Bachelor Thesis
Universität Greifswald | Institute of Geography und Geology

# Mapping Species Composition in the Permafrost Environments of the Central Lena Delta - An Approach using Ordination Analysis and Field Spectroscopy

Submitted by

**Carl Christoph Stadie**

| | |
|---|---|
| Study program: | Geography B.Sc. |
| Registration number: | 160845 |
| Supervisors: | Prof. Dr. Sebastian van der Linden |
| | Dr. Alexandra Runge |

Greifswald, 1st September 2021

**Abstract**

Permafrost landscapes are one of the earth´s landscapes most affected by climate change. Thus, monitoring these landscapes is necessary. Vegetation cover and vegetation composition often serve as an indicator for the state of the permafrost and thus need to be monitored accordingly.

This thesis uses a non-classificatory ordination approach to map species composition in the central Lena delta in Russia. It further describes patterns to be found within the central Lena delta´s species composition as well as which kind of data serves best as a basis for vegetation composition mapping.

Partial Components Analysis (PCA) was used to extract species composition and floristic gradients from a dataset collected during the expedition LENA-2018. The extracted values were combined using different kinds of spectral input data, both hyper- and multispectral, as well as aggregated and non-aggregated, to train a Partial Least Square Regressor (PLSR). Species composition was mapped using a Sentinel 2 image from August 2018.

It was shown that species composition was dominated by Mosses, *Carex chordorrhiza*, *Salix glauca*, *Salix pulchra*, *Eriophorum vaginatum* and *Poaceae*. Each species was linked to a specific type of landform. Hyperspectral, aggregated data performed best as predictor variables, but due to the lack of hyperspectral imagery, the aggregated multispectral data can also be used as it leads to satisfying results.

These results indicated that species composition mapping in permafrost landscapes is possible although to a lesser degree of accuracy compared to lower latitudes. These observations could pave the way further studies resulting in a more detailed description of permafrost properties related to species composition like active layer thickness or greenhouse gas content.

**Contents**

**List of Figures**

**List of Tables**

**Abbreviations**

| | |
|---|---|
| AWI | Alfred-Wegener-Institute |
| Nm | Nanometres |
| PC | Partial Components |
| PCA | Partial Components Analysis |
| PLSR | Partial-Least-Square-Regression |
| $R^2$ | explained variation |
| RMSE | Root-Mean-Square-Error |

**1 Introduction**

Permafrost landscapes are one of the earth's regions most affected by climate change. With active layer deepening (Park et al., 2016), overall soil temperatures rising and permafrost thawing (Boike et al., 2019), permafrost landscapes have long been subject to climate change research. In this respect, low-arctic tundra vegetation has been gaining importance. Covering around six million square kilometres (Atkinson and Treitz, 2012) low arctic vegetation cover and species composition has shown to be strongly linked to permafrost properties like active layer thickness (Walker et al., 2003) or greenhouse gas content (Fuchs et al., 2018), thus serving as a good indicator for monitoring the effects of climate change (Atkinson and Treitz, 2012). Furthermore, changes in vegetation cover itself have had huge impacts on the low arctic environment (Kapfer and Popova, 2021). Considering these aspects, detailed mapping of vegetation cover and vegetation composition is needed (Macander et al., 2017).

Traditional approaches to vegetation cover mapping involved the use of spectrally distinct and ecologically meaningful classes representing different vegetation communities, which although they have proven to be highly useful  (Atkinson and Treitz, 2012), often lack necessary details regarding the depiction of individual species (Bartsch et al., 2016). Furthermore, the resulting maps show hard boundaries, which, although easy to understand, do often not reflect the gradual transitions found in vegetation cover accordingly (Feilhauer et al., 2021). They are also of limited use for long-time monitoring because changes between classes are unlikely, for change tend to occur within the given class boundaries instead of the total transition from one class to the other (Macander et al., 2017). Thus, classificatory approaches may miss subtle changes and new approaches considering mapping of low-arctic tundra vegetation are needed. These should consider vegetation as a continuous field of species composition, rather than a set of sharp vegetation cover classes.

One of these approaches is the use of ordination methods to better reflect the variability of vegetation  (Beamish, 2019). Ordination methods use dimensionality reduction to display multivariate data in a two or three-dimensional space while conserving the similarity or dissimilarity of observations (Syms, 2008). In terms of species composition, sites with a similar species composition will be grouped. Given that species composition influences a surface's reflectance (Buchhorn et al., 2013), spectral measurements of the vegetation

sample sites can be used to train a regressor to map the species composition of the given area (Feilhauer et al., 2010).

Using this non-classificatory approach, gradual changes in species composition are represented as continuous fields rather than sharp boundaries (Schmidtlein et al., 2007). Although ordination analysis has been successfully performed in low arctic environments (Forbes and Sumina, 1999), examples of mapping species composition are hard to find.

Thus, in this thesis projective cover data from 13 sites in the central Lena delta will be subject to a Principal Component Analysis (PCA) and combined with hyperspectral field spectroscopy data using a Partial-Least-Square-Regression (PLSR) to map species composition over Sentinel 2 imagery to answer the following questions:

- Can methods of ordination be used to successfully map the species composition in the central Lena delta?

- Which kind of data is best suited for mapping using ordination techniques?

- How is the central Lena delta´s vegetation composed and what patterns do emerge?

**2 Theoretical Background**

**2.1 Low-Arctic Tundra Vegetation and its Spectral Properties**

Low-arctic vegetation composition on a large scale is determined by temperature and thus divided into five distinct bioclimatic zones A to E (figure 1). The most northern zone A can be characterized by low mean July temperatures of approximately 0-3 °C and mostly barren surfaces, with moss and lichen cover occurring only under favourable conditions. With higher temperatures to the south, subzones B (3-5 °C) and C (5-7 °C) show continuous layers of moss, lichens and herbs, as well as a second layer of prostrate shrubs respectively. The two most southern subzones D (5-9 °C) and E (9-12°C) can be distinguished through the emergence of dwarf shrubs with a maximum height of 40 cm in subzone D and up to 80 cm in subzone E (Walker et al., 2018).

Within these subzones, the spatial distribution of vegetation is controlled by different environmental factors resulting in quite diverse and very heterogeneous vegetation cover (Liu et al., 2017). Soil moisture has been identified as the primary driving factor, forcing species to arrange in distinct, but spectrally quite similar communities along moisture gradients (Walker et al., 2005), thus roughly dividing tundra vegetation into wet and dry tundra (Muster et al., 2012). Furthermore, factors like soil chemistry (Buchhorn et al., 2013), the availability of nutrients (van Wijk et al., 2005), topography (Evans et al., 1989) and the presence of permafrost disturbances (Rudy et al., 2013) have shown to contribute to the intrazonal differentiation of low-arctic vegetation.

Spectral properties of low-arctic tundra vegetation follow these environmental gradients (Buchhorn et al., 2013). The spectral signal of low-artic vegetation is a mixture of signals of vascular and non-vascular plants as well as the soil background (Ulrich et al., 2009). The reflectance within the visible electromagnetic spectrum (400-700 nm) is determined by the three major photosynthetic pigment groups, namely chlorophyll, carotenoid, and anthocyanin (Coops et al., 2003), while cellular and canopy structure control the reflectance values of the near-infrared and red edge regions (Gausman, 1974).
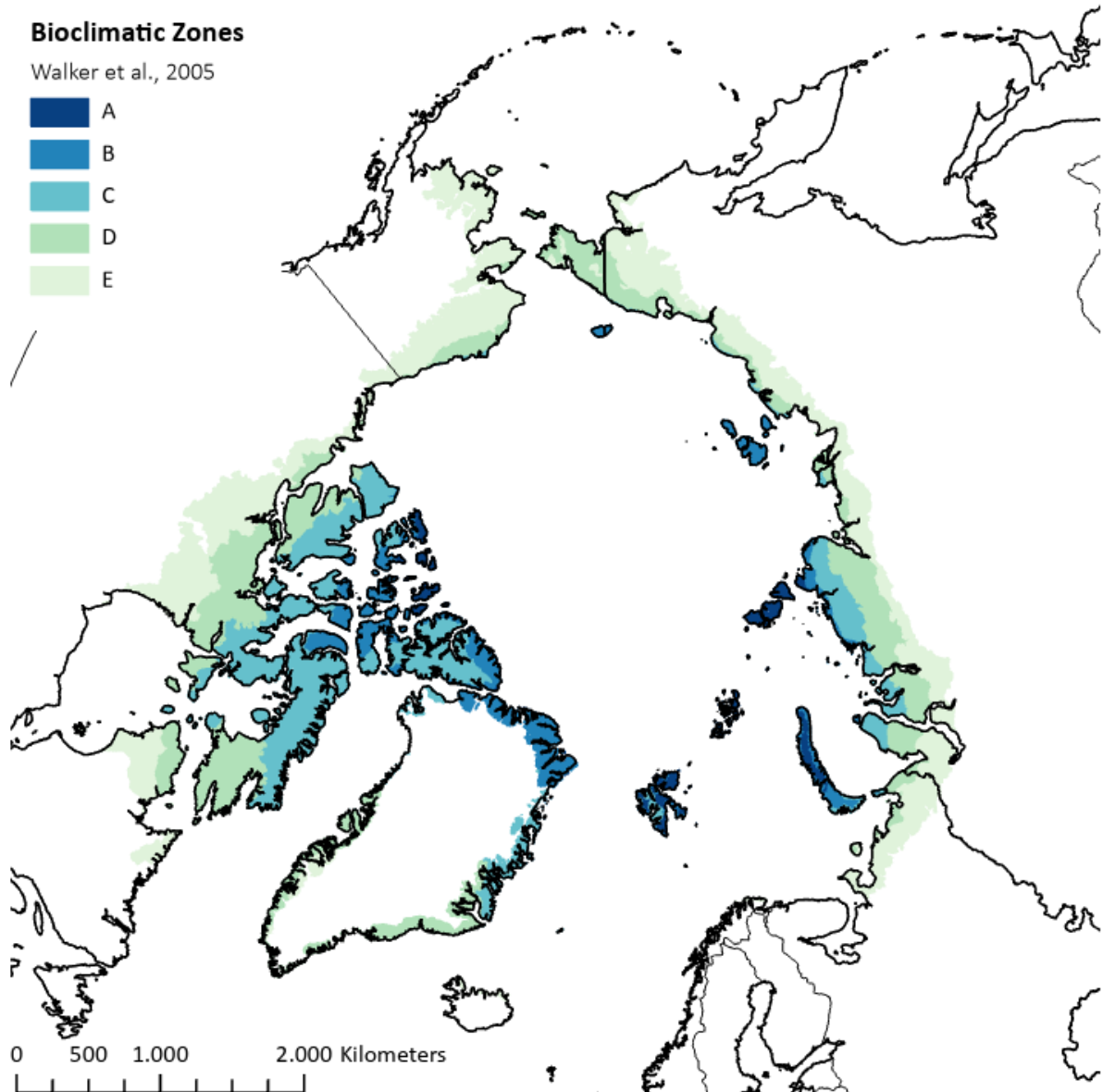
**Figure 1:** Bioclimatic zones of low-arctic vegetation based on Walker et al., 2005

For low-artic vegetation in general this means that the observed reflectance is on average lower than the reflectance observed in denser growing biomes because there are mostly few vascular plants, high soil moisture and a high amount of dead material on the ground as well as intense shadowing due to high sun zenith angles. Thus, when compared to reflectance profiles of lower latitudes, low-arctic vegetation spectral features are more similar to those of soil or vegetation in senescence (Liu et al., 2017). Moreover, there is only a weak green reflectance peak to be observed, as well as a narrow red edge. Additionally, the red edge does not end in a near-infrared-reflectance plateau, instead, the red edge starts to bend and shows a steady increase until it has reached its maximum reflectance at around 1020 nm

(Buchhorn et al., 2013). Reflectance profiles of different species tend to be most discriminative within this range of the near-infrared-region (Bratsch et al., 2016).

However, these general properties can be modified by environmental factors and surface properties (Ulrich et al., 2009). Sites situated further south but being similar to northern sites in terms of species composition, tend to show higher reflectance values, especially in the NIR regions as they possess more above-ground-biomass (Buchhorn et al., 2013). Low-lying wet sites might show higher reflectance in the near-infrared region (Ulrich et al., 2009), but although inhabiting large amounts of biomass does not necessarily show this increase in reflectance as vegetation height and soil moisture tend to balance each other. In terms of soil acidity, moist acidic sites are generally influenced by chlorophyll-rich shrubs and mosses leading to a slight increase in reflectance in the visible green, while moist non-acidic sites tend to show high reflectance in the visible red and the visible blue. The reason for that can be found in the occurrence of soils covered by large amounts of biological soil crust and shrubs with a large amount of non-green dead components (Buchhorn et al., 2013). Yet, these differences in reflectance are very weak and often get superimposed by the effects of changing species community or plant phenology (Beamish et al., 2017).

## 2.2 Using Ordination Analysis for Vegetation Mapping

Methods of ordination have long been an established tool in ecology for pattern identification and visualisation of multivariate data as they order samples according to their similarity (Anderson, 1971). In remote sensing, however, these methods have been used for a long time only for the detection of spectrally distinct units to serve as a basis for later classifications (Thomas et al., 2003). Ordination methods can transform large species-by-plot-matrixes into fewer dimensions (Schmidtlein et al., 2007) by locating each site in a multidimensional ordination space where each axis resembles one kind of species. The sites get located on these axes depending on the fraction of species found at that particular site. These multidimensional coordinates get transformed into ordination scores which represent the position of the site in the multidimensional ordination space and therefore species composition, but in lower usually two to three dimensions all while preserving the distances between the sites in the multidimensional space.

For example, given is an area with only three species x, y and z being present and four sample sites being located, where at sites 1 to 3 one kind of species is dominant and at site 4

all three species are equally distributed: If these sites are plotted in a multidimensional space according to their species composition (here three dimensional as only three species are present), sites 1 to 3 will be located on the axis of their representative species and site 4 will be found between the three extremes (figure 2a). When being transformed to a lower, two-dimensional space the distances between these sites will be preserved with sites 1 to 3 located at opposite corners of the plot and site 4 centred between them (figure 2b). Although the number of dimensions has been lowered, the sites still keep their relative distance from each other (Leyer and Wesche, 2008).
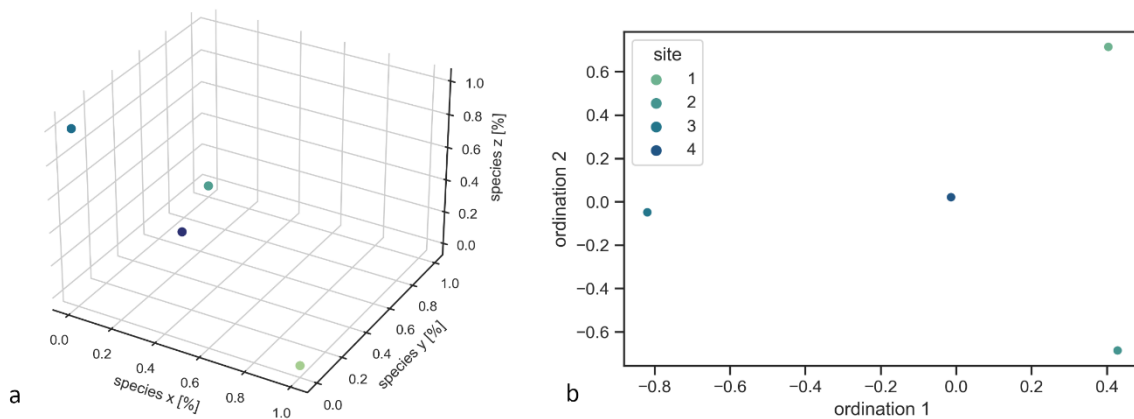


**Figure 2:** Exemplary transformation of multivariate data from a multidimensional into a two-dimensional space. 2a: The tree extremes (sites 1 to 3) were located at their respective place in the multidimensional space, with site 4 being located in between. 2b: When transformed into a lower dimension the sites keep their relative distances to each other with sites 1 to 3 at the extremes and site 4 in the middle.

For achieving dimensionality reduction, different algorithms like Nonmetric Multidimensional Scaling (Kruskal, 1964), Detrended Correspondence Analysis (Hill and Gauch, 1980) or Principal Component Analysis (Hotelling, 1933) can be used. Which algorithm is best suited depends on the nature of the input dataset. NMDS and DCA are useful if the dataset covers a wide range of environmental conditions with species being present at a site on one end of an expected gradient but completely absent on the other end, whereas PCA for example works best with datasets with low variability where sites on opposing ends of the gradient still are largely similar in terms of species composition (Feilhauer et al., 2011). In the resulting graph sites which are similar to each other are grouped as they show a similar species composition and therefore share the same

environmental properties, which can then be extracted (Virtanen et al., 2006). Thus, ordination analysis can extract patterns in species occurrence while still preserving the continuous nature of vegetation (Gleason, 1926).

As reflectance is confirmed to be empirically related to species composition (Feilhauer et al., 2010), spectral measurements (either hyper- or multispectral) in combination with ordination scores can be used to map species composition (Harris et al., 2015) usually utilising a regressor. Because reflectance and gradients both being continuous variables, using a standard regression would be possible, but is highly prone to error especially when using hyperspectral data as an input variable (Kumar et al., 2002), thus other methods like Partial-Least-Square-Regression (PLSR) is used (Wold et al., 2001).

Studies utilising ordination techniques for vegetation mapping have been growing in importance. The approach was first used by (Schmidtlein and Sassin, 2004) to map species composition along environmental gradients in Bavarian grasslands. The approach was further tested in different biotopes such as moors (Schmidtlein et al., 2007), heaths (Feilhauer et al., 2011) and peatlands (Harris et al., 2015). All these studies have used hyperspectral imagery in combination with varying reduction methods. In arctic environments, however, only a few studies exist, namely Döpper et al. (2021), who used ordination techniques to model species composition and active layer depth using Sentinel 2 imagery in central Alaska. However, no examples of studies further north than 65° N exist.

## 3 Material and Methods

### 3.1 Data

During the expedition LENA-2018 extensive field data were collected on the islands of Samoylov and Kurungnakh. Fieldwork included the collection of reflectance spectra of various homogeneous vegetation areas on different permafrost landforms deemed representative for the area (Runge et al., 2022), as well as a sampling of estimated foliage projected cover, originally with the modelling of above-ground biomass in mind (Shevtsova et al., 2021b). Sampling took place during solar noon between August, 7$^{th}$ and August, 18$^{th}$. For each site, a square of 30x30 m was laid out, in which the spectral measurements were randomly taken using a Spectral Revolution 2500 field spectrometer (table 1). On average 100 measurements with a spectral range of 350-2500 nm were acquired. A scanning time of 5 seconds was used to acquire one spectrum, each acquired spectrum represents an average of 10 individual measurements. Applying this averaging procedure increases the signal-to-noise ratio of the resulting reflectance measurement (Ulrich et al., 2009).

**Table 1:** Instrument details for the Spectral Revolutions SR-2500 field spectrometer

| | |
|---|---|
| **Spectral range** | 300-2500 nm |
| **Spectral resolution** | 3.5 nm (350-1000 nm) |
| | 22 nm at 1500 nm |
| | 22 nm at 2100 nm |
| **Sampling bandwidth** | 1 nm increments, 2151 channels |
| **Detectors** | 512 element UV enhanced Si array |
| | 256 element TE cooled extended InGaAs array |

At the start and at the end of each survey the system was referenced by measuring the back reflected radiance from a Zenith Lite $^{TM}$ Diffuse Reflectance Target (Runge et al., 2022). All data were corrected for erroneous measurements and spectral bands containing water absorption features were masked. Furthermore, all measurements were resampled to the spectral resolution of Sentinel 2 MSI using the Q-GIS Plugin EnMap-Box (van der Linden et al., 2015).

11

Projective cover of tall shrubs was estimated on a circular sample plot with a radius of 15 m around the centre of the 30x30m square mentioned above. Within this radius, three individual subplots, 2x2 m in dimension, were placed for projective cover estimation of other species, based on the representativity of the subplot for the whole site (figure 3). The samples of the individual subplots were averaged to be representative of the whole plot (Shevtsova et al., 2021a). In total, the dataset includes 15 sites of various homogeneous vegetation areas on different permafrost landforms, of which 13 were included in this analysis. Two sites were excluded due to unsuitable illumination conditions or the presence of clouds during the sampling, resulting in inconsistent reflectance measurements. An overview of all 13 sampling sites can be found in appendix 1.
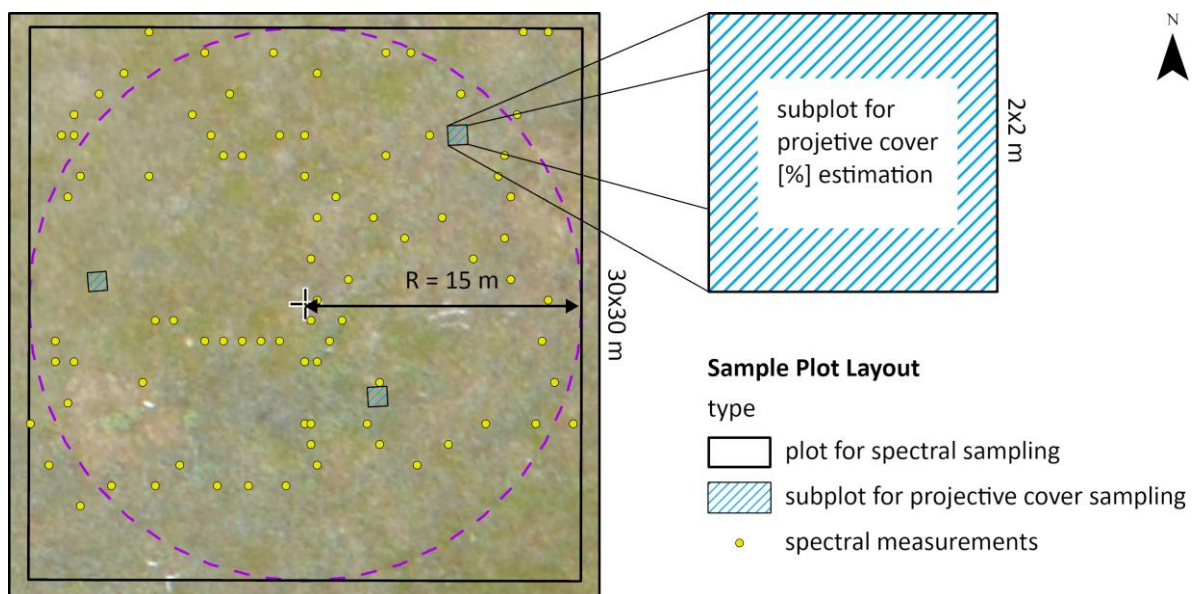


**Figure 3:** Sampling scheme for the LENA 2018 expedition: Spectral sampling took place randomly within a 30x30m square, projective cover for all shrubs were estimated within a 15 m radius of the square´s centre. Projective cover estimation of other species was based on three 2x2 m subplots (figure based on an orthomosaic by Boike et al., 2015).

For later extrapolation of ordination values, a Sentinel-2-1C scene acquired on August, 6th 2018 was corrected for atmospheric influences using the atmospheric correction processor Sen2Cor. Atmospheric correction processing was performed with the default configuration which uses a rural aerosol model with a start visibility parameter of 40 km corresponding to the aerosol optical thickness of 0.20 at 550 nm (Pignatale, 2020). Furthermore, non-vegetated areas like waterbodies and sandbanks were masked.

**3.2 Study Area**

The Lena delta is the largest arctic delta in the world. Located at the Laptev Sea in the zone of continuous permafrost, it is dominated by permafrost landscapes with permafrost depths reaching up to 600 m (Grosse et al., 2007). The delta is geomorphologically active, most of it consists of wetlands with a heterogenous microrelief with thaw depressions with lakes, ice-wedge-polygons as well as pingos (Ulrich et al., 2009). The delta can be divided into three main terraces, with the first being the still active Holocene floodplain dominated by ice-wedge-polygons (Boike et al., 2008). The second terrace is characterized by sandy late-Pleistocene deposits where lakes and thaw depressions are common. Lastly, the third terrace resembles the remnants of the Pleistocene accumulation plane with thick, ice-rich, and highly organic yedoma sediments (Are and Reimnitz, 2000). The study area (figure 4) is located in the central Lena delta and consists mostly of islands belonging to the first terrace, except for Kurungnakh, which yedoma uplands are part of the third terrace.

Climate-wise, the study area is truly arctic, with an annual mean temperature of -12.5 °C. January and February are the coldest months, where temperatures fall on average to -30 °C and -33°C respectively. The warmest months are July and August, with an average air temperature of 7 °C. During summer rainfalls an average of 125 mm can be measured, summing up to an annual mean precipitation of  250 mm with winter snowfall, which is highly variable due to high wind speeds (Raschke and Savelieva, 2017).
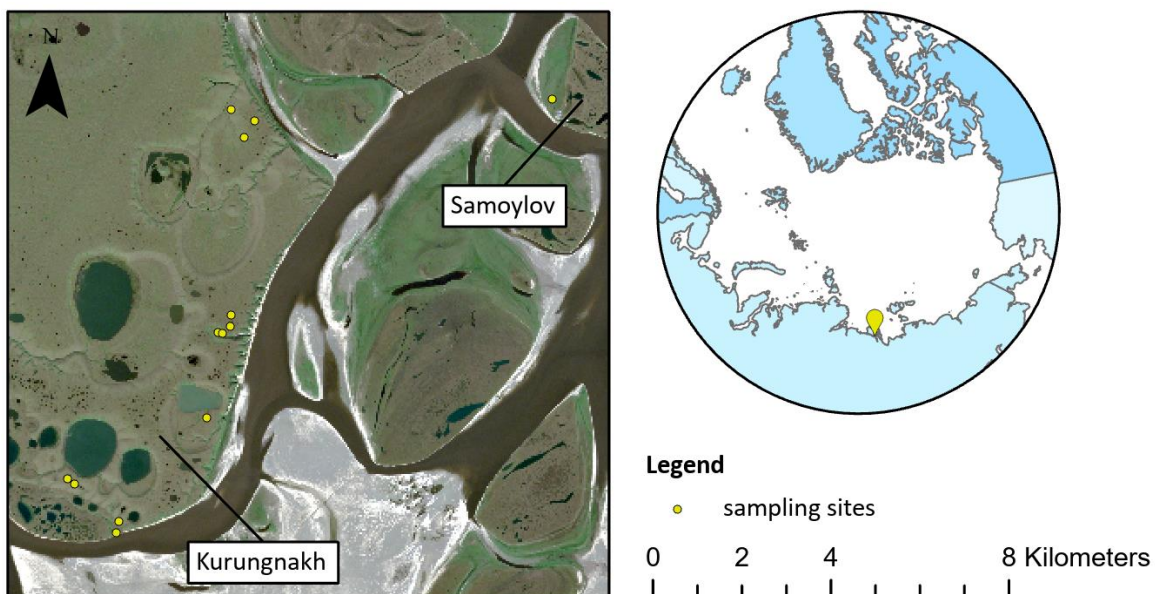


**Figure 4:** Central Lena Delta on a Sentinel 2 image. Kurungnakh as part of the third terrace is dominated through Pleistocene yedoma deposits. Samoylov is part of the active floodplain belonging to the first terrace.

The central Lena Delta can in terms of vegetation be characterised as partly tussock, sedge, dwarf-shrub moss tundra and partly as sedge, moss, and low-shrub wetlands according to the Circumarctic Vegetation Map (Walker et al., 2005). Dominant species are *Salix glauca*, *Carex chordorrhiza* and *Salix pulchra* along with mosses and lichens.  Mosses and Liches reach heights of up to 5cm while vascular plants, although sparsely represented in the area, can reach heights of up to 30 cm.  Vegetation is arranged in distinct species communities, which can roughly be separated in wet and dry polygonal tundra. These communities are although quite homogenous in their composition of dominating species, highly heterogeneous when it comes to their proportions of said species (Raschke and Savelieva, 2017).

**3.3 Methods**

Principal Component Analysis (Hotelling, 1933) has been chosen for the extraction of floristic gradients, as the data shows only a low grade of variation between the sampling sites in terms of dominating species (appendix 2). Principal Component Analysis has been shown to perform best under such circumstances (Feilhauer et al., 2011). The Principal Component Algorithm was applied to a matrix of species projected cover in per cent (columns) and sites (rows).  For the description of the floristic gradients, expert knowledge contributed by Dr Birgit Heim (AWI) was used.

To map the ordination space, the PCA-axis scores of the field plots were regressed against the corresponding collected field spectra as predictor variables (figure 5). To test the performance of different types of input data the collected field spectra were processed to different degrees. The original hyperspectral dataset was resampled to the spectral resolution of the Sentinel 2 MSI sensor. Furthermore, for each dataset, the hyperspectral, as well as the resampled one, all individual measurements of each plot, were aggregated using the median. Aggregation through median was chosen to limit the effects of frequent outliers in the field data. Thus, in the end, four individual datasets, two with individual measurements and two plot-wise median reflections, in both hyper- and multispectral resolution were produced to create the corresponding regression models, which were then to be evaluated.
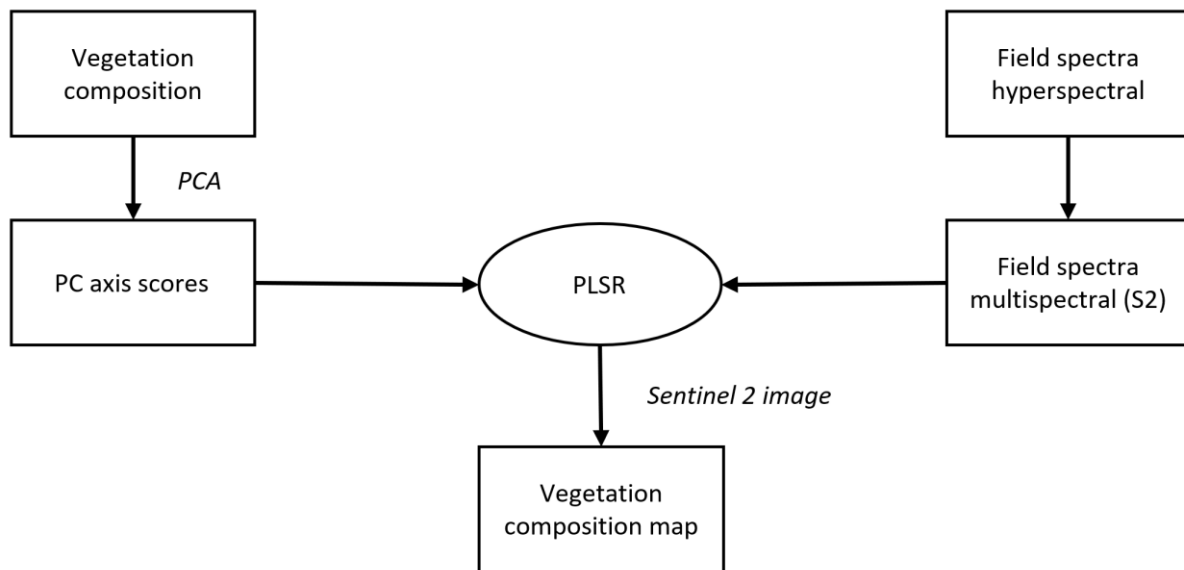
**Figure 5:** Workflow for species composition modelling.

All models were evaluated using the Leave-One-Out cross-validation results of $R^2$ and RMSE. As Regressor a Partial-Least-Square-Regressor (PLSR) was chosen for it has shown to be highly resistant against possible intercorrelations. PLSR finds a linear regression model by projecting the predicted variables and the observable variables to a new space (Wold et al., 2001).

For the final model multispectral aggregated median spectra were used, with reflectance being the predictor variable and the ordination score being the response variable. Both ordination axes were modelled separately. The trained models were applied pixelwise to a Sentinel 2 image from early August corrected for atmospheric influences. Waterbodies, as well as sandbanks, were masked.

All analyses were conducted in the R statistical environment (R Core Team, 2022) using the packages vegan (Oksanen et al.), pls (Liland et al., 2021) and raster (Hijmans, 2022).

## 4 Results

### 4.1 Partial Component Analysis

For the PCA-ordination, the first ordination axis explained 52.4 % of the variation inherent in the original vegetation dataset while the second axis explained 22.5 % of the variation. Any higher axes were only of minor explanatory value. Thus, the original dataset was reduced to a two-dimensional ordination space, which axes explained a variation of 74.9 % in total.

Figure 6 shows the sites´ positions in the two-dimensional ordination space. Each blue point stands for one site, while the arrows indicate the direction of the floristic gradients displayed by each axis. The first axis displays a moisture gradient. It indicates a transition from low-lying wet alases dominated by mosses over moist and dry polygonal tundra found on the yedoma upland and dominated through *Carex chordorrhiza* and Tussocks respectively, to dryer parts like vegetated flood plains and yedoma slopes where shrubs like *Salix glauca* are common. Moreover, at high PC1 and low PC2 scores, shrub fields dominated by *Salix pulchra* can be found.
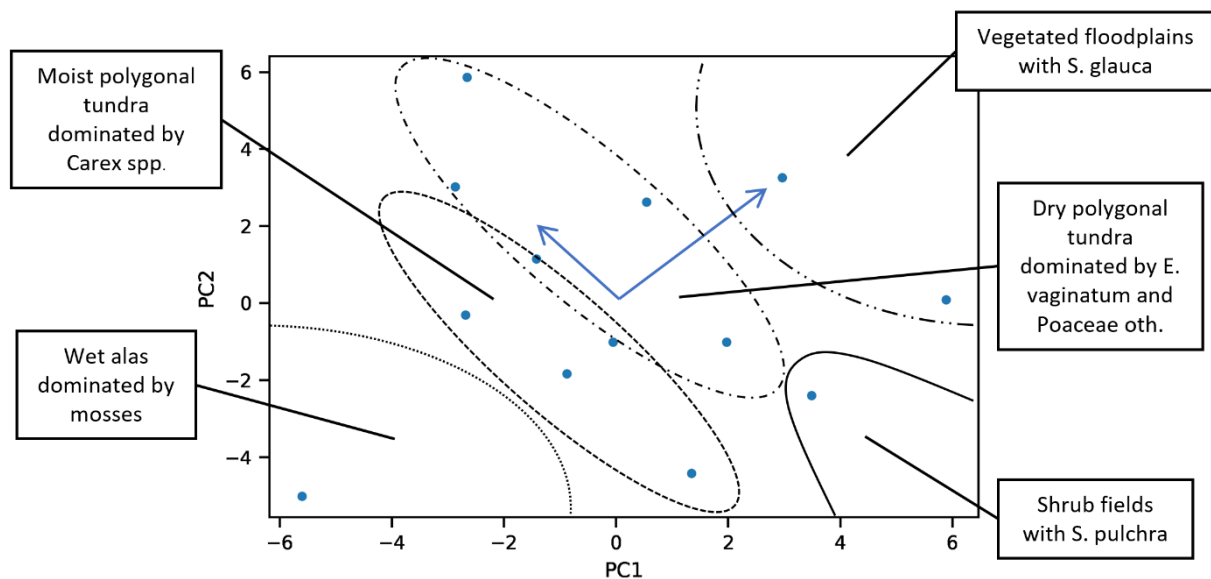


**Figure 6:** First two axes of the ordination space. Delineated areas mark the distribution of major landforms with corresponding vegetation. Arrows show the direction of floristic gradients.

The distribution of sites along the second axis however cannot be attributed to a single environmental factor. A mixture of soil moisture as a remanent of the first axis and availability of nutrients linked to the occurrence of permafrost disturbances is most likely to cause the displayed distribution. Hence, the gradients do not follow their corresponding axes in parallel, but appear to be slightly tilted. Mosses, *Carex chordorrhiza*, *Salix glauca*, *Salix*

*pulchra*, *Eriophorum vaginatum* and *Poaceae* were found to have the most influence on the distribution of sites within the ordination space. Other, less frequent occurring species had an only minor influence on the site´s distribution.

## 4.2 Field Spectroscopy

The results of the processed field spectroscopy can be found in figure 7. On the left (7a) hyperspectral profiles are displayed, while on the right (7b) the corresponding resampled profiles are to be seen. The dark blue line resembles the median reflectance per plot. The light blue areas display the standard deviation as a measure of scattering within the individual measurements per plot. The numbers above the subplots correspond to the assigned IDs of the individual sites and serve the purpose of better identification.

All hyperspectral profiles exhibit comparably low reflectance in the visible green as well as a narrow red edge. Most profiles experience their maximum reflectance of 75 % on average at around 1200 nm. Furthermore, they show the most variability between the individual sites within the near-infrared spectrum. Within other wavelength regions only minor differences between sites are recognizable.

Within the spectral profiles of sites 3, 8, 13 and 26 the red edge starts bending at 700 nm and steadily increases towards maximum reflectance, without forming a near-infrared plateau. The profiles of sites 9, 10, 15 and 14 show a steeper red edge compared to those described previously, however the bend at 700 nm is still present. No formation of a near-infrared plateau is visible, except for site 14.

Sites 5 and 6 are characterized by a steep red edge as well as high reflection throughout the near-infrared, resulting in a broad near-infrared plateau. The previously described bending at 700 nm is still present, but barely visible. Furthermore, these sites show the highest variation within the individual measurements. Lastly, sites 11 and 19 show a steep red edge and a broad near-infrared plateau without the characteristic bending at 700 nm. All four sites are characterised by thick vegetation cover resulting in the mentioned formation of a near-infrared plateau.

When comparing the resampled profiles with their respective counterparts, the loss of detail especially within the near-infrared is obvious. Most of the described features are barely visible. However, the most discriminative features like the steepness of the red edge or the formation of the near-infrared plateau and overall reflection can be recognized.
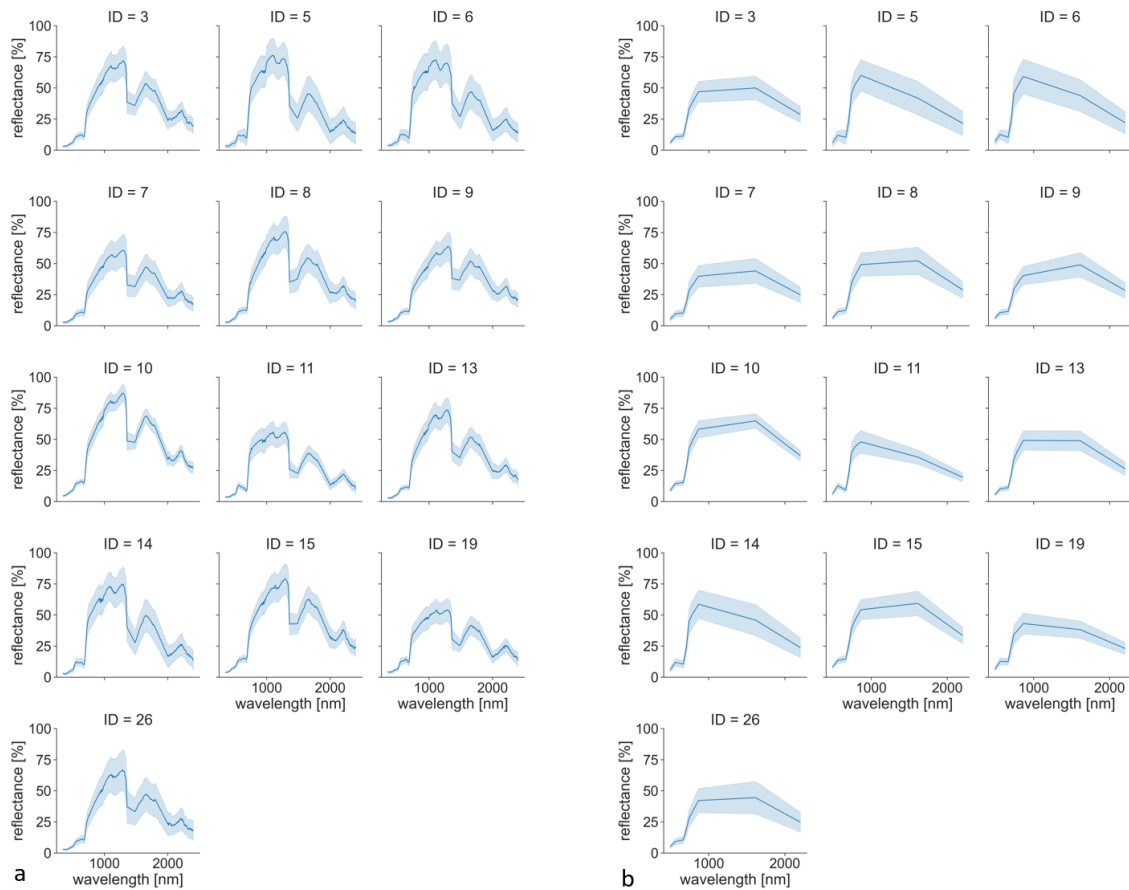
**Figure 7:** Plot-wise spectral profiles. 7a shows hyperspectral profiles, while 7b shows the corresponding profiles resampled to the spectral resolution of Sentinel 2 MSI. The blue line indicates the median profile, while the blue area indicates the standard deviation

## 4.3 Partial-Least-Square-Regression Model Fits

PLSR model fits show different results depending on the input data. The model using all non-aggregated hyperspectral measurements gained a fit of $R^2 = 0.67$ on the first axis (PC1) with an RMSE of 1.89. All available components were used as predictors, with those within the visible green, near-infrared and short-wave infrared being most influential. As seen in figure 8a, which shows the distribution of predicted PC scores as violines with the dashed line indicating a perfect model fit, most predicted axis scores are within range compared to the site´s position on the PC1 axis. However, some outliers are present, namely the extrema at -5.60983 and 5.88993, getting highly over- or, respectively, underestimated. In the case of 5.88993, only some outlier captures the measured PC-score, while at -5.60983 few of the predicted values matches the measured score. The outlying points belong to sampling sites 3 and 6, which were the wettest and driest sites. Furthermore, the effects of single outliers in

the spectral measurements with exceptionally high or low reflection become visible, leading to extreme outliers in the predicted axis scores, as to be seen at 3.48538 for example. These points can be attributed to sampling site 26.

On the second ordination axis (PC2), the model performed with an explained variation of $R^2 = 0.629$ and an RMSE of 2.01, with all components used as predictors, with those within the spectral range of near and short-wave infrared being most influential. PC2 performed slightly worse than the PC1 model fit. As with PC1, most measured values can be found within the range of the 25 % and 75 % quantile.  Also, extrema, to be found at negative axis scores (e.g., -5.01613/-4.42157) tend to be overestimated, while positive ones (e.g., 1.147468 upwards) are largely underestimated (figure 8b). In addition, outliers in the predictor variables again cause outliers in the predicted PC scores, although being less frequent, compared to the model of PC1. Still, outliers can be attributed to sites with exceptionally high scatter within the spectral measuremnets. It can also be seen, that the axis scores estimated by the PLSR are distributed unimodally, with most local maxima lying within the range of the corresponding measured values.
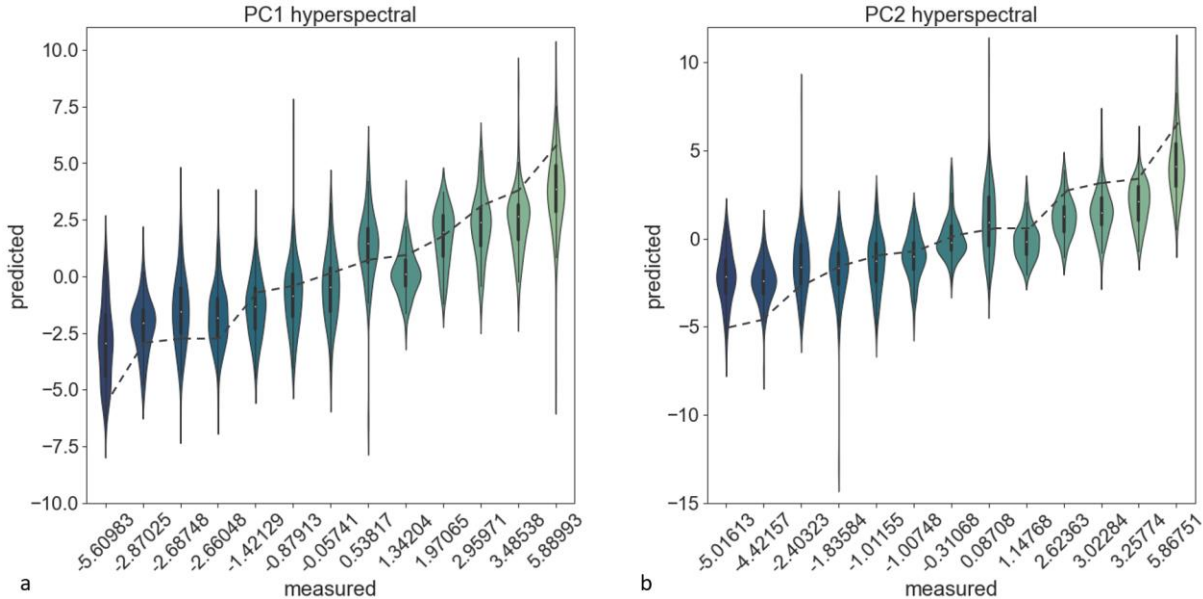


**Figure 8:** Predicted ordination scores for PC1 and PC2 models using single not aggregated hyperspectral input data against measured ordination scores. 7a shows predicted scores for PC1, while 7b shows predicted scores for PC2.

When using all available field measurements but resampled to the spectral resolution of Sentinel 2 MSI, the model gained a fit of $R^2 = 0.44$, while achieving an RMSE of 2.23. Again, all components were used as predictors. Variables within the near-infrared had the most influence on the model. Figure 9a shows that, like the two previous examples, the extrema

on both ends of the ordination axis suffer from over-/underestimation. When compared to the hyperspectral model of PC1, the Sentinel 2 model is less prone to outliers, especially in terms of wrongly underestimated values, like at 5.889933. PC1 hyperspectral clearly shows multiple outliers getting vastly over- or underestimated, which are absent in PC1 Sentinel 2. This resistance against outliers results in an overall trend to bulkier, but still mostly multimodal distributions.

The model of PC2 with all available components for Sentinel 2 used as predictors of which the near-infrared and short-wave infrared were most influential, reached an RMSE of 2.79 and an explained variation of $R^2 = 0.14$. The trend of slight over- and underestimation of extreme values observed in the previously mentioned models continues as to be seen in figure 9b. Furthermore, outliers like those present at -2.40323 or -1.83584 in the hyperspectral model of PC2 get eliminated.
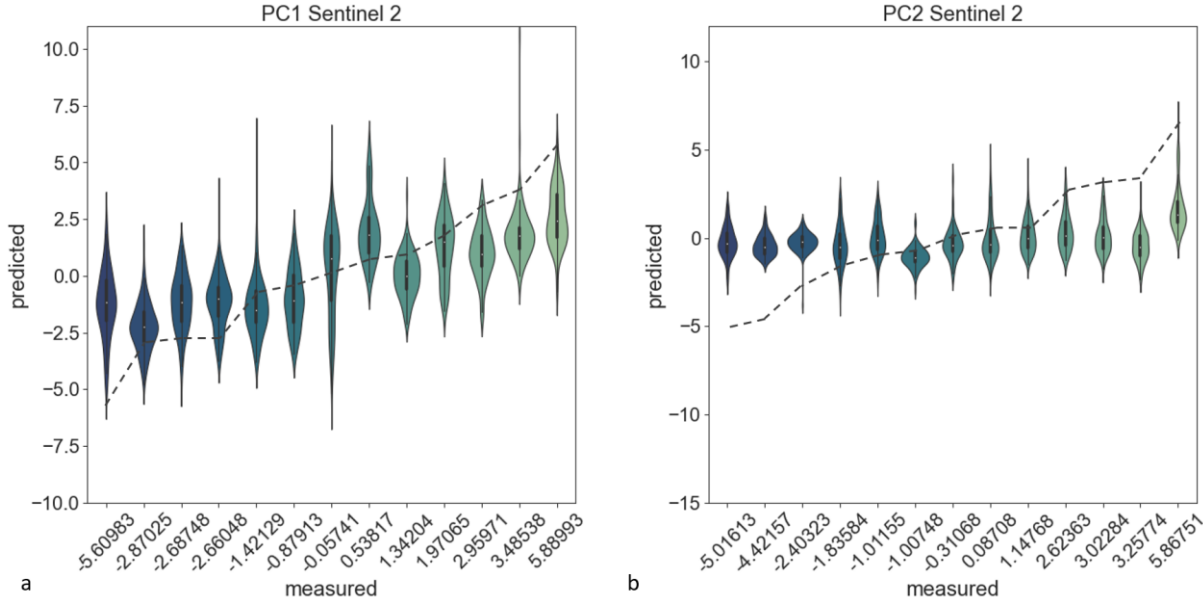


**Figure 9:** Predicted ordination scores for PC1 and PC2 models using single not aggregated multispectral input data against measured ordination scores. 8a shows predicted scores for PC1, while 8b shows predicted scores for PC2.

All models show a high scatter within the predicted variables and have difficulties to depict extreme axis scores correctly. The two hyperspectral models performed slightly better in terms of RMSE and showed a significantly better performance in terms of overall explained variation. The PC1 axis always performed better both in terms of RMSE and $R^2$ compared to their corresponding PC2 axis. This again shows that PC1 is the dominant ordination axis explaining most of the species distribution in the study area.

When using only spectral measurements plot-wise aggregated by median, PC1 scores an RMSE of 2.99 with hyperspectral data as predictor variables. The fitted model explains 94.11 % of the variation inherent to the first ordination axis. These results were achieved by using six components as predictor variables, of which those within the spectral range of the visible green, near-infrared and short-wave infrared had the most influence. Figure 10a shows the predicted PC1 scores plotted against the measured ordination scores. About half of the predicted scores get over estimated while the other half gets underestimated, with the extrema being badly represented. An outlier is present at -2.66048, which can be attributed to site 26 with high variability within the spectral measurements due to a late date of recording with starting senescence.

For PC2 the hyperspectral PLSR model fit reached an RMSE of 3.49 while explaining 78.77 % of the variation. Best results were reached when using five components. Predictors within the visible green and near-infrared had the largest influence on the regression outcome. As shown in figure 10b, ordination scores near the end of the PC2 axis are badly estimated by the regressor, continuing the trend already mentioned with previously described models.
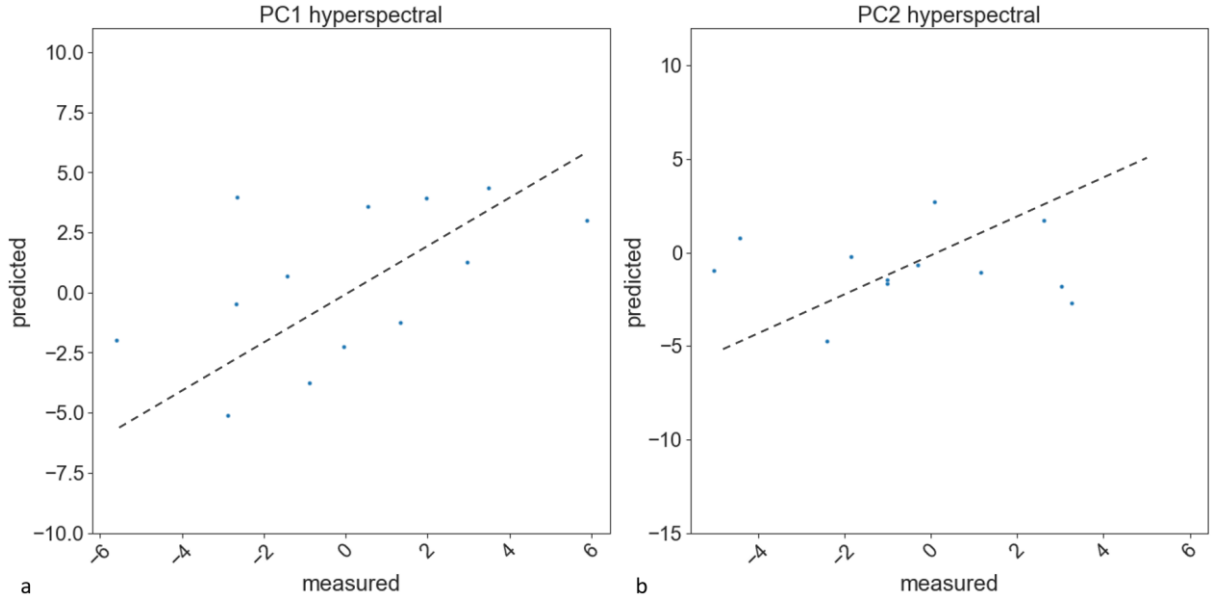


**Figure 10:** Predicted ordination scores for PC1 and PC2 models using aggregated hyperspectral input data against measured ordination scores. 8a shows predicted scores for PC1, while 8b shows predicted scores for PC2.

The PLSR model using plot-wise median spectra resampled to the spectral resolution of Sentinel 2 MSI explained 72.96 % of the variance for PC1, with an RMSE of 2.72. Four components were used, with the spectral ranges of visible green and near-infrared were of most importance for the regressor. Extrema are again badly depicted and an outlier can

again be found at -2.66048 (figure 11a). However, the model greatly improved the prediction of some PC1 scores, like -1.42129 and -0.053817, estimating them almost to the point.

For the second PC axis, the PLSR model reached an RMSE of 3.26 and explained 63.98 % of the variance inherent to the PC2 axis. Best results were achieved when using four components with variables within the red, near- and short-wave infrared being most influential. As figure 11b shows, the model captured the negative extreme better than the previously described models did. The positive extreme, however, follows the described trend and gets vastly overestimated.
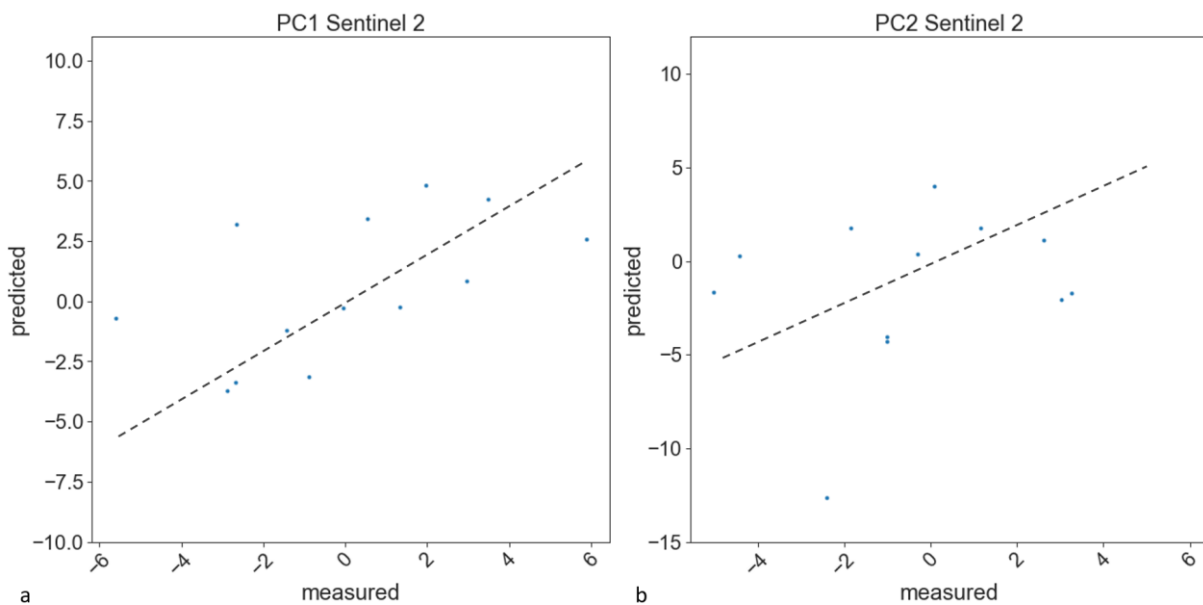


**Figure 11:** Predicted ordination scores for PC1 and PC2 models using aggregated multispectral input data against measured ordination scores. 8a shows predicted scores for PC1, while 8b shows predicted scores for PC2.

When comparing the described median-based models, it can be said that the models using data resampled to Sentinel 2 MSI perform better than those using hyperspectral data in terms of RMSE but show an overall worse performance considering the explained variation. Correct estimation of extreme axis scores remains an issue.

Taking all described models into account, those using aggregated spectra as predictor variables can explain more variation, but perform slightly worse when it comes to RMSE values. All models reach their limits when estimating extremely negative or positive PC scores. Moreover, outliers or high heterogeneity within the predictor variables have been shown to cause outliers in the prediction. For further analysis and mapping only models with predictor variables resampled to Sentinel 2 and plot-wise aggregation by median will be used as those offer the best trade-off between explained variation and RMSE (table 2).

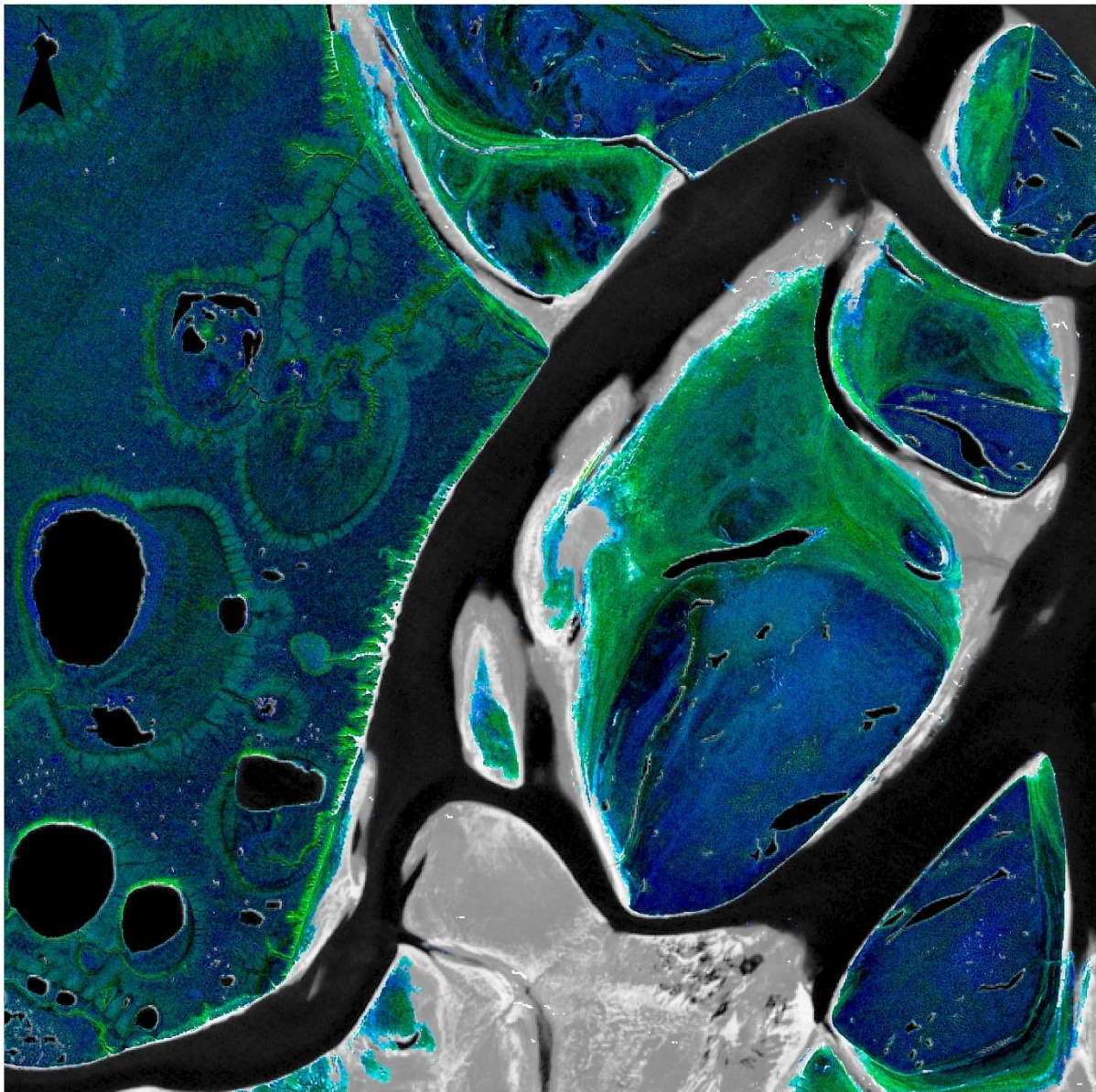**Table 2:** Comparison of PLSR model fits

| model | axis | RMSE | R² |
|---|---|---|---|
| Hyperspectral, non-aggregated | PC1 | 1.89 | 0.67 |
| Hyperspectral, non-aggregated | PC2 | 2.01 | 0.629 |
| Hyperspectral, aggregated | PC1 | 2.99 | 0.941 |
| Hyperspectral, aggregated | PC2 | 3.49 | 0.787 |
| Multispectral, non-aggregated | PC1 | 2.23 | 0.44 |
| Multispectral, non-aggregated | PC2 | 2.79 | 0.14 |
| Multispectral, aggregated | PC1 | 2.72 | 0.729 |
| Multispectral, aggregated | PC2 | 3.26 | 0.639 |

## 4.4 Mapping

Figure 12 depicts the false colour RGB-composite of the mapped PC-axes. The first PC axis has been assigned to the green channel, the second the blue, while the red channel has been left out, as only two ordination axes were used. Non-vegetated areas like waterbodies and sandbanks were masked and hence are displayed as a grey scale image of the Sentinel 2 scene used as a basis for the mapping of the axes scores. Different colours indicate a different species composition in line with PC scores.

The legend illustrates the colour ranges of different vegetation types and the distribution of characteristic species according to their scores determined by the Partial Components Analysis. The depiction of landforms and vegetation types is included to help with interpretation and shall serve as an indicator of general trends, as most vegetation types do not feature a consistent colour and thus are difficult to outline in the map.

Black or dark greenish/bluish indicates wet conditions mostly found in low-lying parts of alases, which are dominated by mosses. A rich bright blue indicates low scores on the PC1 axis but high ones on PC2, thus standing for moist areas, but low on nutrients like they can be found on Yedoma uplands. Different species of sedges are dominant. A bright green, on the other hand, indicates dryer, but nutrient rich conditions, which can be found on low lying shrub fields and are occasionally flooded by the Lena River. These are dominated by *Salix pulchra*. High scores in PC axes one and two are represented by turquoise colours, indicating old vegetated floodplains with *Salix glauca*.
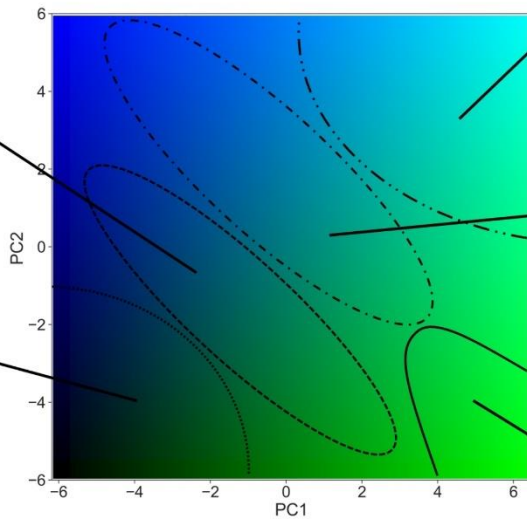
**Figure 12:** Mapped PC axes in the related colour space, defined by PC scores (red: none, green: PC1, blue: PC2)

The colour gradient from dark greens and blues towards turquoise describes the change from moist to dryer conditions and a change from mosses to shrubs in terms of vegetation. The gradient from green to blue describes changes expressed in PC2, like nutrient content of the soil or the presence of permafrost disturbances. In terms of vegetation composition, it indicates a change from shrubs to sedges. Hues not attributed to one of the described types as well as hues shared by multiple vegetation types show transitions between the neighbouring vegetation types.

The described transitions in colour reflect changes in the PC axes scores and thus changes in species composition. Figures 13 shows a detailed view of the mapped ordination axis near a gully in the south of Kurungnakh. The most prominent features are the shrub fields along the sites of the gully. According to the displayed colours, the area is dominated by *Salix pulchra*, which corresponds to the conditions found on the site. Furthermore, dark hues at the lowest parts near the centre and the end of the gully indicate wetter conditions and thus a change to mossier species composition.
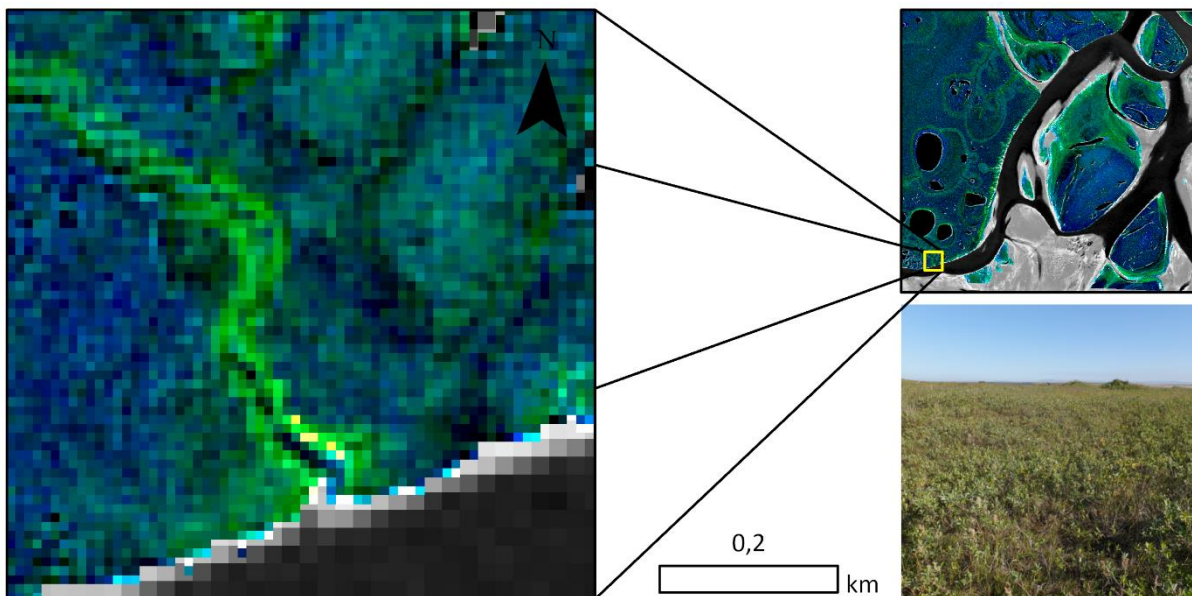


**Figure 13:** Detailed view of the mapped ordination axes in the south of Kurungnakh. Colours indicate shrub fields, which are to be found on site. Moreover, wetter conditions near the centre of the gully are appropriately depicted in darker tones.

The area to be seen in figure 14 is an old vegetated floodplain in the western part of the island of Samoylov. One can clearly see the difference between the moist *Carex chordorrhiza* dominated polygonal tundra in deep blue in the east and the floodplain in the west in green. This indicates that the western part is dominated by shrubs, however, when compared to

the floodplain described previously, instead of *Salix pulchra*, *Salix glauca* is the dominant species. This change in species composition is expressed in the primarily turquoise accent, which is especially prominent near the coastline and shows the dominance of dry and less nutritious conditions. Field observations show the domination of shrubs in the area, which are recognizable as *Salix glauca* by the cotton-like bushes.
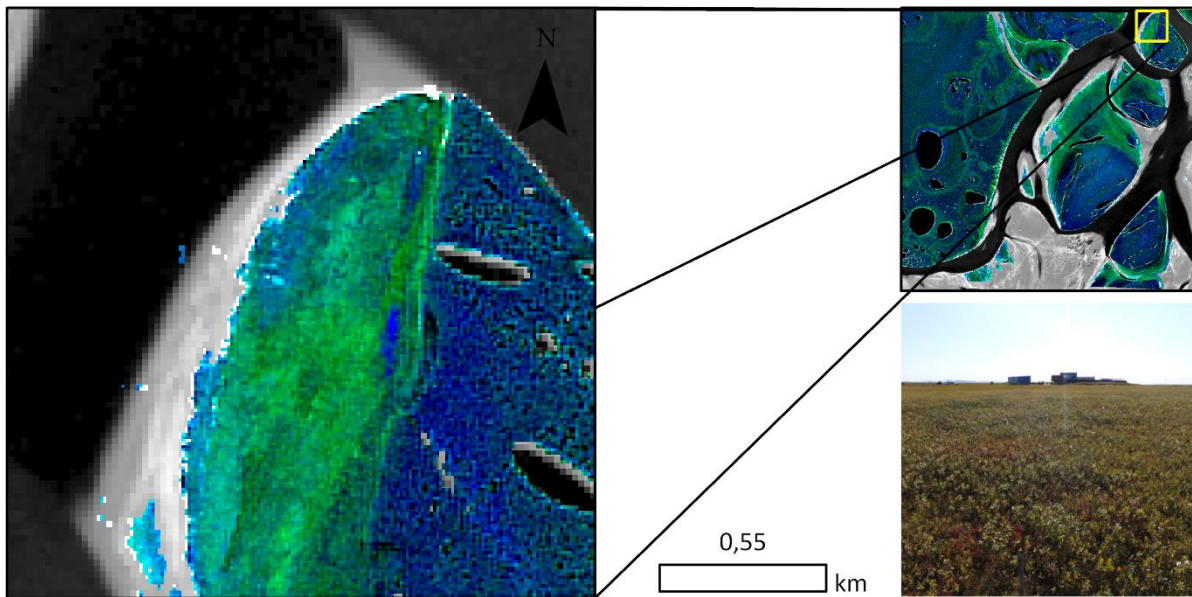


**Figure 14:** Detailed view of the mapped ordination axes in the northwest of Samoylov. The change in dominating species is expressed through the change in colour from green to turquoise especially visible near the coastline.

Figure 15 shows another detail to be found in an alas near the east coast of Kurungnakh. The darker areas at the bottom of the slopes surrounding the alas as well as within the deepest part of the alas southwest of the depicted lake indicate wet conditions, where mosses are most dominant. Field observations, as shown in the picture, confirm this impression. However, some portions of shrubs and sedges are still to be found. This observation explains the mixture of dark blues and greens found around the nearly black areas. It indicates the transition from wetter moss-dominated areas to areas where mosses still make up a considerable portion of the species composition, but other species, usually indicating dryer conditions are gaining importance. Figure 15 further shows an example of the representation of abrupt changes in species composition. The gully situated in the southeast of the map presents a hard change in environmental conditions, from the relatively moist and undisturbed yedoma uplands to the disturbed surfaces of the gully. This change corresponds with change in species composition from *Carex chordorrhiza* dominated moist

tundra to a field of *Salix pulchra* at the sides and at the bottom of the gully, visible by the abrupt change from blue to green.
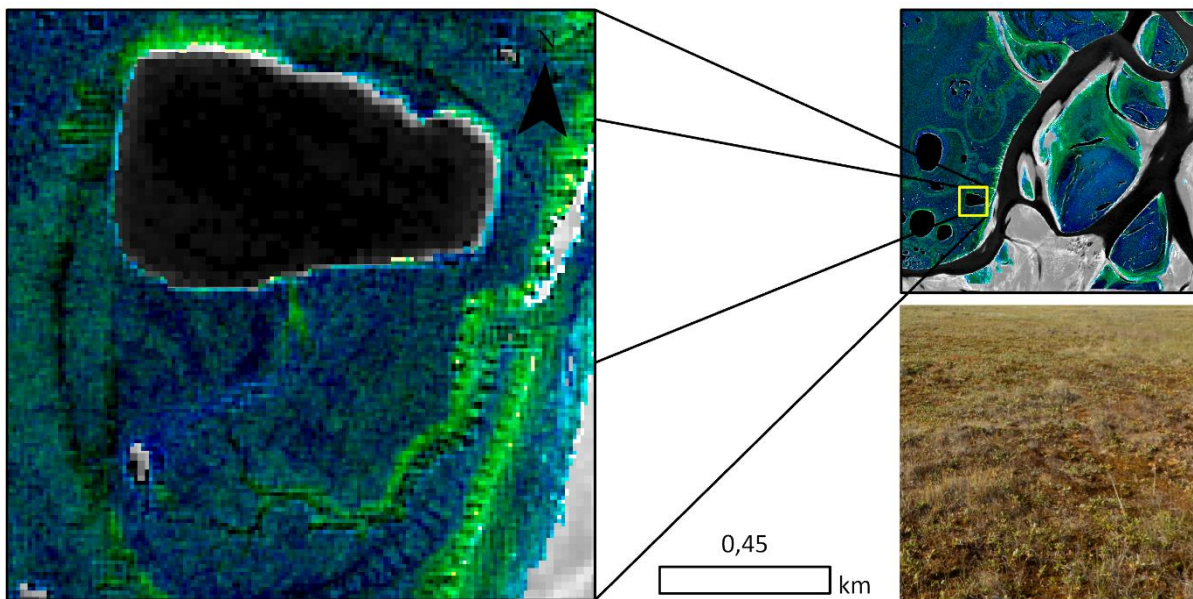


**Figure 15:** Detailed view of the mapped ordination axes in the southeast of Kurungnakh. Dark tones at the bottom of slopes indicate wet conditions, where mosses are dominant. Dark greens and blues show growing portions of shrubs and sedges and dryer conditions.

A good example of the continuous representation of changes in environmental conditions and species composition can be found in figure 16. It presents an area in the northeast of Kurungnakh where the yedoma upland slightly slopes towards a valley with a small stream. The top of the yedoma upland is characterized by a deep blue colour, which changes over a mixture of grey-blue-greenish towards dark blue and green at the bottom of the valley. Species wise this indicates a change from *Carex chordorrhiza* dominated moist tundra over dry tundra with *Eriophorum vaginatum* and *Poaceae* towards wet and mossy conditions near the stream. This gradient can be seen in the picture included in figure 14. The foreground shows tussocks of *Poaceae* and *Eriophorum vaginatum* with its cotton lie bushes, while the background shows the mossy areas around the creek.
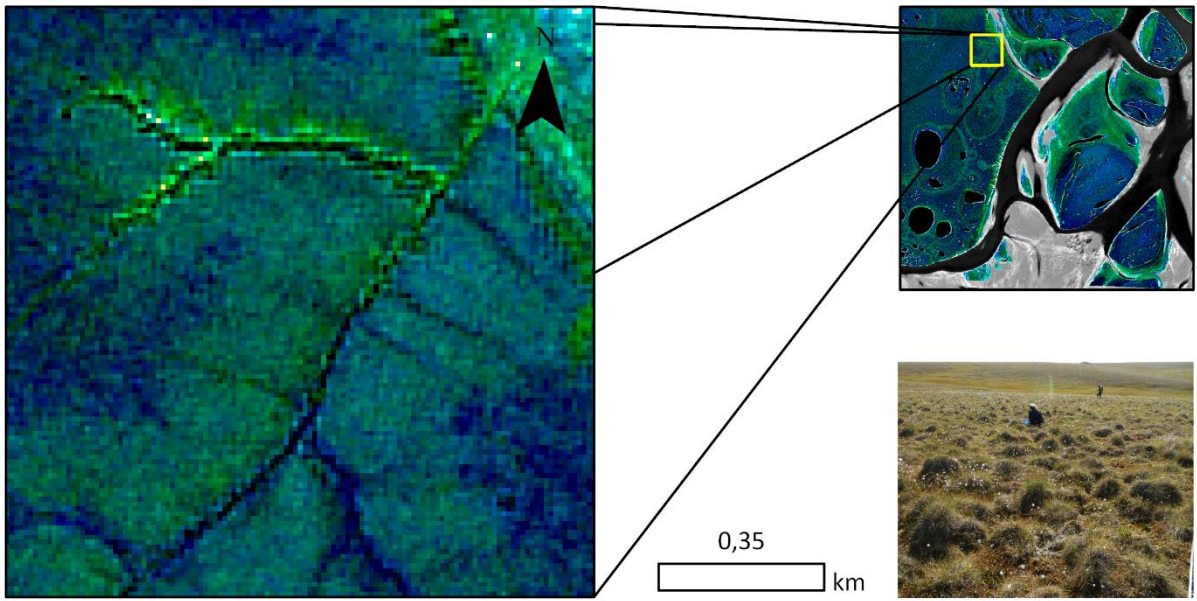
**Figure 16:** Detailed view of the mapped ordination axis in the northeast of Kurungnakh. A gradual change in environmental conditions and species composition is represented through a change in colour from dark blue via grey/green/blue to dark green/blue.

**5 Discussion**

The goal of this thesis was to answer the questions if the mapping of species composition of the central Lena delta using ordination techniques is possible, which data is best suited to do so, as well as how the central Lena delta´s vegetation is composed and which patterns do emerge.

As the Partial Components Analysis showed, species composition in the central Lena delta can be characterized as a mixture of moss, *Carex chordorrhiza*, *Eriophorum vaginatum*, *Poaceae*, *Salix pulchra* and *Salix glauca*. Other species are of lesser importance in terms of composition, as they are of minor occurrence. Both ordination axis combined explained 74.9 % of the variation inherent to the original dataset, which suggests a good performance when compared to other studies using ordination techniques (Forbes and Sumina, 1999; Döpper et al., 2021; Feilhauer et al., 2011). Moreover, it performed best compared to other methods of ordination like Detrended Correspondence Analysis, which were tested preliminary. It outperformed them not only in term of explained variance, but also depicted the distribution of species in the ordination space most unambiguously.

The most influential environmental factor can be identified as a moisture gradient. Furthermore, other environmental factors like the availability of nutrients coupled with the occurrence of permafrost disturbances are expressed in the second ordination axis, however, no clear pattern emerges. This could be caused by a mixture of environmental factors being present in the second axis as well as the soil moisture gradient expressed in the first axis being so pronounced that it is still present in the second axis.

These results reflect the current status in literature. Walker et al. (2005) described the central Lena delta´s vegetation as a mixture of mosses, sedges, tussocks and shrubs. The most common species are *Salix glauca*, *Carex chordorrhiza* and *Salix pulchra* along with mosses (Raschke and Savelieva, 2017). Furthermore, the distribution of vegetation is described as being determined by soil moisture (Schickhoff et al., 2002; Virtanen et al., 2006), explaining the strong moisture gradient observed in the ordination space. Muster et al. (2012) described the resulting division of tundra vegetation into moist and dry tundra, which is also emphasized by the presented results.

The spectral profiles serving as predictor variables show the abundance of the reflectance peak in the visible green, the bend in the red edge at 700 nm and the partial abundance of the near-infrared plateau. The spectral profiles are most discriminative within the near-

infrared which explains why the near-infrared region consistently was one of the most influential components used in the PLSR models. However, in other wavelength regions and especially in those profiles resampled to Sentinel 2, the profiles are quite alike, despite changes in species composition. In this regard, it can be concluded that due to the high discriminability in the near-infrared, which was previously described by Bartsch et al. (2016), and the high importance of said spectral region for the PLSR, a high spectral resolution within the near-infrared could be beneficial for further investigations.

In terms of their spectral characteristics the spectral profiles can be described as typical for low-arctic vegetation (Buchhorn et al., 2013) except for reflectance being higher than expected from the results of previous studies (Beamish, 2019; Buchhorn et al., 2013; Ulrich et al., 2009). This could be explained as a consequence of the spectral sampling taking place in densely and continuously vegetated areas, where the influence of soil and other factors was minimal. The low variability between the spectral profiles of different sites can be attributed to the abundance of vascular plants and the effect of high soil moisture in combination with dense vegetation, which balances the influences of these two surface properties (Ulrich et al., 2009).

When combining Partial Components Analysis with spectral information, the individual models show mixed results. The models using single instead of aggregated spectra as predictor variables perform better in terms of RMSE but worse in terms of explained variance. Furthermore, PC scores with high variation in their corresponding spectral measurements are more likely for higher scatter in the predicted PC scores. When comparing hyperspectral and multispectral models, hyperspectral models perform better both in terms of explained variation and RMSE.

Therefore, one can conclude that aggregated hyperspectral spectra are best suited as a basis for regression, as it explains most of the variance inherent to the original dataset and is more resistant to outliers. This conclusion is supported by the frequent use of hyperspectral imagery in similar studies (Schmidtlein and Sassin, 2004; Schmidtlein et al., 2007; Feilhauer et al., 2011). However, hyperspectral imagery is rarely available for arctic regions. Thus, as shown in this thesis, aggregated multispectral data can be used as an alternative. However, with hyperspectral data becoming widely available with the arrival of spaceborne hyperspectral systems like EnMAP, this situation might change.

When comparing the model fits presented in this thesis with those of similar studies of lower latitudes, it becomes obvious, that although hard to compare, while showing good results in terms of explained variance, the models used lack the needed accuracy in terms of RMSE. Not only extreme axis scores are nearly always over- or underestimated, but also are sites with a high scatter within the input data prone to the formation of outliers especially when using non-aggregated single measurements. The frequency and distance of outliers is reduced to some part by the use of multispectral data, but, although roughly following the trend of measured values, predicted and measured ones rarely match. Nevertheless, it has to be taken into account that the presented thesis represents the northernmost attempt to map species composition using ordination methods, which comes with extreme challenges like a high spectral similarity of vegetation surfaces. High sun angles and short vegetation periods resulting in quick both physical and spectral changes in vegetation further impair the quality of collected data. Döpper et al. (2021) experienced a drop in both RMSE and explained variance compared to studies conducted in more temperate climate zones (Harris et al., 2015), thus it is only logical to expect a further drop of these parameters 600 km further north.

Against these odds, the mapped PC scores were able to accurately show distinct patterns in the species composition of the central Lena delta. Low lying, wet areas such as alases or the bottom part of yedoma slopes are dominated mosses. The yedoma uplands themselves are dominated by *Carex chordorrhiza* in moist areas, while dryer parts and yedoma slopes are characterized through a combination of Poaceace and *Eriophorum vaginatum*. Furthermore, shrub fields dominated by *Salix pulchra* and vegetated floodplains populated by *Salix glauca* are recognizable. Within and between these patterns, transitions and changes in species composition can be seen through a gradual change in colour. However, despite the continuous nature of the mapped PC axes, abrupt changes in species composition are still represented as sharp boundaries.

With regard to the poor model fits, precaution is required concerning the mapped scores. Especially extremely high or low axis scores have a high probability of being vastly over- or underestimated. Hence, it can be said, that the PC scores presented in the map indicate the general trend of species composition in that particular area rather than an exact representation. Furthermore, the results suffer from the low amount of sampling points. While other studies outside the arctic used between 17 (Feilhauer et al., 2011) and 133

sampling points per km² (Schmidtlein et al., 2007) and Döpper et al. (2021) still using 0.3 points per km², here species composition was estimated using only 0.09 sampling points per km². Therefore, the mapped species composition can barely be deemed representative for the study area as a whole. This lack in representation is a direct result of the organising of sampling sites, which were not planned with the needs of the mapping of species composition with ordination analysis in mind. This is the major limitation of this work.

Consequently, further research is necessary to accurately map species composition by utilising ordination techniques. Those should consider a higher amount of sampling sites to achieve better spatial coverage and thus achieve results more representative for the area of interest. Moreover, possibilities to improve the fits of PLSR models should be explored to accomplish a more accurate representation of species composition. In further steps the feasibility of using species composition maps derived from ordination analysis as a base for upscaling of greenhouse gas emissions like it has been done on the basis of land cover classifications (Schneider et al., 2009) should be tested. Due to the continuous nature of ordination derived species composition maps, there is a high chance to enhance the accuracy of greenhouse gas emission models in comparison to the use of rough classifications. Similar studies executed in lower latitudes indicated that such an approach is feasible and promising (Lopatin et al., 2019).

**6 Conclusions**

This thesis´ objective was to explore if species composition mapping via ordination methods in the central Lena delta is possible, which data is best suited to do so and what patterns of species composition emerge. Therefore, different variations of field spectroscopy data in combination with ordination scores were fed into regression models, which were evaluated. The best performing model was used to map species composition over a Sentinel 2 image. Species composition of the central Lena delta can be described as a mixture of mosses, *Carex chordorrhiza*, *Poaceae*, *Eriophorum vaginatum*, as well as *Salix pulchra* and *Salix glauca*. The distribution of these species is determined by a soil moisture gradient leading to characteristic patterns. Mosses are dominant in alases and other low-lying areas, while *Carex chordorrhiza* dominates the wetter parts of yedoma uplands and moist polygonal tundra. *Poaceae* and *Eriophorum vaginatum* can be found on dryer parts of yedoma uplands as well as on yedoma slopes, while *Salix pulchra* and Salix glauca form shrub fields on old vegetated floodplains. Mixtures of and changes between these vegetation types were indicated by a change of colour. Regression models using aggregated hyperspectral data as predictor variables showed best performance. As hyperspectral imagery is rarely available in arctic regions, the use of data resampled to multispectral resolution is also suitable. A high spectral resolution of the near-infrared region is advisable as it showed to have great influence on the regressor´s performance.

This thesis demonstrated that ordination methods can be used as an alternative to land cover classifications for the mapping of species composition. However, it was also demonstrated that in order to achieve more accurate model fits further research is needed. Apart from this, the use of species composition data as a basis for greenhouse gas emission modelling needs to be explored. The accurate representation of species composition as a continuum instead of crisp classes provides the chance for a more accurate modelling of greenhouse gas emissions. Therefore, it could help to better describe the current and future impact of permafrost landscapes on our changing climate.

**Acknowledgements**

# References

Anderson, A.J.B. (1971). Ordination Methods in Ecology. *Journal of Ecology*, *59*, 713–726.

Are, F., & Reimnitz, R. (2000). An Overview of the Lena River Delta Setting: Geology, Tectonics, Geomorphology, and Hydrology. *Journal of Coastal research*, *16*. https://journals.flvc.org/jcr/article/download/80930/78072/92244.

Atkinson, D., & Treitz, P. (2012). Arctic Ecological Classifications Derived from Vegetation Community and Satellite Spectral Data. *Remote Sensing*, *4*, 3948–3971.

Bartsch, A., Höfler, A., Kroisleitner, C., & Trofaier, A. (2016). Land Cover Mapping in Northern High Latitude Permafrost Regions with Satellite Data: Achievements and Remaining Challenges. *Remote Sensing*, *8*, 979.

Beamish, A., Coops, N., Chabrillat, S., & Heim, B. (2017). A Phenological Approach to Spectral Differentiation of Low-Arctic Tundra Vegetation Communities, North Slope, Alaska. *Remote Sensing*, *9*, 1200.

Beamish, A.L. (2019). Hyperspectral remote sensing of the spatial and temporal heterogeneity of low Arctic vegetation.

Boike, J., Nitzbon, J., Anders, K., Grigoriev, M., Bolshiyanov, D., Langer, M., Lange, S., Bornemann, N., Morgenstern, A., Schreiber, P., Wille, C., Chadburn, S., Gouttevin, I., Burke, E., & Kutzbach, L. (2019). A 16-year record (2002–2017) of permafrost, active-layer, and meteorological conditions at the Samoylov Island Arctic permafrost research site, Lena River delta, northern Siberia: an opportunity to validate remote-sensing data and land surface, snow, and permafrost models. *Earth System Science Data*, *11*, 261–299.

Boike, J., Veh, G., Stoof, G., Sachs, T., Busse, H., & Muster, S. (2015). Near-infrared orthomosaic of Samoylov Island, Siberia, summer 2014 (854 MB).

Boike, J., Wille, C., & Abnizova, A. (2008). Climatology and summer energy and water balance of polygonal tundra in the Lena River Delta, Siberia. *Journal of Geophysical Research*, *113*.

Bratsch, S., Epstein, H., Buchhorn, M., & Walker, D. (2016). Differentiating among Four Arctic Tundra Plant Communities at Ivotuk, Alaska Using Field Spectroscopy. *Remote Sensing*, *8*, 51.

Buchhorn, M., Walker, D., Heim, B., Raynolds, M., Epstein, H., & Schwieder, M. (2013). Ground-Based Hyperspectral Characterization of Alaska Tundra Vegetation along Environmental Gradients. *Remote Sensing*, *5*, 3971–4005.

Coops, N.C., Stone, C., Culvenor, D.S., Chisholm, L.A., & Merton, R.N. (2003). Chlorophyll content in eucalypt vegetation at the leaf and canopy scales as derived from high resolution spectral data. *Tree physiology*, *23*, 23–31.

Döpper, V., Panda, S., Waigl, C., Braun, M., & Feilhauer, H. (2021). Using floristic gradient mapping to assess seasonal thaw depth in interior Alaska. *Applied Vegetation Science*, *24*.

Evans, B.M., Walker, D.A., Benson, C.S., Nordstrand, E.A., & Petersen, G.W. (1989). Spatial interrelationships between terrain, snow distribution and vegetation patterns at an arctic foothills site in Alaska. *Ecography*, *12*, 270–278.

Feilhauer, H., Faude, U., & Schmidtlein, S. (2011). Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. *Remote Sensing of Environment*, *115*, 2513–2524.

Feilhauer, H., Oerke, E.-C., & Schmidtlein, S. (2010). Quantifying empirical relations between planted species mixtures and canopy reflectance with PROTEST. *Remote Sensing of Environment*, *114*, 1513–1521.

Feilhauer, H., Zlinszky, A., Kania, A., Foody, G.M., Doktor, D., Lausch, A., & Schmidtlein, S. (2021). Let your maps be fuzzy!—Class probabilities and floristic gradients as alternatives to crisp mapping for remote sensing of vegetation. *Remote Sensing in Ecology and Conservation*, *7*, 292–305.

Forbes, B.C., & Sumina, O.I. (1999). Comparative Ordination of Low Arctic Vegetation Recovering from Disturbance: Reconciling Two Contrasting Approaches for Field Data Collection. *Arctic, Antarctic, and Alpine Research*, *31*, 389–399.

Fuchs, M., Grosse, G., Strauss, J., Günther, F., Grigoriev, M., Maximov, G.M., & Hugelius, G. (2018). Carbon and nitrogen pools in thermokarst-affected permafrost landscapes in Arctic Siberia. *Biogeosciences*, *15*, 953–971.

Gausman, H.W. (1974). Leaf reflectance of near-infrared. *Photogram Eng*, *40*, 183–191.

Gleason, H.A. (1926). The Individualistic Concept of the Plant Association. *Bulletin of the Torrey Botanical Club*, *53*, 7.

Grosse, G., Schirrmeister, L., Siegert, C., Kunitsky, V.V., Slagoda, E.A., Andreev, A.A., & Dereviagyn, A.Y. (2007). Geological and geomorphological evolution of a sedimentary periglacial landscape in Northeast Siberia during the Late Quaternary. *Geomorphology*, *86*, 25–51.

Harris, A., Charnock, R., & Lucas, R.M. (2015). Hyperspectral remote sensing of peatland floristic gradients. *Remote Sensing of Environment*, *162*, 99–111.

Hijmans, R.J. (2022). raster: Geographic Data Analysis and Modeling. https://CRAN.R-project.org/package=raster.

Hill, M.O., & Gauch, H.G. (1980). Detrended Correspondence Analysis: An Improved Ordination Technique. In E. van der Maarel (Ed.), *Classification and Ordination* (pp. 47–58). Dordrecht: Springer Netherlands.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*, 417–441.

Kapfer, J., & Popova, K. (2021). Changes in subarctic vegetation after one century of land use and climate change. *Journal of Vegetation Science*, *32*.

Kruskal, J.B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.

Kumar, L., Schmidt, K., Dury, S., & Skidmore, A. (2002). Imaging Spectrometry and Vegetation Science. In F.D. van der Meer, M. Abrams, P. Curran, A. Dekker, S.M. de Jong, & M. Schaepman (Eds.), *Imaging Spectrometry* (pp. 111–155). Dordrecht: Springer Netherlands.

Leyer, I., & Wesche, K. (2008). *Multivariate Statistik in der Ökologie. Eine Einführung.* Berlin, Heidelberg: Springer.

Liland, K.H., Mevik, B.-H., & Wehrens, R. (2021). pls: Partial Least Squares and Principal Component Regression. https://CRAN.R-project.org/package=pls.

Liu, N., Budkewitsch, P., & Treitz, P. (2017). Examining spectral reflectance features related to Arctic percent vegetation cover: Implications for hyperspectral remote sensing of Arctic tundra. *Remote Sensing of Environment*, *192*, 58–72.

Lopatin, J., Kattenborn, T., Galleguillos, M., Perez-Quezada, J.F., & Schmidtlein, S. (2019). Using aboveground vegetation attributes as proxies for mapping peatland belowground carbon stocks. *Remote Sensing of Environment*, *231*, 111217.

Macander, M., Frost, G., Nelson, P., & Swingley, C. (2017). Regional Quantitative Cover Mapping of Tundra Plant Functional Types in Arctic Alaska. *Remote Sensing*, *9*, 1024.

Muster, S., Langer, M., Heim, B., Westermann, S., & Boike, J. (2012). Subpixel heterogeneity of ice-wedge polygonal tundra: a multi-scale analysis of land cover and

evapotranspiration in the Lena River Delta, Siberia. *Tellus B: Chemical and Physical Meteorology*, *64*, 17301.

Oksanen, J., Simpson, G., F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R.B. O'Hara, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, Matt Barbour, Michael Bedward, Ben Bolker, Daniel Borcard, Gustavo Carvalho, Michael Chirico, Miquel De Caceres, Sebastien Durand, Heloisa Beatriz Antoniazi Evangelista, Rich FitzJohn, Michael Friendly, Brendan Furneaux, Geoffrey Hannigan, Mark O. Hill, Leo Lahti, Dan McGlinn, Marie-Helene Ouellette, Eduardo Ribeiro Cunha, Tyler Smith, Adrian Stier, Cajo J.F. Ter Braak, & James Weedon. *vegan: Community Ecology Package*.

Park, H., Kim, Y., & Kimball, J.S. (2016). Widespread permafrost vulnerability and soil active layer increases over the high northern latitudes inferred from satellite remote sensing and process model assessments. *Remote Sensing of Environment*, *175*, 349–358.

Pignatale, F.C. (2020). S2 PGDS. http://step.esa.int/thirdparties/sen2cor/2.10.0/docs/S2-PDGS-MPC-L2A-SRN-V2.10.0.pdf.

R Core Team (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Raschke, E.A., & Savelieva, L.A. (2017). Subrecent spore–pollen spectra and modern vegetation from the Lena River Delta, Russian Arctic. *Contemporary Problems of Ecology*, *10*, 395–410.

Rudy, A.C., Lamoureux, S.F., Treitz, P., & Collingwood, A. (2013). Identifying permafrost slope disturbance using multi-temporal optical satellite images and change detection techniques. *Cold Regions Science and Technology*, *88*, 37–49.

Runge, A., Fuchs, M., Shevtsova, I., Landgraf, N., Heim, B., Herzschuh, U., & Grosse, G. (2022). Hyperspectral field spectrometry of Arctic vegetation units in the central Lena Delta: PANGAEA. https://doi.pangaea.de/10.1594/PANGAEA.945982.

Schickhoff, U., Walker, M.D., & Walker, D.A. (2002). Riparian willow communities on the Arctic Slope of Alaska and their environmental relationships: A classification and ordination analysis. *Phytocoenologia*, *32*, 145–204.

Schmidtlein, S., & Sassin, J. (2004). Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing of Environment*, *92*, 126–138.

Schmidtlein, S., Zimmermann, P., Schüpferling, R., & Weiß, C. (2007). Mapping the floristic continuum: Ordination space position estimated from imaging spectroscopy. *Journal of Vegetation Science*, *18*, 131–140.

Schneider, J., Grosse, G., & Wagner, D. (2009). Land cover classification of tundra environments in the Arctic Lena Delta based on Landsat 7 ETM+ data and its application for upscaling of methane emissions. *Remote Sensing of Environment*, *113*, 380–391.

Shevtsova, I., Herzschuh, U., Heim, B., Schulte, L., Stünzi, S., Pestryakova, L.A., Zakharov, E.S., & Kruse, S. (2021a). Recent above-ground biomass changes in central Chukotka (Russian Far East) using field sampling and Landsat satellite data. *Biogeosciences*, *18*, 3343–3366.

Shevtsova, I., Laschinskiy, N., Heim, B., & Herzschuh, U. (2021b). Foliage projective cover of 26 vegetation sites of central Lena Delta from 2018: PANGAEA.

Syms, C. (2008). Ordination. In B.D. Fath, & S.E. Jørgensen (Eds.), *Encyclopedia of ecology* (pp. 2572–2581). Oxford: Elsevier.

Thomas, V., Treitz, P., Jelinski, D., Miller, J., Lafleur, P., & McCaughey, J. (2003). Image classification of a northern peatland complex using spectral and plant community data. *Remote Sensing of Environment*, *84*, 83–99.

Ulrich, M., Grosse, G., Chabrillat, S., & Schirrmeister, L. (2009). Spectral characterization of periglacial surfaces and geomorphological units in the Arctic Lena Delta using field spectrometry and remote sensing. *Remote Sensing of Environment*, *113*, 1220–1235.

van der Linden, S., Rabe, A., Held, M., Jakimow, B., Leitão, P., Okujeni, A., Schwieder, M., Suess, S., & Hostert, P. (2015). The EnMAP-Box—A Toolbox and Application Programming Interface for EnMAP Data Processing. *Remote Sensing*, *7*, 11249–11266.

van Wijk, M.T., Williams, M., & Shaver, G.R. (2005). Tight coupling between leaf area index and foliage N content in arctic plant communities. *Oecologia*, *142*, 421–427.

Virtanen, R., Oksanen, J., Oksanen, L., & Razzhivin, V.Y. (2006). Broad-scale vegetation-environment relationships in Eurasian high-latitude areas. *Journal of Vegetation Science*, *17*, 519–528.

Walker, D.A., Daniëls, F.J., Matveyeva, N.V., Šibík, J., Walker, M.D., Breen, A.L., Druckenmiller, L.A., Raynolds, M.K., Bültmann, H., Hennekens, S., Buchhorn, M., Epstein, H.E., Ermokhina, K., Fosaa, A.M., Heiðmarsson, S., Heim, B., Jónsdóttir, I.S., Koroleva, N., Lévesque, E., MacKenzie, W.H., Henry, G.H., Nilsen, L., Peet, R., Razzhivin, V., Talbot, S.S.,

Telyatnikov, M., Thannheiser, D., Webber, P.J., & Wirth, L.M. (2018). Circumpolar Arctic Vegetation Classification. *Phytocoenologia*, *48*, 181–201.

Walker, D.A., Jia, G.J., Epstein, H.E., Raynolds, M.K., Chapin III, F.S., Copass, C., Hinzman, L.D., Knudson, J.A., Maier, H.A., Michaelson, G.J., Nelson, F., Ping, C.L., Romanovsky, V.E., & Shiklomanov, N. (2003). Vegetation-soil-thaw-depth relationships along a low-arctic bioclimate gradient, Alaska: synthesis of information from the ATLAS studies. *Permafrost and Periglacial Processes*, *14*, 103–123.

Walker, D.A., Raynolds, M.K., Daniëls, F.J., Einarsson, E., Elvebakk, A., Gould, W.A., Katenin, A.E., Kholod, S.S., Markon, C.J., Melnikov, E.S., Moskalenko, N.G., Talbot, S.S., Yurtsev, B.A., & Team, T.o.m.o.t.C. (2005). The Circumpolar Arctic vegetation map. *Journal of Vegetation Science*, *16*, 267–282.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109–130.

Appendix 1: Description of Sampling sites (Runge et al., 2021)

| ID | Date of recording | Landform | Description | Picture |
|---|---|---|---|---|
| **3** | 07.08.2018 | Alas | Alas with moss and sometimes sedges |  |
| **5** | 08.08.2018 | Drained Lake Basin | Mosses with sedges and tussocks and single shrubs |  |
| **6** | 08.08.2018 | Drained Lake Basin | Mosses with sedges and tussocks and single shrubs |  |
| **7** | 08.08.2018 | Yedoma Upland | Dry and wet areas |  |
| **8** | 09.08.2018 | Yedoma Upland | Dry and wet areas |  |

| 9 | 09.08.2018 | Yedoma Upland | Dry and wet areas |  |
|---|---|---|---|---|
| 10 | 09.08.2018 | Yedoma Slope | Mosses and ground covering plants |  |
| 11 | 09.08.2018 | Shrub Field | Shrubs and dwarf shrubs |  |
| 13 | 10.08.2018 | Yedoma Upland | Wet parts: mosses, moist parts: sedges, dry parts: shrubs |  |
| 14 | 10.08.2018 | Yedoma Upland near slope | Polygonal tundra, wet |  |

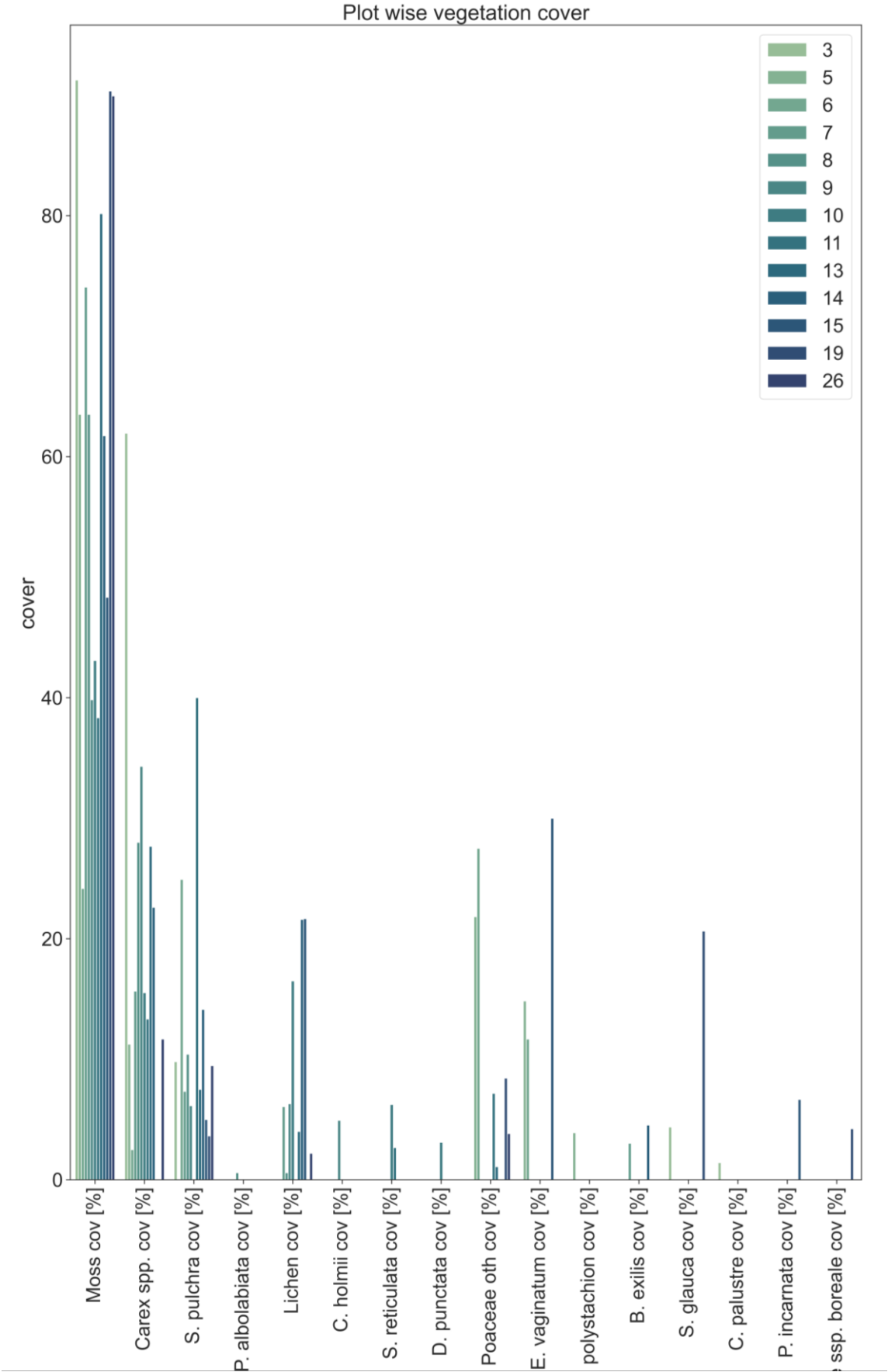| 15 | 10.08.2018 | Yedoma Slope | Mosses and E. vaginatum | |
| 19 | 26.08.2018 | Floodplain | Old floodplain with shrubs, mainly S. glauca |  |
| 26 | 28.08.2018 | Yedoma Upland | Moist and dry polygonal tundra. Mosses and dwarf shurbs |  |

Appendix 2: Distribution of species along sampling sites (Shevtsova et al., 2021b)



Plot wise vegetation cover

**Declaration on my honour**

I, Mr Carl Stadie, born on 10.08.2000 in Rostock, hereby declare that I have written this Bachelor's thesis independently, using the sources and aids indicated. This thesis has not been submitted in any other examination procedure.

Greifswald 20.08.2022