cover map. I understand they are the same thing but need to be consistent.

- Lisovski et al., 2022 is in the process to be renamed on PANGEA to have consistent naming (Habitat Map).

Line 306, it is not clear where/what are the hierarchical level 1 in Figure 3a.

- The hierarchy refers to the dendrogram in Figure 3 (now 4) a. Level one is the baseline and each split in the tree is another hierarchical level. We added an explanation in the text.

Figure 3: Why do you use surface reflectance for the classification for producing datase4, but then top of atmosphere reflectance for dataset5? It seems to me it is hard to spectrally differentiate the different habitat classes looking at (b), makes me more suspicious about the overall high accuracy of the classification: I do not think the validation scheme is reliable.

- We used TOA images since at the time of the analysis the northern S2 tiles were not yet processed to BOA images. While BOA is generally better, TOA has been proven to be of sufficient quality to derive high accuracy land cover products.

Figure 3: what are B and C inlets, they're not sand probabilities.

- We think you are referring to figure 4 (now figure 5). Indeed, we labelled inlet B and C incorrectly. These do not show the sand probability but regional examples of the Lena Delta habitat map. We corrected the figure caption. Thanks.

  Figure 5*: Lena Delta habitat classes (Dataset 5). The entire Lena Delta on the left with three regional examples: A showing the seasonal sand probability and B and C regional examples of the habitat classes.*