ENVIRONMENTAL AND ANTHROPOGENIC FACTORS CO-SHAPE COMMUNITY-LEVEL PLANT SPECIES RICHNESS ACROSS THE WESTERN SIBERIAN ARCTIC

Factors co-shape W. Siberian tundra PSR

Abstract

Aim

The Arctic ecosystem and its species are exposed to amplified climate warming and in some regions to rapidly developing economic activity. This study strives to identify, model and map the patterns of community-level plant species richness in the Western Siberian Arctic and to estimate the relative impact of environmental and anthropogenic factors driving these patterns. With our results and methods, we aim at contributing towards conservation efforts for arctic species richness.

Location

Western Siberian Arctic, Russia.

Methods

We investigated the relative impact of environmental, topographic and anthropogenic factors on community-level plant species richness of the Western Siberian Arctic, using macroecological models trained with an extensive, newly assembled geobotanical dataset. We included vascular plants, mosses and lichens in our analysis, as non-vascular plants substantially contribute to species richness and ecosystem functions in the Arctic.

Results

We found that the mean community-level plant species richness in this vast Arctic region does not decrease with increasing latitude. Instead, we identified an increase in species richness from South-West to North-East, which can be explained by climatic, topographical and anthropogenic factors. We found that lowest species richness is associated with a medium (30 to 50 km) distance to infrastructure while neighboring (0 to 10 km) and remote (≈ 100 km) areas

have relatively high species richness. We also show that the existing protected areas cover only a small part of the areas with the highest species richness.

Conclusions

Our results reveal complex spatial patterns of community-level species richness in the Western Siberian Arctic. We suggest that the impact of economic activities on species richness is ambiguous and not limited to areas directly affected by infrastructure. We show that while community-level plant species richness is mostly driven by non-anthropogenic factors, human activities also contribute to the heterogeneous distribution of community-level species richness on a broad scale. Our approach and results can be used to develop nature protection strategies for other arctic regions facing similar challenges.

Keywords: Arctic, macroecological modeling, Arctic Vegetation Archive, community-level plant species richness, anthropogenic impact on biodiversity, tundra vegetation, anthropogenic change.

Introduction

The documentation of Arctic plant diversity and its distribution under global change is one of the key priorities of international science and policy agendas as coordinated by the Conservation of Arctic Flora and Fauna (CAFF, 1997) of the Arctic Council and the International Arctic Science Committee (IASC). This information is urgently needed for the identification of Arctic biodiversity hotspots, which are a major target for nature protection and conservation (UN Convention on Biological diversity, CBD) (CBD, 1992). Plant diversity in the Arctic is usually studied at regional (hundreds of square kilometres), local (square kilometres) and community (square meters) levels. Despite Arctic regional and (to a lesser extent) local plant diversity being relatively well documented, the community-level distribution of plant diversity across broad spatial extents and its drivers remains understudied, especially in the Siberian part of the Arctic (Walker et al. 1994;

Daniëls et al. 2005; Daniëls et al., 2013; Khitun et al. 2016). Yet, disentangling the relative impact of the environmental and anthropogenic drivers of diversity at the community level is important for understanding the distribution of Arctic vegetation and predicting its response to global change.

Species richness across plant communities in the Arctic is determined by an interaction of local abiotic factors such as soil moisture, meso- and microrelief, wind speed and exposure, permafrost and soil conditions (Schultz, 2005; Iturrate-Garcia et al., 2016; Walker et al., 2019), which can promote high heterogeneity among communities at small spatial scales. This heterogeneity is often larger than intra-regional differences between communities belonging to the same vegetation type (Khitun, 1998; Khitun and Rebristaya, 1998). Furthermore, anthropogenic factors play an increasingly important role in shaping Arctic vegetation, changing community composition, threatening some local species (especially, lichens and mosses) and simultaneously increasing total plant species richness through introduction of new species and habitat change (Forbes 1995; 1997; Rebristya and Khitun, 1998; Nellemann et al., 2001; Red book of YANAO, 2010; Daniëls et al. 2013, Povoroznyuk et al. 2022).

The Western Siberian tundra is a rapidly transforming region of the Arctic (Kumpula et al. 2011, 2012; Walker et. al, 2012; Kozlova, 2013). The combination of multiple interacting factors including climate change, infrastructure expansion, fossil fuel extraction (Skipin, 2014), reindeer pressure (Kryazhimskii et al., 2011; Egelkraut et al., 2020; Veselkin et al., 2021) and species invasions, contributes to large-scale ecosystem degradation within and beyond areas directly affected by economic activity (Forbes et al., 2009; Golovatin et al., 2012). The high landscape homogeneity (Rebristaya, 2013) at broad spatial scales (about 300’000 km^2) contrasting with the uneven spatial distribution of anthropogenic impacts make the Western Siberian tundra a

natural laboratory for studying the relative impact of environmental and anthropogenic drivers on tundra flora and vegetation across biological, temporal, and spatial scales.

Most of the botanical research in the Western Siberian tundra was conducted at the site level, following the 'local flora' methodology (Rebristaya et al., 1989; Rebristaya and Khitun, 1994, 1998; Khitun and Rebristaya, 1998; Khitun, 2002, 2003; Rebristaya, 2013; Khitun et al., 2016). This methodology is based on a complete assessment of vascular plant species in an area of 100-300 km^2. There are 42 local floras described across the Western Siberian tundra, but their distribution is uneven: about two thirds of the local floras were described on the Yamal peninsula, while other areas are poorly sampled. Local species pools vary widely: from 215 species in Layakha, west of Taz peninsula (Fig. 1), subzone E (Rebristaya et al.,1989; CAVM, 2003), and 209 species in Chugoryakha, south-west of Gydan (Fig.1), subzone E (Rebristaya and Khitun 1994; CAVM, 2003), to 75 species on Bely island, subzone B (Rebristaya, 1995; CAVM, 2003). Generally, regional species richness declines with latitude, but areas at the same latitude at Gydan have richer floras than at Yamal by 20 to 30 species (Khitun, 1998; Khitun, 2016; Rebristaya 2013). Although overall summer warmth has been identified as the main contributing factor to floristic richness gradients, other factors such as soil acidity, glaciation and sea level history of the area and its flora are also considered important (Walker et al. 2005; Khitun, 1998, 2016).

While earlier studies based on local floristic data provide important insight into regional vascular plant species richness, we still lack an understanding of which factors are structuring the species richness at the fine-grained community level across the Western Siberian Arctic and how climate, topographic and anthropogenic factors combine to impact community species richness across broad spatial extents. Large-scale quantitative studies of community-level species richness have not been carried out in Western Siberia, where existing studies rely either

on traditional geobotanical methods or are limited to smaller areas (Khitun, 1998; Forbes and Sumina, 1999; Rebristaya 2013). Here, based on a newly assembled, large geobotanical data set, we aim to identify the main drivers and map the patterns of community-level plant species richness, including vascular plants, mosses and lichens, in the Western Siberian tundra and the environmental and anthropogenic forces driving these patterns. We estimate the relative impact of different modern and historical environmental factors on plot-level community species richness using macroecological models. We model the distribution of mean plant species richness across the area and compare these predictions to previous geobotanical studies. We also discuss how economic activities contribute to the community-level patterns of plant species richness. In our study, we posit four core hypotheses: 1) climate factors (including paleoclimate) are more important in explaining patterns of community-level species richness than topographic factors, 2) anthropogenic factors are as important predictors as natural factors, 3) community-level plant species richness in the area follows the latitudinal diversity gradient , and 4) current protected areas do not sufficiently well cover regions with high species richness.

Methods

The objective of our research is estimating the distribution of plant species richness at the community level across the Western Siberian tundra. To this end, we used macroecological models, predicting mean plot-level plant species richness as a function of environmental factors (Guisan and Rahbek, 2011; Guisan et al., 2017). We also estimated the role of anthropogenic factors, using distance from infrastructure as a proxy for anthropogenic impact.

Study area

The Western Siberian tundra is located in the northern part of the Western Siberian plain and covers slightly more than 300’000 km^2. The area has a low plant species richness at the

126 regional level because of its landscape properties and geoclimatic history. According to the
127 floristic classification by Yurtsev (1994), used in the Circumpolar Arctic Vegetation map (CAVM
128 team, 2003), the area belongs to the European-West-Siberian province (Yamal-Gydan
129 subprovince) of the Arctic floristic region (Yurtsev, 1994). In comparison with neighboring
130 subprovinces, Yamal-Gydan is characterized by almost complete absence of endemism,
131 extremely low vascular plant species richness (the lowest in continental Russia), and a lack of
132 many montane species (Yurtsev, 1994; Khitun, 1998; Sekretareva, 1999; Rebristaya, 2013;
133 Daniëls et al., 2013). In total, the province harbors about 450 species of vascular plants
134 (Koroleva et al., 2011), 276 species of mosses (Chernyadyeva, 2001; Voronova and
135 Dyachenko, 2018) and 250 species of lichens (Magomedova et al., 2006). The flora of the area
136 was shaped by Quaternary climate oscillations as well as glaciations and marine
137 transgressions, which had an especially strong impact on Yamal (Rebristaya 2013; Stewart et
138 al., 2016). Landscape homogeneity, high soil acidity, and absence of bedrock exposure also
139 contribute to observed low species richness (Khitun 1998; Rebristaya, 2013).
140

141 ## Geobotanical plots

142 To estimate community-level species richness we used geobotanical data from the Russian
143 Arctic Vegetation Archive (Ermokhina et al., 2022). These data consist of 1483 Braun Blanquet
144 plots established in homogenous vegetation collected during 2005-2017 field campaigns in the
145 Western Siberian tundra (Fig. 1) (Zemlianskii et al., 2023). The data were collected following the
146 standard international Arctic Vegetation Archive protocol (Walker et al., 2013, 2016, 2018) and
147 include full species lists of vascular plants and (contrary to most other existing floristic studies of
148 the area) also bryophytes and lichens (Elven, 2011; Raynolds et al., 2013). For the 12 major
149 sites (100-150 km^2 sub-areas in which more than 60 plots were sampled), we collected data
150 representative for all vegetation types present in the area (at least 5 plots of related

communities per site). In addition, we used 10 minor sites with 4-21 plots per site. The plot size varied from 25 to 100 m^2 depending on community characteristics (Matveeva, 1998). We divided plots into two classes, large plots ($100m^2$) and small plots (less than $100m^2$), to test for the effect of plot size on species richness.

The plot-level species richness, which we calculated by plot-wise summing species richness of vascular plants, mosses and lichens (liverworts data was omitted because of uneven identification quality across the database) was used to build regression-type macroecological models. The response variable of our models was total species richness per community. To estimate latitudinal trends at the site-level, we also inferred lichen, moss, vascular plants and total species richness for each major site.

Environmental variables

For each geobotanical plot, we extracted corresponding data from a set of 48 environmental predictors describing climate, topography, vegetation, and anthropogenic impact (Appendix Table S1). Climatic predictors included wind speed from the Global Wind Atlas (https://globalwindatlas.info/), 19 bioclimatic variables (seasonal and annual statistics of temperature and precipitation) from CHELSA (Karger et al., 2017), mean ground temperature from ESA Global permafrost project (Obu, et al., 2019), and annual statistics of climate moisture index, total cloud cover, potential evapotranspiration, site water balance, and growing degree days from CHELSA-BIOCLIM+ (Brun et al., 2022). Topographic predictors included altitude (incl. standard deviation of altitude), topographic position index, log-transformed slope, and aspect, which were derived from the Arctic digital elevation model (Morin et al., 2016; Porter et al., 2018) and topographic wetness index (Marthews et al., 2015). Mean normalized difference vegetation index (NDVI) for the period July-August 2019-2020 as observed by MODIS was used

as vegetation-related predictor (https://modis.gsfc.nasa.gov/). A detailed list of predictors can be seen in Appendix Table S1. All predictors were reprojected, resampled, and aligned to a consistent grid in ESRI:102025 projection with 1000 m horizontal resolution. From this set of predictors, we selected 14 predictors for analysis based on univariate predictive performance (see Appendix Table S1), limited collinearity (absolute pairwise Pearson correlation coefficients <0.7), and ecological relevance (different types of factors such as temperature, wind, relief and precipitation were represented with at least one predictor, and factors with direct effects were preferred over factors with indirect effects). We also classified the plots to small (less than $100m^2$) and large ($100m^2$) to correct for the potential effect of plot size on species richness. Finally, we used distance to infrastructure as a proxy for anthropogenic impact, combining industrial activities and the resulting species invasion and increase of reindeer pressure into one single predictor. To this end, we downloaded all the available data for roads, railroads, settlements, industrial sites, and airports from OpenStreetMap (https://www.openstreetmap.org) and converted them to points. Then, we calculated the distance between each raster cell and the closest infrastructure point. Raster layers were reprojected in QGIS (version 3.12, https://www.qgis.org/) while the resampling and predictor selection was conducted in R (version 4.1.2, R Core Team, 2021) using the package raster (Bivand et al., 2021).

In addition to modern climate, we investigated the effect of five paleoclimatic variables (mean annual temperature, annual precipitation sum, paleo-elevation, distance to land ice, and maximum (latest) year in time-series where the location was covered by land ice) from the Last Glacial Maximum period using the CHELSA-TraCE21k dataset (Karger et al., 2021). The data for our geobotanical plots was obtained through an R script (https://github.com/jakobjassmann/cryo_db_v2). We assessed the univariate predictive performance of the historical predictors and assured limited collinearity to modern climate predictors.

Fitting macroecological models

We modeled species richness as a function of non-anthropogenic predictors using four different model algorithms: random forest (RF), gradient boosting machines (GBMs), generalized linear models (GLMs), and generalized additive models (GAMs) (see Appendix Table S1). For RF, we fitted 500 regression trees, considering three predictors for each tree. For GBMs, we set the number of trees to 80, the minimum number of data points per leaf to 10, the learning rate to 0.1 and the error distribution was to 'poisson'. For GLMs and GAMs we assumed a Poisson error distribution and used the 'log' link function. For GLMs, we defined a linear and a quadratic term for each predictor. For GAMs, we used smooth terms with four degrees of freedom. For GLMs and GAMs, we step-wise optimized the Akaike information criterion by removing uninformative predictor terms from the model equation. Macroecological models were fitted in the R environment (version 4.1.2) using the packages randomForest (Liaw and Wiener, 2002), gbm (Greenwell et al., 2020), and gam (Hastie, 2020).

Model validation and performance

We used 5-fold cross-validation to estimate model performance. Agreement between observed and predicted species richness was assessed using Spearman correlation coefficients and mean absolute error (MAE).

Spatial projections

We produced plant species richness maps based on ensembles of spatial projections. For all fitted models that fulfilled the performance criterion (Table 1), we generated ensemble spatial projections of species richness across the study area. Ensembles were generated using the mean of modelled species richness among the different algorithms. In addition, we derived the

model disagreement between algorithms as the prediction span (i.e., maximum - minimum predicted species richness among algorithms per pixel). Finally, we intersected the obtained richness map with a shapefile of existing protected areas borders (CAFF, 2010).

Estimation of anthropogenic impact related to economic activities

To assess the role of human impact related to economic activity, we tested the predictive power of distance to infrastructure derived from Open Street Maps (https://www.openstreetmap.org/) for species richness and examined its response curve (Appendix Fig. 1). To this end, we added distance to infrastructure as a predictor to macroecological models and updated the predictions of species richness. To disentangle anthropogenic impact from environmental factors we produced maps with anthropogenic impact set to zero (we indicated the maximum distance to infrastructure which was found to be 115.285 km) and compared the obtained maps with predictions made using environmental data only. To estimate potential influence of economic activity on model disagreement, we assessed the correlation between distance to infrastructure and model disagreement both for each plot and for a set of randomly sampled 10 x 10 km grid cells.

Results

Model performance

The average performance of our models was 0.58 Spearman correlation (0.59 Spearman correlation or 8.3 mean absolute error without GLM) and 8.1 mean absolute error. The best model was GBM which had 0.61 Spearman correlation and the lowest mean absolute error (7.9). For model fits to be considered for spatial projections and subsequent analyses, we defined a Spearman correlation coefficient of 0.55 as minimum model performance. We therefore removed GLM model from the ensemble predictions, du to too low performance.

Role of selected predictors

The predictive power of environmental variables shows that climate-related factors are better predictors of mean community-level plant species richness than factors related to topography or of human influence (Table 2). Our results confirm that community-level species richness in the Arctic is strongly linked to warmth, but the actual dependence is unexpected: lower mean ground temperature and fewer growing degree days are associated with higher species richness (found primarily in the more continental Gydan peninsula) (Appendix Fig. 2). The trend is different for low mean temperature of driest quarter (spring) where highest species richness is associated with intermediate temperatures (Appendix Fig. 2). Moisture factors are also important: the maximum and range of the climate moisture index as well as potential evapotranspiration (min) significantly contribute to model quality (Appendix Fig. 2). Cloud area fraction and mean wind speed show moderate predictive power (5 and 7% respectively). High species richness is associated with relatively low wind speed and cloud fraction (Appendix Fig. 2). Topographic relief factors are generally less important for community-level species richness than climate variables. Plant species richness has a positive correlation with slope and standard deviation of altitude (Appendix Fig. 2). They are the only two topographic relief predictors which have predictive power higher than 5%. Altitude, aspect, topographic wetness index and roughness (topographic position index), on the other hand, have very low predictive power. Plot size was omitted as a predictor during GAM and GLM stepwise variable reduction, so we consider the plot size effect as minor as long as the plot size is not enlarged further. Multivariate GBM and RF also show the same model performance with and without the use of plot size.

Paleoclimatic predictors

Paleoclimatic predictors show high predictive power, partly even higher than any contemporary temperature predictor used, yet they are strongly correlated with modern climate predictors. The

strongest paleoclimate predictor found is temperature from 12.1 thousand years ago, which alone explains 24.6% of the deviance (4.5% higher than that of the actual mean annual air temperature). The next four strongest paleo-predictors are also temperatures from the Pleistocene-Holocene boundary period (11.2-12.7 thousand years ago), which all demonstrate high predictive power (24.5-25.5% explained deviance). At the same time, all the strongest paleoclimatic predictors (temperature, precipitation, distance to land ice) exhibit high correlation with current mean ground temperature (0.95, 0.83 and 0.72 respectively for a 12.1-thousand-year-old time point) and often among each other. The data show no evidence of either land ice or sea water at the plot locations throughout the entire time period, though the altitude data seems to have low accuracy. The closest points to the continental ice are found to be Bely Island (180 km) and Nakhodka bay (Southern Yamal) (200 km).

Spatial patterns of community-level plant species richness

Our model results show a highly heterogeneous distribution of community-level plant species richness across the Western Siberian tundra (Fig. 2). Mean species richness of the model ensemble map varies from 20 species on Eastern Yamal, Bovanenkovo railroad area (Fig. 2, 4), to more than 40 in the Gydan national park area. The Yamal peninsula shows generally lower species richness than Gydan. Furthermore, longitudinal differences between Yamal and Gydan are generally higher than latitudinal differences within both peninsulas. Protected areas (except Gydan National park) generally cover areas with low species richness. At the same time, the main part of the species-rich area in Northern Gydan remains unprotected as well as smaller species-rich areas in Northern and Eastern Yamal.

Community-level plant species richness across the Western Siberian tundra does not decline with latitude

It is widely recognized that landscape-level or regional plant species richness in the Arctic tundra is strongly dependent on summer warmth and hence declines with latitude (Hypothesis 3). At the community level, we found an opposing trend: median species richness of lichens, mosses, and vascular plants increases with latitude (reduced warmth) (Fig. 2). As single predictor, latitude is a relatively strong predictor of community-level plant species richness across the area with 15% of deviance explained (Appendix Table S1). Given the high correlation of latitude with other, more direct predictors, it was not used in the final model but we clearly see a temperature-richness effect that is opposed to the expected decline on richness with latitude and associated decrease in temperatures.

Model disagreement and its spatial patterns

The ensemble of models shows low model disagreement (less than 6 species) in most parts of Gydan, Taz peninsula and some areas of Northern, Southern and coastal Eastern Yamal (Fig. 3). We identified high model uncertainty (7-14 species) in the Western and Central Yamal, around Novy port and the South Tambey gas field at the eastern coast of Yamal, and on both the Taz bay and on Bely island. All the areas with high model uncertainty apart from Bely Island are exposed to high anthropogenic impacts, so we additionally assessed the correlation between distance to infrastructure and model disagreement. Low positive correlation is present for both plot locations (0.39) and among randomly selected sites (0.3).

Anthropogenic impact as a predictor

To estimate the influence of anthropogenic impact on species richness we tested distance to infrastructure as a predictor, which has significant predictive power (12% explained deviance). The GAM response curve shows that the dependence of species richness from distance to infrastructure is non-linear whereby the closest and furthest located sites are associated with

highest species richness (Appendix Fig. 2). By contrast, the lowest species richness is associated with sites at intermediate distances (30 to 50 km).

The role of anthropogenic factors in shaping plant species richness is underlined by model projections. We tested two mapping options: (1) using the actual distance to infrastructure raster (Fig. 4) and (2) using a hypothetical zero human impact raster (distance to infrastructure was set to maximum value) (Fig. 5). We also calculated maximum disagreement between projections of these 2 options and their uncertainties (Appendix Fig. 3). In contrast to the 'environmental only' and the 'actual distance to infrastructure' projections, the map of 'zero human impact' shows relatively high species richness across vast areas of Yamal (especially in the South) and low species richness in South-Eastern Gydan. The 'actual distance to infrastructure' projection shows circular patterns of species richness around some infrastructure objects clearly present in Western and Central Yamal, but also in some areas in Taz and Gydan (Fig. 4). Model disagreement within the 'zero human impact' projection is significantly higher (up to more than 30 species in some areas of Southern Gydan and Central Yamal) than the 'actual distance to infrastructure' and the 'environmental-only' projections which have similar numbers (up to 12 species) and spatial patterns (Figs 3-5).

An additional test on model disagreements between 'actual distance to infrastructure', 'zero human impact' and 'environmental-only' projections shows that highest model uncertainty is associated with intermediate and sometimes neighboring distances (Appendix Fig. 3).

Discussion

Our models show a highly heterogeneous distribution of community-level plant species richness across the Western Siberian tundra. In the study, we tested four hypotheses. As we expected in H.1, climate factors such as ground temperature and precipitation play a key role in shaping community species richness while topographic relief plays a secondary role (Table 2). Contrary to the pattern common in the Arctic on the regional level, we do not find a decrease in

community-level species richness from South to North, but rather a consistent increase (H.2, Fig. 2). While being less important than environmental factors, anthropogenic impacts (measured as spatial distance to infrastructures) contribute to the spatial heterogeneity of species richness, whereby their impact is strongest in the southwestern part of the area and some parts of southern Gydan (H.3, Fig. 4-5; Appendix Fig. 3). It is identified as one of the factors that removes uncertainty in the results, increasing species richness close to infrastructure and decreasing it at intermediate distances (Fig. 4-5, Appendix Fig. 3) compared to large distances. Our results reveal the complex interplay of factors driving community species richness in the Western Siberian tundra, which include both natural and anthropogenic forces interacting with each other. Finally, in accordance with H.4, our results suggest that the most species-rich areas remain largely unprotected (Fig. 2).

Our results suggest increasing mean community-level species richness from South-West to North-East which does not fit well with the common view of a clear negative latitudinal richness gradient in the Arctic (Daniëls et al. 2000; Shultz, 2005; Walker et. al, 2005; Daniëls et al., 2013), but is in agreement with some earlier studies made on the site-level in Western Siberia (Khitun, 1998; Rebristaya, 2013). Potential explanations relate to the history and geography of the region. Northern Gydan, which contained refugia during the last ice age (Khitun, 1998), has a higher richness than the Yamal peninsula. The latter was completely covered by water during the middle Pleistocene and mostly during the late Pleistocene transgressions, while Gydan kept the terrestrial connections with relatively rich Taymyr and Central Siberian floras (Khitun, 1998; Rebristaya, 2013). The North-east of Gydan is also characterized by soils generally less acidic than Yamal, permitting several arctic-alpine species to migrate from the east and settle there (Khitun, 1998; CAVM, 2003). These idiosyncratic histories and connections may partly explain the reversal of the South-North trend there (Khitun, 1998). However, we do not detect an equally strong and consistent South-to-North increase in Yamal (Fig. 2, Fig. 4). But even here,

the projected community species richness is lowest in the southern part, which is at the same time the most well-connected area of the peninsula, adjacent to the Polar Urals. Hence, the complex factors are influencing this pattern, which were confirmed by our analysis.

The testing of historical climate predictors indicate that paleoclimate had a significant impact on plant species richness distribution, with paleoclimate predictors such as temperature and precipitation sometimes being more significant than the similar contemporary climate predictors which indicates a legacy effect of past climate on the contemporary, community-level richness patterns. According to the CHELSA-TraCE21k dataset our study area was not affected by glaciation or sea level change over the past 21,000 years, which differs from previous research on the region that indicated some sea transgressions during the Boreal age of Holocene (9,200-8,200 years ago), although not as significant as those in the Pleistocene (Rebristaya, 2013). It is challenging to separate the influence of current climate from the history of the area, as demonstrated at the example of Gydan, where the high species richness is mostly attributed to its historical development (Khitun, 1998).

The patchy but rather low predicted species richness in the southern and western parts of the study area (especially in southern Yamal) might also be partly a result of anthropogenic vegetation change through a combination of intense reindeer herding and gas extraction (Golovatin et al., 2010; Ektova and Morozova, 2015; Forbes, 2013; Golovnev et al., 2016, Veselkin et al., 2021). The expansion of oil and gas infrastructure drives changes in ecological communities on a broad spatial scale, through associated species invasions and reducing reindeer grazing ground (down to 50-90% within 3-10 km) at short distances and resulting increase of reindeer density at intermediate distances (Kryazhimskii et al., 2011; Nellemann et al. 2001). Gas drilling is currently expanding in Northern Gydan which could further endanger the most species-rich ecosystems only partially protected by Gydan National Park. Our model

results indicate that more targeted research is needed to investigate the direct and indirect impact of human activities such as industrial expansion and related herding density change, and how they affect plant species richness. Social studies suggest that in Yamal increased pressure on pastures could be driven by land deprivation, pushing the herders to graze reindeers on smaller fields (Forbes, 2013) which in turn could lead to plant diversity loss.

Plant species richness across the Western Siberian tundra is shaped by a combination of environmental and anthropogenic factors, whereby the influence of climate factors is the strongest. The capacity of nature reserves to protect plant species is limited because of their insufficient spatial coverage, low connectivity, sole focus on animal protection in some cases, and often weak protection status (Kalyakin et al. 2000; Barry et al., 2017). Therefore, plant diversity protection requires a complex social-ecological approach that is up to be developed. Reducing industrial activity and active participation of Nenets people should be part of the approach.

Supporting information

APPENDIX FIGURE 1: Distance to infrastructure (km) map.

APPENDIX FIGURE 2: GAM response curves.

APPENDIX FIGURE 3. Additional distance to infrastructure test results.

APPENDIX Table S1: Full list of tested predictors

APPENDIX S2: ODMAP protocol

References

Arctic Council, Conservation of Arctic Flora and Fauna Working Group (2010). CAFF Map No.58 - Protected Areas in the Arctic classed after their IUCN category.

Arefev, S.P., Bogdanov, V.D., Golovatin, M.G., Gorbunov, P.Yu., Grigoreva, O.V., Dyachenko, A.P., Ermakov, A.M., Ishenko, V.G., Zinovev, E.V., Zoteeva, E.A. and Knyazev, M.S., 2010. Red book of Yamalo-Nenets Autonomous District: animals, plants, fungi. [Arefev, S.P., Bogdanov, V.D., Golovatin, M.G., Gorbunov, P.Yu., Grigoreva, O.V., Dyachenko, A.P., Ermakov, A.M., Ishenko, V.G., Zinovev, E.V., Zoteeva, E.A. and Knyazev, M.S., 2010. Krasnaya kniga Yamalo-Neneckogo avtonomnogo okruga: zhivotnye, rasteniya, griby.]

Barry, T., Hólmgrímur, H., & Guðmundsdóttir S. (2017). Arctic protected areas in 2017: status and trends. *Biodiversity*, *18*(4), 186–195. https://doi.org/10.1080/14888386.2017.1390496

Bivand, R., Keitt, T., & Rowlingson, B. (2021). rgdal: Bindings for the 'Geospatial'Data Abstraction Library https://CRAN.

Broennimann, O., Di Cola, V., Petitpierre, B., Breiner, F., Scherrer, D., Manuela, D., Randin, C., Engler, R., Hordijk, W., Mod, H., & Pottier, J. (2014). Package 'ecospat'.

Brun, P., Zimmermann, N.E., Hari, C., Pellissier, L., & Karger, D.N. (2022). Global climate-related predictors at kilometre resolution for the past and future, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2022-212, in review.

CAVM team (2003). Circumpolar Arctic vegetation map." Conservation of Arctic Flora and Fauna (CAFF) Map No 1.

Chernyadyeva, I.V. (2001). Moss flora of Yamal Peninsula (West Siberian Arctic). *Arctoa*, *10*, 121. 10.15298/arctoa.10.13

Conservation of Arctic Flora and Fauna (CAFF). (1997) CPAN Progress Report 1997. 39.

Daniëls, F.J.A., Bültmann, H., Lünterbusch, C., & Wilhelm, M. (2000). Vegetation zones and biodiversity of the North American Arctic. *Ber. Reinh. Tuxen Ges.*, *12*, 131–151.

Daniëls, F.J.A., Elvebakk, A., Talbot, S.S., & Walker, D.A. (2005). Classification and mapping of arctic vegetation. *Phytocoenologia*, *35*, 715–1079. 10.1127/0340-269X/2005/0035-0715

Daniëls, F.J.A., Gillespie, L.J., Poulin, M., Afonina, O.M., Alsos, I.G., Aronsson, M., Bültmann, H., Ickert-Bond, S.M., Konstantinova, N., Lovejoy, C., & Väre, H. (2013). Chapter 9. Plants., *Arctic Biodiversity Assessment. Status and Trends in Arctic Biodiversity*, (pp. 258–301). Conservation of Arctic Flora and Fauna (CAFF).

Egelkraut, D., Barthelemy, H., & Olofsson, J. (n.d.). Reindeer trampling promotes vegetation changes in tundra heathlands: Results from a simulation experiment. Journal of Vegetation Science, 31(3), 476–486.

Ektova, S. N., & Morozova, L. M. (2015). Rate of recovery of lichen-dominated tundra vegetation after overgrazing at the Yamal Peninsula. Czech Polar Reports, 5(1), 27–32.

Elven, R. (2011). Annotated checklist of the Panarctic Flora (PAF).

Forbes, B.C. (1995) Tundra disturbance studies. III. Short-term effects of aeolian sand and dust, Yamal Region, northwest Siberia, Russia. Environmental Conservation 22: 335-344. https://doi.org/10.1017/S0376892900034901.

Forbes, B.C. (1997) Tundra disturbance studies. IV. Species establishment on anthropogenic primary surfaces, Yamal Peninsula, northwest Siberia, Russia. Polar Geography 21: 79-100. https://doi.org/10.1080/10889379709377619.

Forbes, B.C. and O.I. Sumina. (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. https://doi.org/10.1080/15230430.1999.12003323

Forbes, B.C. (2013). Cultural resilience of social-ecological systems in the Nenets and Yamal-Nenets Autonomous Okrugs, Russia: a focus on reindeer nomads of the tundra. *Ecology and society*, *18*(4). http://dx.doi.org/10.5751/ES-05791-180436

Forbes, B.C., Stammler, F., Kumpula, T., Meschtyb, N., Pajunen, A., & Kaarlejärvi, E. (2009). High resilience in the Yamal-Nenets social–ecological system, West Siberian Arctic, Russia. *Proceedings of the National Academy of Sciences*, *106*(52), 22041–22048. https://doi.org/10.1073/pnas.0908286106

Golovatin, M. G., Morozova, L. M., Ektova, S. N., & Paskhalny, S. P. (2010). The change of tundra biota at Yamal peninsula (the North of the Western Siberia, Russia) in connection with anthropogenic and climatic shifts. Tundras: Vegetation, wildlife and climate trends. New York: Nova Publishers, 1–46.

Golovatin, M., Morozova, L., & Ektova, S. (2012). Effect of reindeer overgrazing on vegetation and animals of tundra ecosystems of the Yamal peninsula. *Czech Polar Reports*, *2*(2), 80–91. https://doi.org/10.5817/CPR2012-2-8

Golovnev, A.V., Garin, N.P., & Kukanov, D.A. (2016). Olenevody Yamala (materialy k Atlasu kochevyh tekhnologij) [Reindeer herders of Yamal (materials for the Atlas of nomadic technologies)]. UrO RAS Publ.

Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers (2020). gbm: Generalized Boosted Regression Models. R package version 2.1.8.

Guisan, A., & Rahbek, C. (2011). SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, *38*(8), 1433-1444. https://doi.org/10.1111/j.1365-2699.2011.02550.x

493 Guisan, A., Thuiller, W., & Zimmermann, N.E. (2017). *Habitat suitability and distribution models:*
494 *with applications in R.*, Cambridge University Press. https://doi.org/10.1017/9781139028271

495 Hastie, T. (2020). gam: Generalized Additive Models. R package version 1.20.

496 Hijmans, R.J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro,
497 O.P., Bevan, A., Racine, E.B., Shortridge, A., & Hijmans, M.R.J. (2015). Package 'raster'.

498 Iturrate-Garcia, M., O'Brien, M.J., Khitun, O., Abiven, S., Niklaus, P.A., & Schaepman-Strub, G.
499 (2016). Interactive effects between plant functional types and soil factors on tundra species
500 diversity and community composition. *Ecology and evolution*, *6*(22), 8126–8137.
501 https://doi.org/10.1002/ece3.2548

502 Kalyakin, V.N., Romanenko, F.A., Molochaev, A.V., Rogacheva, E.V., & Syroechkovskii, E.E.
503 (2000). Gydanskiy zapovednik [Gydan natural reserve], in Zapovedniki Sibiri [Natural reserves
504 of Siberia], *2*, pp. 47–55.

505 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann,
506 N.E., Linder, H.P., & Kessler, M. (2016). CHELSA climatologies at high resolution for the earth's
507 land surface areas (Version 1.0).

508 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann,
509 N.E., Linder, H.P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land
510 surface areas. *Scientific data*, *4*(1), 1–20. https://doi.org/10.1038/sdata.2017.122

511 Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2021):
512 CHELSA-TraCE21k v1. 0. Downscaled transient temperature and precipitation data since the
513 last glacial maximum. Climate of the Past Discussions, 1-27.

514 Khitun, O.V. (1998). Sravnitelnyy analiz lokalnykh i partsialnykh flor v dvukh podzonakh
515 Zapadnosibirskoy Arktiki (p-ova Gydanskiy i Tazovskiy) [Comparative analysis of local and
516 partial floras in two subzones of the West Siberian Arctic (the Gydan and Taz Peninsulas)].
517 *Study of biological diversity by the methods of comparative floristics: Materials of the IV*
518 *workshop on comparative floristics.*, 173–201.

519 Khitun, O.V. (2002). Vnutrilandshaftnaya struktura flory okrestnostey ustia reki Tinikyakha
520 (severnyye gipoarkticheskiye tundry Gydanskogo poluostrova) [Intralandscape structure of flora
521 of the surroundings of the Tinikyakha River mouth (northern hypoarctic tundra in the Gydanskii
522 Peninsula)]. *Botanicheskii zhurnal*, *87*(8), 1–24.

523 Khitun, O.V. (2003). Analiz vnutrilandshaftnoy struktury flory srednego techeniya reki
524 Khalmeryakha (Gydanskiy p-ov) [Analysis of the intralandscape structure of flora of the
525 Khalmeryakha middle reaches (Gydansky Peninsula)]. *Botanicheskii zhurnal*, *88*(10), 21–39.

526 Khitun, O.V. (2016). Intralandscape differentiation of the local flora in the central part of the
527 Gydansky Peninsula (West Siberian Arctic). *Czech Polar Reports*, *6*(2), 191–209.
528 https://doi.org/10.5817/CPR2016-2-18

529 Khitun, O.V., & Rebristaya, O.V. (1998). Rastitelnost' i ekotopologicheskaya struktura flory
530 okrestnostey mysa Khonorasale (arkticheskiye tundry Gydanskogo poluostrova). [Vegetation
531 and ecotopological structure of Khorosale cape surroundings flora (arctic tundra of Gydan
532 peninsula)]. *Botanicheskii Zhurnal*, *83*(12), 21–37.

533 Khitun, O.V., Koroleva, T.M., Chinenko, S.V., Petrovsky, V.V., Pospelova, E.B., Pospelov, I.N.,
534 & Zverev, A. (2016). Applications of local floras for floristic subdivision and monitoring vascular
535 plant diversity in the Russian Arctic. *Arctic Science*, *2*(3), 103–126. https://doi.org/10.1139/as-
536 2015-0010

Koroleva, T.M., Zverev, A.A., Katenin, A.E., Petrovskii, V.V., Pospelova, E.B., Rebristaya, O.V., Khitun, O.V., Khodachek, E.A, & Chinenko, S.V. (2011). *Botanicheskii zhurnal*, *96*(2) (Dolgotnaya geograficheskaya struktura lokalnykh i regionalnykh flor Aziatskoy Arktiki. 2. [Longitudinal geographic structure of local and regional floras of the Asian Arctic. 2.]), 145–169.

Kozlova, A.E. (2013). Antropogennoe preobrazovaniye relief v usloviakh khozyaistvennogo osvoeniya territorii poluostrova Yamal. [Anthropogenic relief change during the economic development of Yamal]. *Izvestiya Rossiyskoy akademii nauk*, *4,* 87–94.

Kryazhimskii, F. V., Maklakov, K. V., Morozova, L. M., & Ektova, S. N. (2011). System analysis of biogeocenoses of the Yamal Peninsula: Simulation of the impact of large-herd reindeer breeding on vegetation. Russian Journal of Ecology, 42(5), 351–361. Kumpula, T., Forbes, B.C., Stammler, F., & Meschtyb, N. (2012). Dynamics of a coupled system: multi-resolution remote sensing in assessing social-ecological responses during 25 years of gas field development in Arctic Russia. *Remote Sensing*, *4*(4), 1046–1068. https://doi.org/10.3390/rs4041046

Kumpula, T., Pajunen, A., Kaarlejärvi, E., Forbes, B.C., & Stammler, F. (2011). Land use and land cover change in Arctic Russia: Ecological and social implications of industrial development. *Global Environmental Change*, *21*(2), 550–562. https://doi.org/10.1016/j.gloenvcha.2010.12.010

Liaw, A. & Wiener, M. (2002). Classification and Regression by Random Forest. *R News*, *2*, 18-22.

Magomedova, M.A., Morozova, L.M., Ektova, S.N., Rebristaya, O.V., Chernyadyeva, I.V., Potemkin, A.D., & Knyazev, M.S. (2006). *Poluostrov Yamal: rastitelniy pokrov [The Yamal Peninsula: vegetation cover]*, City-press.

559 Marthews, T. R., Dadson, S. J., Lehner, B., Abele, S., & Gedney, N. (2015). High-resolution
560 global topographic index values. NERC Environmental Information Data
561 Centre. https://doi.org/10.5285/6b0c4358-2bf3-4924-aa8f-793d468b92be

562 Matveeva, N.A. (1998). *Zonalnost' v rastitelnom pokrove Arktiki. [Zonation in the plant cover in*
563 *the Arctic]*, BIN RAN.

564 Morin, P., Porter, C., Cloutier, M., Howat, I., Noh, M.J., Willis, M., Bates, B., Willamson, C. &
565 Peterman, K. (2016). ArcticDEM: a publically available, high resolution elevation model of the
566 Arctic. *Egu general assembly conference abstracts.*

567 Nellemann, C. et al. (2001) GLOBIO Global methodology for mapping human impacts on the
568 biosphere: the Arctic 2050 scenario and global application. UNEP/DEWA Technical Report 3,
569 47 pp.

570 Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H.H., Dashtseren, A.,
571 Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., & Khomutov, A. (2019). Northern
572 Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km^2 scale. *Earth-*
573 *Science Reviews*, *193*, 299-316. https://doi.org/10.1016/j.earscirev.2019.04.023

574 Porter, C., Morin, P., Howat, I., Noh, M.J., Bates, B., Peterman, K., Keesey, S., Schlenk, M.,
575 Gardiner, J., Tomko, K., & Willis, M. (2018). ArcticDEM, V1, Harvard Dataverse.
576 https://doi.org/10.7910/DVN/OHHUKH

577 Povoroznyuk, O., Vincent, W.F., Schweitzer, P., Laptander, R., Bennett, M., Calmels, F.,
578 Sergeev, D., Arp, C., Forbes, B.C., Roy-Léveillée, P. & Walker, D.A. (2022). Arctic roads and
579 railways: Social and environmental consequences of transport infrastructure in the circumpolar
580 North. Arctic Science. https://doi.org/10.1139/AS-2021-0033

QGIS Development Team (2022). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Version 3.12. Retrieved June 30, 2022, from http://qgis.osgeo.org.

Raynolds, M.K., Breen, A.L., Walker, D.A., Elven, R., Belland, R., Konstantinova, N., Kristinsson, H., & Hennekens, S. (2013). The Pan-Arctic Species List (PASL). *Arctic Vegetation Archive (AVA) Workshop.*

Rebristaya, O.V. (1995). Vascular plants of the Bely Island (Kara Sea). *Botanicheskii zhurnal*, *80*(7), 16–36.

Rebristaya, O.V. (2013). *Flora poluostrova Yamal: sovremennoye sostoyaniye i istoriya formirovaniya. [Flora of the Yamal peninsula: modern state and history of the formation]*, Izd-vo SPbGETU «LETI».

Rebristaya, O.V., & Khitun, O.V. (1998). Botaniko-geograficheskiye osobennosti flory Tsentralnogo Yamala [Botanical-geographic peculiarities of the flora of Central Yamal]. *Botanicheskii zhurnal*, *83*(7), 37–52.

Rebristaya, O.V., & Khitun, O.V. (1994). Flora sosudistykh rasteniy nizovyev reki Chugoriakha (yugo-zapadnaya chast Gydanskogo poluostrova. Zapadnosibirskaya Arktika). [Vascular plants flora of the lower reaches of the Chugoryakha River (south-western part of the Gydansky Peninsula, Western Siberian Arctic).]. *Botanicheskii Zhurnal*, *79*(8), 68–77.

Rebristaya, O.V., Tvorogov, V.A., & Khitun, O.V. (1989). Flora Tazovskogo poluostrova (sever Zapadnoy Sibiri) [Flora of the Tazovsy Peninsula (north of the Western Siberia)]. *Botanicheskii zhurnal*, *74*(1), 22–35.

Schultz, J. (2005). *The ecozones of the world.*, Springer. https://doi.org/10.1007/3-540-28527-X

603 Sekretareva, N.A. (1999). *The vascular plants of the Russian Arctic and adjacent territories.*,
604 Pensoft.

605 Skipin, L.N., Galyamov, A.A., Gaevaya, E.V., & Zakharova, E.V. (2014). Ekologicheskaya
606 otsenka sostoyaniya pochvennogo pokrova oleniikh pastbish poluostrova Yamal [The ecological
607 assessment of reindeer pastures soils of the Yamal peninsula]. *Agroprodovolstvennaya politika*
608 *Rossii*, *4,* 29–32.

609 Stewart, L., Alsos, I.G., Bay, C., Breen, A.L., Brochmann, C., Boulanger-Lapointe, N.,
610 Broennimann, O., Bültmann, H., Bøcher, P.K., Damgaard, C., & Daniëls, F.J. (2016). The
611 regional species richness and genetic diversity of Arctic vegetation reflect both past glaciations
612 and current climate. *Global Ecology and Biogeography*, *25*(4), 430–442.
613 https://doi.org/10.1111/geb.12424

614 Veselkin, D., Morozova, L., & Gorbunova, A. (2021). Decrease of NDVI values in the southern
615 tundra of Yamal in 2001–2018 correlates with the size of domesticated reindeer population.
616 *Mod. Probl. Remote Sens. Earth Space*, *18*, 143-155. https://doi.org/10.21046/2070-7401-2021-
617 18-2-143-155

618 Voronova, O.G., & Diachenko, A.P. (2018). Bryophyte flora of hydrocarbon deposit sites in the
619 Yamalo-Nenets Autonomous Okrug. *Vestnik Tomskogo gosudarstvennogo universiteta*.
620 *Biologiya, 42*. https://doi.org/10.17223/19988591/42/6

621 Walker, D.A., Breen, A.L., Raynolds, M.K., & Walker, M.D. (2013). Arctic Vegetation Archive
622 (AVA) Workshop.

623 Walker, D.A., Daniëls, F.J., Matveyeva, N.V., Šibík, J., Walker, M.D., Breen, A.L.,
624 Druckenmiller, L.A., Raynolds, M.K., Bültmann, H., Hennekens, S., & Buchhorn, M. (2018).

Circumpolar arctic vegetation classification. *Phytocoenologia*, *48*(2), 181–201. https://doi.org/10.1127/phyto/2017/0192

Walker, D.A., Daniëls, F.J.A., Alsos, I., Bhatt, U.S., Breen, A.L., Buchhorn, M., Bültmann, H., Druckenmiller, L.A., Edwards, M.E., Ehrich, D., & Epstein, H.E. (2016). Circumpolar Arctic vegetation: a hierarchic review and roadmap toward an internationally consistent approach to survey, archive and classify tundra plot data. *Environmental Research Letters*, *11*(5), 055005. https://doi.org/10.1088/1748-9326/11/5/055005

Walker, D.A., Epstein, H.E., Raynolds, M.K., Kuss, P., Kopecky, M.A., Frost, G.V., Daniëls, F.J.A., Leibman, M.O., Moskalenko, N.G., Matyshak, G.V., & Khitun, O.V. (2012). Environment, vegetation and greenness (NDVI) along the North America and Eurasia Arctic transects. *Environmental Research Letters, 7*(1), 015504. https://doi.org/10.1088/1748-9326/7/1/015504

Walker, D.A., Epstein, H.E., Šibík, J., Bhatt, U., Romanovsky, V.E., Breen, A.L., Chasníková, S., Daanen, R., Druckenmiller, L.A., Ermokhina, K., & Forbes, B.C. (2019). Vegetation on mesic loamy and sandy soils along a 1700-km maritime Eurasia Arctic Transect. *Applied vegetation science*, *22*(1), 150–167. https://doi.org/10.1111/avsc.12401

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. (2005). The circumpolar Arctic vegetation map. *Journal of Vegetation Science*, *16*(3), 267–282. https://doi.org/10.1111/j.1654-1103.2005.tb02365.x

Walker, M.D., Daniëls, F.J.A., & van der Maarel, E. (1994). Circumpolar arctic vegetation – Introduction and perspective. *Journal of Vegetation Science*, *5*, 758–764. https://doi.org/10.1111/j.1654-1103.1994.tb00395.x

Yurtsev, B.A. (1994). Floristic division of the Arctic. *Journal of Vegetation Science*, *5*(6), 765–776. https://doi.org/10.2307/3236191

Zemlianskii, V. , Ermokhina, K., Schaepman-Strub, G., Matveyeva, N., Troeva, E., Lavrinenko, I., Telyatnikov, M., Pospelov, I., Koroleva, N., Leonova, N., Khitun, O., Walker, D., Breen, A., Kadetov, N., Lavrinenko, O., Ivleva, T., Kholod, S., Petrzhik, N., Kurysheva, M., Gunin, Y., Lapina, A., Korolev, D., Kudr, E., & Plekhanova, E.. Russian Arctic Vegetation Archive – a new database of plant community composition and environmental conditions. *Global Ecology and Biogeography*, in review.

Zurell, D., Franklin, J., König, C., Bouchet, P.J., Serra-Diaz, J.M., Dormann, C.F., Elith, J., Fandos Guzman, G., Feng, X., Guillera-Arroita, G., Guisan A., Leitão, P.J., Lahoz-Monfort, J.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Thuiller, W., Yates K.L., Zimmermann, N.E., Merow, C. (2020). A standard protocol for describing species distribution models. *Ecography 43*, 1261-1277. DOI: 10.1111/ecog.04960

Global Wind Atlas. Retrieved June 30, 2022, from https://globalwindatlas.info/

Ermokhina K., Zemlianskii V., Kurysheva M., & Korolev D. Russian Arctic Vegetation Archive website. Retrieved June 30, 2022, from https://avarus.space/

Open Street map. Retrieved June 30, 2022, from https://www.openstreetmap.org

ORNL DAAC 2018. MODIS and VIIRS Land Products Global Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Retrieved Semtember 23, 2021 https://modis.gsfc.nasa.gov/

(1992) UN Convention on Biological diversity, CBD.Assmann J. CHARTER WP4 Holocene Cryo DB v2. Retrieved January 25, 2023, from https://github.com/jakobjassmann/cryo_db_v2

Data availability statement: Data available from the Dryad Digital Repository (https://datadryad.org/stash/share/bFWEuics4IXhXfj2xvo4or1sUYa-WriskoaRUuoVdeU) (Zemlianskii et al., 2022).

Figures and tables

Table 1: Model performance statistics from 5-fold cross-validation (GLM = generalized linear model, GAM = generalized additive model, GBM = gradient boosting machine). Red color shows GLM, which we omitted because of its comparably low performance (<0.55 Spearman correlation). The numbers in brackets indicate the change in performance upon adding the "distance to infrastructure" predictor.

Model	Spearman correlation	Mean Absolute Error	Root Mean Square Error
GLM	0.53	8.4 (8.3)	10.8 (10.7)
Random Forest	0.59	8.1	10.3
GAM	0.59	8.0	10.2
GBM	0.60 (0.61)	7.9	10.1 (10.0)

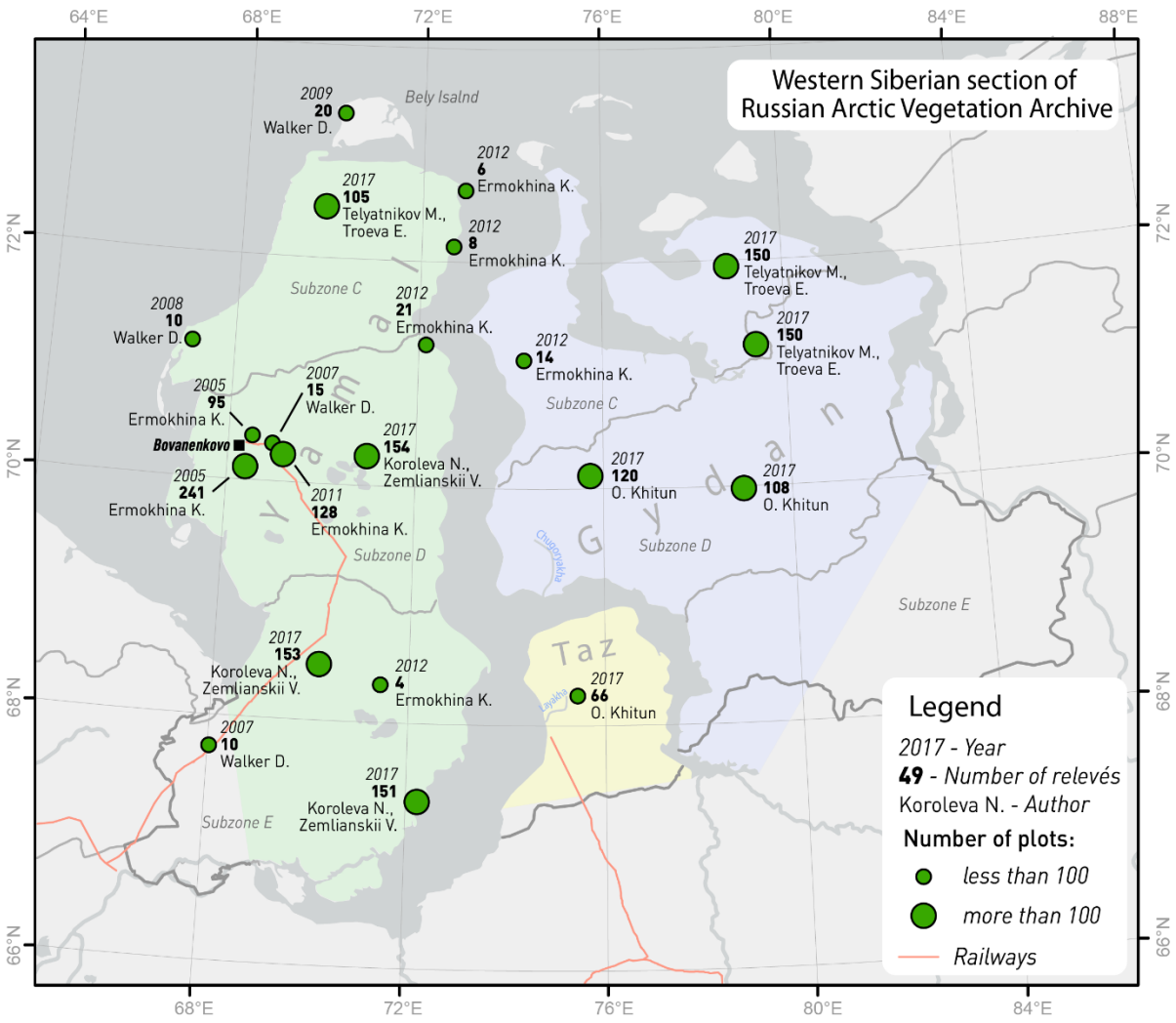
679 **Table 2:** Environmental variables used in the model. The full list of evaluated variables is
680 presented in Appendix, Table 1.

N	Predictors	Predictive power (%)	Original spatial res. (m)	Source
1	Mean Ground Temperature (2000-2016)	19	1000	ESA Global permafrost project (Obu, et al., 2019)
2	Potential evapotranspiration (min)	17	30 arcsec (<1000)	CHELSA new (Brun et al., 2022)
3	Mean temperature of driest quarter	14	30 arcsec (<1000)	CHELSA Bioclim (Karger et al., 2016)
4	Climate moisture index (max)	13	30 arcsec (<1000)	CHELSA new (Brun et al., 2022)
5	Distance to infrastructure	11	1000	OSM based (https://www.openstreetmap.org/)
6	Growing degree days above 5°C	11	30 arcsec (<1000)	Extended Bioclim (Karger et al., 2017)
7	Climate moisture index (range)	11	30 arcsec (<1000)	CHELSA new (Brun et al., 2022)

8	(log transformed) slope	10	10	ArcticDEM based (Morin et al., 2016; Porter et al., 2018)
9	Cloud area fraction	7	30 arcsec (<1000)	CHELSA Bioclim (Karger et al., 2016)
10	Mean wind speed	5	100	Global Wind Atlas (https://globalwindatlas.info/)

681

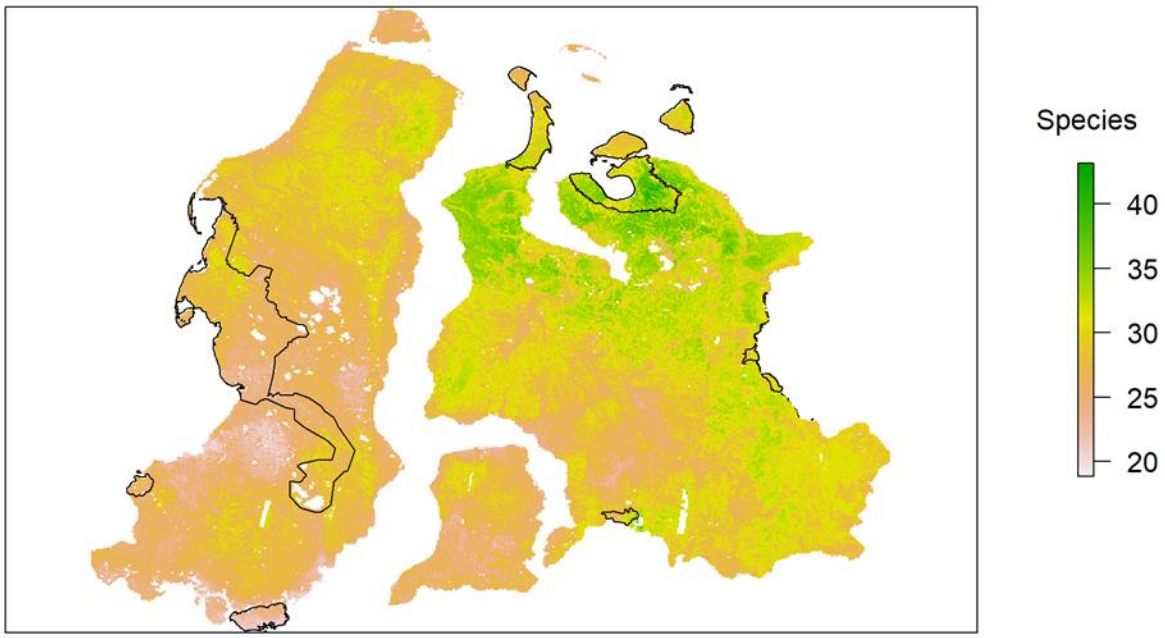
682



683

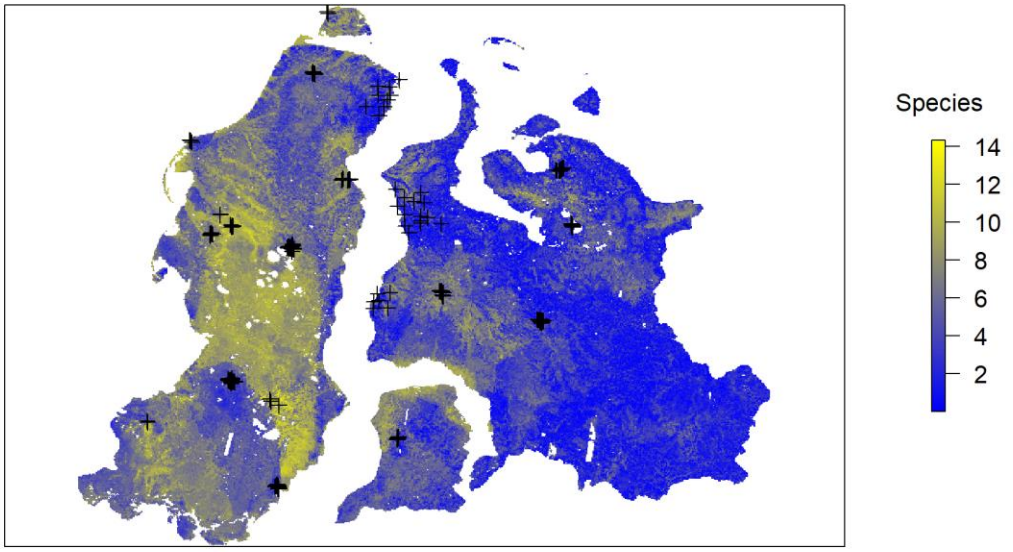
684 **Figure 1:** Western Siberian study area, including the location of the major study sites and respective number of geobotanical plots

685 per site (= number of relevés). The Yamal peninsula is shaded in green, Taz in yellow and Gydan in blue.



686

687 **Figure 2:** Mean plant species richness distribution in the Western Siberian tundra as predicted by a macroecological model
688 ensemble based on a general additive (GAM), gradient boosting machine (GBM) and random forest (RF) model. Black borders
689 show existing protected areas.



690

691 **Figure 3:** Model disagreement map indicating maximum difference in species number between GAM, GBM and Random forest

692 ('environmental only' projection). Black crosses are geobotanical plot locations.

693

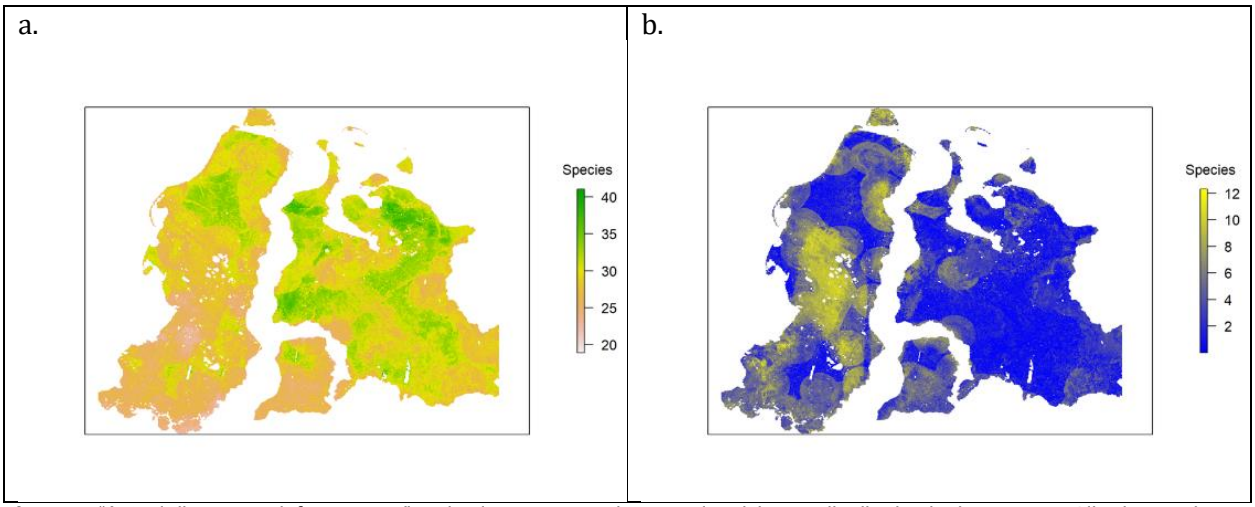


Figure 4. "Actual distance to infrastructure" projection. **a.** Mean plant species richness distribution in the Western Siberian tundra as predicted by a macroecological model ensemble based on a general additive (GAM), gradient boosting machine (GBM) and random forest model, similar to Figure 2, but additionally including the predictor 'distance to infrastructure'. **b.** Model disagreement map indicating maximum difference in species number between general additive (GAM), gradient boosting machine (GBM) and random forest model) similar to Figure 3, but additionally including the predictor 'distance to infrastructure'.

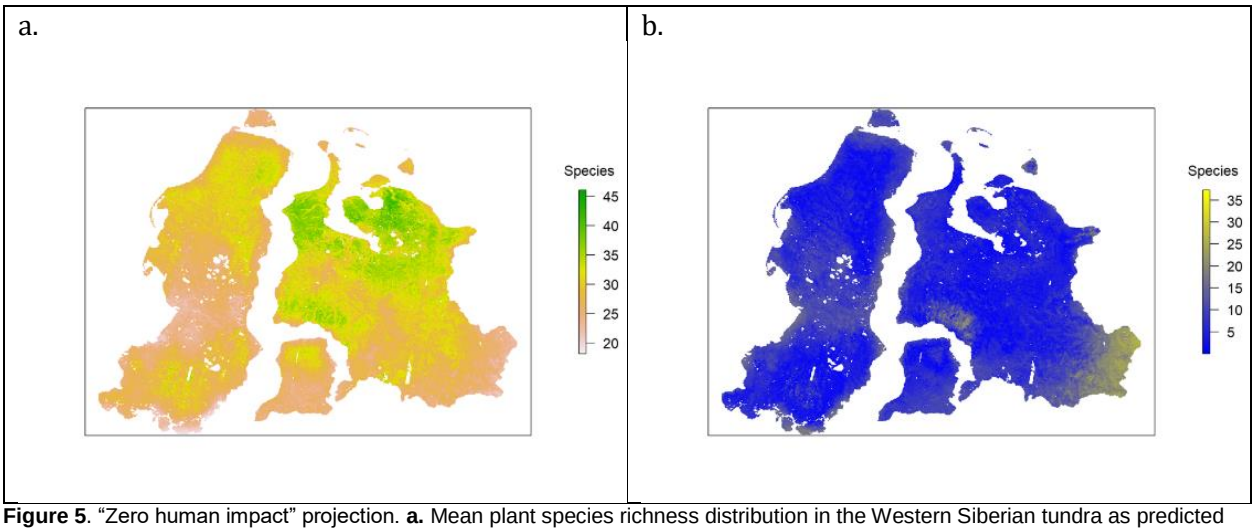


Figure 5. "Zero human impact" projection. **a.** Mean plant species richness distribution in the Western Siberian tundra as predicted by a macroecological model ensemble based on a general additive (GAM), gradient boosting machine (GBM) and random forest model, similar to Figure 2, but additionally including the predictor 'distance to infrastructure' set to the maximum value (115.285 km). **b.** Model disagreement map indicating maximum difference in species number between general additive (GAM), gradient boosting machine (GBM) and random forest model similar to Figure 3, but additionally including the predictor 'distance to infrastructure' set to the maximum value (115.285 km).