Poor transferability of machine learning classification models: Classifying vegetation on palsa complexes in Arctic Sweden

Holly Dancer

University of Nottihngham

Keywords: palsa, peatland, Arctic, vegetation mapping, model transferability, multispectral

Abstract

The Arctic is under great pressure from a warming climate, and areas at particular risk are those underlain by permafrost. Palsas are raised peatlands situated on permafrost, and are currently greatly degrading and releasing methane, a powerful greenhouse gas, into the atmosphere. As palsas degrade, vegetation and methane production change, and measuring these changes is crucial for estimating the emissions from these ecosystems. Assessing the ability of machine learning (ML) vegetation classification models to transfer between different palsa complexes across the region is important for mapping these ecosystems efficiently. In this project, two ML classifiers (random forest (RF) and support vector machines (SVM)) were trained on data from three palsa complexes and tested on a further three sites. Despite high overall accuracies on two of the training sites (>80% overall accuracy for the two most accurate models), both models performed poorly when transferred to the testing sites, despite two different sampling methods being trialled (overall accuracies for all testing sites <45%). The testing sites closer to a training site had better overall accuracies and were most impacted by the removal of the closest training site from the training data, with minimum decreases of 5.9% and 12% for these two sites, suggesting variation in vegetation classes across the region.

There were multiple possible reasons for the poor performance of these models, including the small number of training sites and training data, and the limitations of the ML methods themselves. Due to the mixtures of species that make up the vegetation classes, a model that allows for the input of swatches of image rather than the point values that are necessary for the ML methods is suggested to improve accuracy, such as a Convolutional Neural Network (CNN).

1. Introduction

The Arctic region is experiencing warming at a rate four times faster than the global average (Rantanen et al., 2022), meaning that understanding and quantifying the impact of warming is crucial, especially due to the presence of permafrost. Permafrost is any material found in the ground (soil, bedrock, etc.) that remains at or below 0 °C for a minimum of two years. The soil above permafrost that thaws seasonally is known as the active layer, where biological activity happens. The location and thickness of the permafrost (and active layer) is determined by a variety of site-specific factors that ultimately impact ground temperature (e.g. vegetation (Rouse, 1984), snow cover (Desrochers and Granberg, 1988), etc.), and as such, the same is true for the degradation of permafrost.

Permafrost is warming at a rate following the increase in air temperature, with global permafrost temperature increasing by 0.29 ± 0.12°C in the period from 2007 to 2016 (Biskaborn et al., 2019), leading to permafrost degradation. The degradation of permafrost has an impact on ecosystems up to the global scale due to the large amount of carbon stored in the Arctic region.

Approximately half of global soil carbon is stored in the Arctic, making it an area with potential to emit large amounts of greenhouse gases (GHGs) (Hugelius et al., 2020; Köchy et al., 2015). An ecosystem of particular importance in terms of carbon

storage are wetlands, as these ecosystems are estimated to cover around 3.7 ± 0.5 million km^2 of land, storing 415 ± 150 Pg C and
nearly half of the area covered and its carbon stocks are affected by permafrost (Hugelius et al., 2020).

Palsa complexes are systems of particular importance in the Arctic as they are greatly impacted by permafrost degradation due to their reliance on permafrost. Palsa complexes consist of raised plateaus of peat formed around a frozen core of peat which are commonly bordered by areas of former palsa which has now degraded, becoming waterlogged, and sometimes thermokarst lakes (Ballantyne, 2018; Zuidhoff and Kolstrup, 2000). Permafrost degradation in this ecosystem has been seen over large areas of the Arctic (Åkerman and Johansson, 2008; Borge et al., 2017; de la Barreda-Bautista et al., 2022; Luoto and Seppälä, 2003; Olvmo et al., 2020; Sannel et al., 2016; Sannel and Kuhry, 2011; Sjögersten et al., 2023), and such

degradation causes subsidence of the peat surface and eventually waterlogging (Olvmo et al., 2020; Sjöberg et al., 2015). Degradation of permafrost sees an increase in active layer (Åkerman and Johansson, 2008), which in palsa peatlands allows for an increased volume of peat within which biological activity can occur, such as the activity of methanogenic organisms which produce methane (CH_4). The shift from raised peat mounds to waterlogged wetland demonstrates a shift in ecosystem functioning in palsas where methane emissions greatly increase, as has been observed across a large number of sites in the Arctic (Glagolev et al., 2011; Miglovets et al., 2021; Sjögersten et al., 2023; Varner et al., 2022; Walter et al., 2006; Walter Anthony et al., 2016).

Methane is a GHG which has seen a large increase in atmospheric concentrations since the 1980's (Turner et al., 2019) and which have accelerated from 2020 to 2021 (Feng et al, 2023). The output of methane into the atmosphere is of great importance due to its high warming potential, 16-38 times that of carbon dioxide (CO_2) (IPCC, 2021). The factor most responsible for this recent acceleration is the interaction between a warmed atmosphere and wetlands causing an increased release of methane from these ecosystems (Turner et al., 2019; Zhang et al., 2023).

Variation in methane emissions in peatlands have been linked to vegetation (Sjögersten et al., 2023), so vegetation maps over a region may help in identifying areas of high methane emission. The variation in methane emissions is in part due to the species-specific difference of plants in emitting methane (Hodgkins et al., 2014; McLaughlin et al., 2023; Sjogersten et al., 2023; Strom et al., 2005).

Mapping of palsa vegetation on this scale would not be possible without the now wide-scale use of UAVs for Earth Observation (EO), which allow for affordable data capture at the centimetre scale (Assmann et al., 2018; Koh and Wich, 2012). UAVs allow for data capture even in such challenging and remote environments as the Arctic (Assmann et al., 2018; Fraser et al., 2016; Malenovský et al., 2017; Mora et al., 2015). Siewert and Olofsson (2020) emphasise how the use of UAVs allows for the capture of fine detail which cannot be captured from satellite images. The ability to capture a whole palsa complex in a short period of time is of great value where previously this level of detail would only be possible by an individual walking the whole site over a much longer period. The use of a regional model and UAV would eliminate the time-consuming need to collect ground testing data from every site, saving valuable time which may allow for more data collection. This is especially important in the Arctic, where vegetation is only present for a small portion of the year while snow is not present, meaning that efficiency of data capture is crucial as the system is changing quickly (Rantanen et al., 2022).

Classification of vegetation within peatlands will allow for the area of each vegetation class, and so methane flux over the area, to be easily calculated. Some classifiers often used in literature include Support Vector Machines (SVM; a classification using training data that lies closest to the boundary between classes in feature space to define where a hyperplane that separates classes lies) and Random Forest (RF; an ensemble learning method that employs multiple decision trees (DTs), hierarchical classifiers that split and classify data based on feature values into subtrees).

Knowing how much methane peatlands are emitting allows for accurate estimates of methane emissions from this crucial ecosystem and may give knowledge of which peatlands are the most impacted by climate change within the study region.

The aims of this project were to: i) assess the suitability of two machine learning (ML) models for classifying palsa vegetation types when transferred between regional sites, ii) compare two sampling types for training data across the two ML models, iii) ascertain whether there are any differences in vegetation within the region, and iv) identify which classes are the most and least successfully classified.

2. Literature review

Mapping of wetlands and peatlands finds its roots in mapping the extent of these ecosystems and their subtypes (Krankina et al., 2008; Peregon et al., 2009). A variety of methods have been used to classify wetlands by type, including semi-automated methods using statistical classifiers such as maximum-likelihood classification (MLC) (Bronge and Naslund-Landenmark, 2002). After the development of machine learning (ML) classifications, these classifiers were compared with statistical classifiers, as can be seen in a study by Sanchez-Hernandez et al. (2007), who compared support vector machines (SVM), an ML method, and MLC, finding SVM to be superior in locating coastal saltmarsh habitats.

However, classification within wetlands was not possible until methods with better spatial resolution were developed, such as the use of unoccupied aerial vehicles (UAVs). This allowed for characteristics within wetlands, such as vegetation type, to be mapped and enabled this data to be combined with other forms of data to study climatic effects on peatlands. A recent example of this can be seen from De la Barreda-Bautista et al. (2022) who combined vegetation maps from three palsa peatlands in Northern Sweden with InSAR data to study the subsidence of palsas on a variety of scales.

SVM is a popular classification method due to its high performance with little training data, and its application has been seen across a variety of studies to classify vegetation. For example, De la Barreda-Bautista et al.

(2022) classified palsa wetlands using SVM, getting overall accuracy values ranging from 70 to 82%. In addition, Sjögersten et al. (2023) used SVM to classify three palsa complexes in Northern Sweden, achieving a minimum accuracy of 70% in order to train a neural network to classify vegetation from satellite images. However, due to the UAV vegetation map not being the end product of this study, the classification may have the potential to reach higher overall accuracies, but this was not explored.

Random forest (RF) classification is another method often used for vegetation mapping and was found to be favourable for mapping vegetation communities in comparison to Convolutional Neural Network (CNN) classifiers by Bhatnagar et al. (2020). This was due to the need to retrain new CNN classifiers for different topologies, seasons, and atmospheric conditions, however, this study was based in Ireland, where the climate is vastly different to Northern Sweden. Another example of the use of RF classifiers can be seen in the study from Räsänen et al. (2019) from Northern Finland, who mapped vegetation communities within a wetland, comparing soft and hard RF classifications, reaching an overall accuracy of 72%. In this study, the use of textural features was found to be key in increasing overall accuracy.

Various other methods have also been used, such as the Bagging Trees (BT) classifier, used by Bhatnagar et al., (2021) to classify a raised bog in Ireland, achieving an overall accuracy of 87%. Another classification method comes through the use of multiresolution segmentation and an Object-Based Image Analysis, which has been used to reach overall classification accuracy of 86% when mapping dominant bog microforms for a site in South Patagonia, Argentina (Lehmann et al., 2016).

Classification models mapping vegetation type are often developed on a case-by-case basis and fitted to the specific site(s) being mapped (Bhatnagar et al., 2020; Bhatnagar et al., 2021; De la Barreda-Bautista et al., 2022; Lehmann et al., 2016; Räsänen et al., 2019; Sjögersten et al., 2023), and a regional classification method for palsa vegetation has not been attempted in the Arctic region. In fact, studies examining the transfer of models within and between regions are few and far between, but model transferability has been tested in a variety of environments for several different purposes. Foody et al. (2002) tested the transferability of a model to predict biomass across the moist zone tropical forest region, finding that when testing models on sites other than the one they were trained on, predictive accuracy declined. A more recent example comes from Venter et al. (2024), who tested the transferability of biodiversity models between two preserves in Northern South Africa, finding that both generalised linear models and generalized boosted models were consistently poor regardless of modelling approach. For vegetation classification, transferability of models has been assessed using RF to classify coastal vegetation (Juel et al., 2015). When testing transferability of RF models across the coastline of Denmark, Juel et al. (2015) achieved a maximum accuracy of 54.2%.

Despite these examples, transferability of vegetation classification ML models has not been assessed on palsa complexes in the region of Arctic Scandinavia, and the two popular ML classifications of SVM and RF have not been directly compared for this purpose.

3. Methodology

3.1 Study area

Six peatland sites were the focus of this study, located approximately 200km North of the Arctic Circle within Sweden and Finland (68N 21E). Three sites were located in close proximity to the border between the Kiruna municipality of Sweden and the Enontekiö municipality of Finland. Two sites were near the village of Nikkaluokta, in the Gällivare Municipality, and the final site was located near Abisko, within the Kiruna municipality in northwestern Lapland (figure 1).

The mean annual temperature (MAT) in Abisko from the period of 1990-2020 was 0.3°C and the mean annual precipitation (MAP) was 347.2mm, and for the same period, the mean monthly temperature (MMT) in July (when data was collected) was 12.2°C, while in 2024, the MMT rose to 14.6°C (http://www.smhi.se, last access: 13 August 2024). The MMT in Nikkaluokta in July for the 1991-2020 period was 12.7°C, while in July 2024 it was 14.7°C (http://www.smhi.se, last access: 13 August 2024). In Enontekiö, temperature and precipitation values are taken over a larger area, spanning the three study sites on the Sweden-Finland border. The MAT from 1981-2010 was < -1°C and the MMT in July in 2022 was 12.8°C, while the MAP 1981-2010 was 450-500mm (https://en.ilmatieteenlaitos.fi, last access: 7 August 2024). There is a difference of >1°C in MAT between the more Northern sites on the Sweden-Finland border and the Southern sites in Nikkaluokta, and this difference in temperature may impact vegetation composition on palsa complexes and palsa degradation.

The sites lie in the sub-Arctic climatic zone within the Arctic Circle, and as such experience 24 hours of sunlight during summer months. As all sites are within a zone of discontinuous or sporadic permafrost (Brown et al., 1997) this means that 50-90% (discontinuous) or 0-50% (sporadic) of the land surface is underlain by permafrost.

The sites consist of areas of raised palsa usually underlain with permafrost which degrade into areas of fen wetland which can be submerged beneath the water table, where no permafrost is present. Raised palsa areas tended to be drier, and as such the vegetation was dominated by dwarf shrubs, including *Betula nana*, and

Empetrum nigrum and some herbaceous species, *Rubus chamaemorus* in particular. In addition, numerous small areas were dominated by lichens and bryophytes (*Sphagnum* spp.). The flooded fen areas mainly tended to be vegetated by a mixture of bryophytes (*Sphagnum* spp.), sedges (*Carex* and *Eriophorum* species), deciduous shrubs (*Betula nana* and *Salix* species) and herbaceous species, mainly consisting of *Comarum palustre*. The degrading edges of palsa were vegetated by a mix of the wetter and drier species, with the drier species dominating the top of slopes, transitioning to the water edge, where the sedge and bryophyte species dominated.

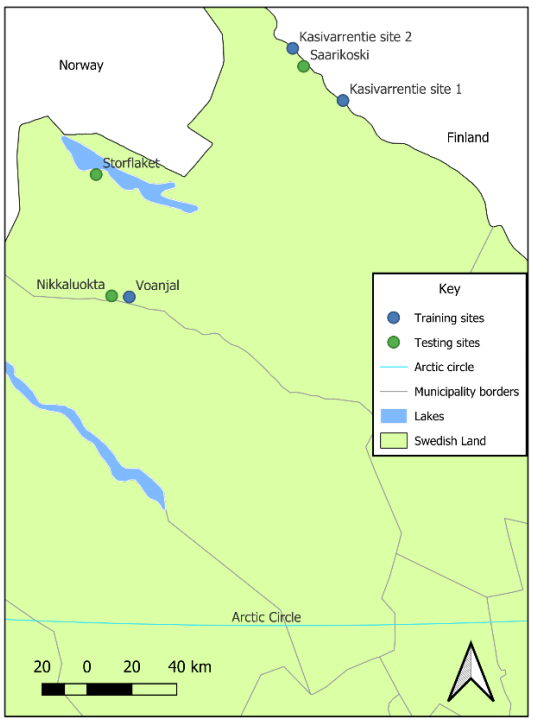

Figure 1. A map displaying the testing and training sites used in this project, showing their locations in comparison to each other and the Arctic Circle. Base map from www.lantmateriet.se under the Open Data License; Creative Commons, CC0.

3.2 Drone image collection

Images were collected from the study sites during July 2022 (Kasivarrentie sites 1 and 2, Saarikoski and Storflaket) and July 2024 (Nikkaloukta and Voanjal). A DJI Phantom 4 Multispectral unoccupied aerial vehicle (UAV) flew over each site taking images through six different sensors – one RGB sensor to image visible light and five monochrome sensors (Blue: 450 nm ± 16 nm; Green: 560 nm ± 16 nm; Red: 650 nm ± 16 nm; Red edge (RE): 730 nm ± 16 nm; Near-infrared (NIR): 840 nm ± 26 nm) to take multispectral images. The number of images collected at each site ranged from 1601 to 6551 (see table 1). At each site, 5-6 UAV targets were placed around the edge of the UAV path to be used as ground control points (GCPs) during processing. Radiometric calibration was not attempted at any of the sites.

3.3 Ground verification point collection

During the same period of fieldwork, ground verification points were collected using a differential GPS (dGPS; Trimble R8s; vertical error of 1.5 cm), where the vegetation class or abiotic class and location were recorded. A random sampling method was used to collect ground verification points from each of the sites. The number of ground verification points per class collected from each site (see table 2) varied depending on the accessibility and abundance of the classes, for example, some parts of the site were so heavily waterlogged they were inaccessible. The dGPS was also used to take a location from the centre of each of the UAV targets so the UAV images may be accurately geolocated.

3.4 Classes

Nine vegetation classes and two abiotic classes were used to characterise the vegetation found at the sites, based on the vegetation types used in Sjögersten et al. (2023) (figure 2). Three vegetation classes were assigned to vegetation on intact or degrading palsas, separating them into six classes depending on the state of the palsa, these being dry lichen (or dry lichen subsiding), dwarf shrub (or dwarf shrub subsiding) and moist moss (or moist moss subsiding). Three vegetation classes were found in areas where the palsa was entirely degraded and became partially flooded, and consisted of willow wetland, sedge/*Eriophorum* wetland (shortened to sedge wetland) and *Sphagnum* wetland. The two abiotic classes were bare ground and water. Detailed descriptions of the vegetation classes can be found in De la Barreda-Bautista et al. (2022).

Figure 2. Images from the six main vegetation types from De la Barreda-Bautista et al. (2022). Top row: vegetation types that can be found on intact and degrading palsas. Bottom row: fen vegetation types found in waterlogged wetland areas.

3.5 Reflectance map and orthoimage production

All images were manually assessed to remove any images with defects which would reduce the quality of the orthoimage to be produced (e.g. blurring, significant changes in lighting). Images were then imported into Pix4D mapper, where geometrically verified matching was used to estimate the orientation of the images and create automatic tie points. Ground control points (GCPs) were added at this stage, and manually located to ensure accurate geolocation of reflectance maps and digital surface model (DSM) (see figure 3). Varying numbers of images were calibrated and used in this step, meaning they were used in the final orthoimages. Next, a Dense Point Cloud was generated, from which a DSM was produced (figure 4). Reflectance maps were then produced for each spectral band and an orthoimage was produced using the RGB imagery.

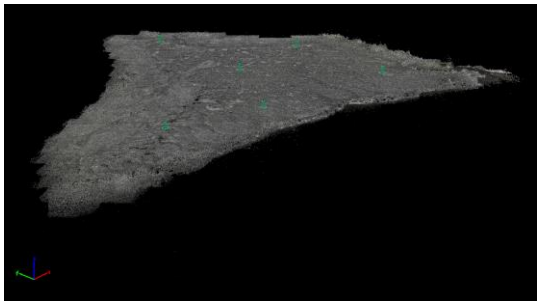

Figure 3. A processing image from Pix4D mapper of the automatic tie points (grey dots) and ground control points (green and blue circle and arrows) for one of the testing sites, Voanjal.

Figure 4. A processing image from Pix4D mapper of the Dense Point Cloud and ground control points (green and blue circle and arrows) for Voanjal.

3.6 Sample point selection

A number of terrain characteristics and vegetation indices were calculated in RStudio using the Terra package and added to a raster stack for each site consisting of the multispectral reflectance maps (red, green, blue, RE and NIR), vegetation indices (Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI) and Enhanced Vegetation Index (EVI)), DSM, and terrain characteristics (slope, Terrain Ruggedness Index (TRI), Topographic Position Index (TPI), flow direction and aspect). The DSM and terrain characteristic layers were resampled to 1m resolution using bilinear resampling to allow the terrain characteristics to capture the larger forms of the palsas rather than smaller variations within.

Three sites were selected to act as training sites so that a site in each location could be left untrained by the models to test the model's performance on sites it had not been trained on. The training sites are Kasivarrentie sites 1 and 2; and Voanjal, leaving Saarikoski, Storflaket and Nikkaluokta for testing. For each training site, a minimum of 15 polygons were constructed in QGIS containing each class from which sample points were then randomly selected in RStudio. It was ensured that the areas within the polygons did not include the ground verification points so that testing and training data did not overlap. Sample points were selected randomly from within the polygons.

Both a stratified and equally weighted training dataset was produced. The number of ground verification points per site for each class was used to weight the number of sample points collected from each site for the stratified sampling dataset, equalling 600 points for each class over all sites. For the equally weighted dataset, 200 points were sampled from each class within each site. This number was selected as it means that the number of training points is above the number recommended by Mather and Koch (2022) based on the number of spectral bands the classification uses. Only the equally weighted training data was used when the removal of a training site was tested to ensure that 400 training samples for each class remained from which to train the models.

3.7 Modelling

The caret package within RStudio was used to classify the vegetation using two different ML methods – RF, using the randomForest package; and SVM, using the kernlab and e1071 packages. Preprocessing was used to centre and scale the SVM models, and a grid search was used to select the optimal hyperparameters (C and sigma) for this model. The same method was used to select the optimal hyperparameter for the RF model (mtry).

One SVM and RF model was trained using data from all training sites for both the equally weighted and stratified sampling data. In addition, the equally weighted data was split by site and used to train an SVM and RF model on data from two sites in every combination.

3.8 Assessing accuracy

The accuracy of the two models was assessed at each site by getting the models to predict the class of the ground verification points and comparing the prediction to the class recorded at the location. Overall

Site name	Date of image capture	Total number of images	Number of calibrated images	Average Ground Sampling Distance (GSD) (cm)	Number of Ground Control Points (GCPs)
Kasivarrentie site 1	27/07/2022	1602	1601	4.31	5
Kasivarrentie site 2	28/07/2022	1614	1614	4.22	5
Nikkaluokta	12/07/2024	4895	4895	4.3	6
Saarikoski	26/07/2022	2737	2725	6.73	5
Storflaket	19/07/2023	4370	4356	4.15	5
Voanjal	13/07/2024	6552	6551	4.32	6

Table 1. A table displaying the details of the orthoimage production for each site.

Site name	Number of ground verification points per class											Total
	Dry lichen	Dry lichen subsiding	Dwarf shrub	Dwarf shrub subsiding	Moist moss	Moist moss subsiding	Sedge wetland	Sphagnum wetland	Willow wetland	Bare ground	Water	
Kasivarrentie site 1	5	9	16	15	1	11	6	19	6	8	0	96
Kasivarrentie site 2	17	6	29	8	20	12	15	14	0	0	0	121
Nikkaluokta	23	14	36	44	27	13	15	35	30	19	14	270
Saarikoski	13	10	16	34	22	20	47	15	43	0	0	220
Storflaket	33	32	47	55	37	37	14	42	42	0	15	354
Voanjal	26	22	26	37	13	13	42	26	29	23	21	278
Total	117	93	170	193	120	106	139	151	150	50	50	1339

Table 2. A table displaying the number of GVPs collected at each site split by class.

accuracy was then calculated in RStudio for the model as a whole and at each site. Confusion matrices were used to assess the class-wise predictions of each model for all sites combined and for each site separately.

Due to the similarity in overall accuracy between the models, z-scores were calculated to test whether there was a significant difference in the accuracies between the models.

4. Results

4.1 Overall accuracy

Overall accuracy across all sites was highest when predicted by the SVM model with equally weighted sampling (52.5%). This model was significantly more accurate than the SVM model with stratified sampling (48.5%) with a z-score of 4.37, above the threshold for 99% confidence of a significant difference in predictive accuracy between the two models. For the RF models, no significant difference was found when comparing overall accuracy across all sites for the equally weighted (51.5%) and stratified (51.9%) sampling methods (z-score = 0.23).

Comparing between the two different models (SVM and RF) resulted in a z-score of 0.73 when comparing between the models with equally weighted sampling, indicating no significant difference in predictive accuracy between the two different models.

The overall accuracies of each site for the models are displayed in table 3, and it is evident that Kasivarrentie site 1 and Voanjal have the highest overall accuracies in every model, which was anticipated as the training data was collected from these sites. Voanjal had a smaller range of overall accuracies, all >80%, while Kasivarrentie ranged from 60% to 89.4%, with greater overall accuracies from the models with equally weighted sampling, as this site had a smaller number of ground verification points, meaning that less training data was taken from this site for the stratified sampling training dataset.

The overall accuracy of Kasivarrentie site 2 ranged from 33.9% to 43.8% (Table 3), despite one third of the training data originating from this site. When models were trained with data from only Kasivarrentie site 1 and Voanjal, the overall accuracy of the models decreased to 49.2% for the SVM model, a decrease in accuracy of 3.3%, and 51.2% for the RF model, showing a smaller decrease of 0.3%. When the same models were trained without training data from either of the other two training sites, the overall accuracy decreased by a larger degree, ranging from 42.3% to 44.1%.

The overall accuracy of all the testing sites combined followed a similar pattern to the overall accuracy combining all sites. The highest overall accuracy for the testing sites was from the SVM method with equally weighted sampling, achieving 41.9% overall accuracy, followed by the RF model with stratified sampling, with 40.2% overall accuracy. The RF model with equally weighted sampling and the SVM model with stratified sampling had the lowest overall accuracies, getting 38.4% and 37.4% respectively. Z-scores were used to see if the models performed significantly differently. Comparing the two different sampling methods for RF saw no significant difference (z-score = 1.52) as the z-score was below the 95% confidence threshold. Similarly, there was no significant difference between the two models with the highest accuracy (z-score = 1.07). There was 95% confidence that the SVM model with equally weighted sampling had significantly better predictive accuracy than the RF model with stratified sampling (z-score = 2.11), and there was 99% confidence that the performance of the SVM model with equally weighted sampling was significantly better than the SVM model with stratified sampling (z-score = 3.97).
Of the testing sites, Nikkaluokta had the highest overall accuracy, averaging approximately 50% (Table 3), and had the highest overall accuracy when predicted by the SVM model with equally weighted sampling, and the lowest overall accuracy from the SVM model with stratified sampling. Saarikoski had a greater range of overall accuracy values than Nikkaluokta, ranging from 35.9% to 45.5%. The model with the lowest overall accuracy for Saarikoski was the equally weighted RF model, while the highest overall accuracy for this site was reached using the equally weighted SVM model. The testing site with the lowest overall accuracy was Storflaket, ranging from 23.9% to 31.9% with the same pattern as Saarikoski.

When the models were trained with data from only two sites, the absence of one site from the training data led to significant decreases in the overall accuracies of the two testing sites in close proximity to one or more training site (Saarikoski and Nikkaluokta). For Saarikoski, the decrease in overall accuracy of this site was seen with the removal of data from Kasivarrentie site 2, as the models trained with the use of data from this site ranged from 37.3% to 45.9%, and the models without data from this site had overall accuracies of 28.2% (RF) and 31.4% (SVM). For Nikkaluokta, the removal of training data from Voanjal resulted in a decrease of overall accuracy from values ranging from 50.9% to 53.6% down to 39.6% (RF) and 38.9% (SVM). Overall accuracy of Storflaket showed no trends across both classifiers due to the removal of training data from a single site.

4.2 Class-wise accuracy

When looking at the differences in classification of the different classes, there is a clear difference in sensitivity between the biotic and abiotic classes. Sensitivity of the bare ground class ranged from 73.7% to 84.2%, and for water, ranged from 86.2% to 93.1%

Model	Sampling technique	Overall accuracy (%)						
		All sites	Kasivarrentie site 1	Kasivarrentie site 2	Saarikoski	Storflaket	Nikkaluokta	Voanjal
SVM	Stratified	48.4	60.0	33.9	41.4	24.9	49.4	83.2
SVM	Equally weighted	52.5	83.5	33.9	45.5	31.9	51.3	81.7
RF	Stratified	51.9	70.6	43.8	40.9	27.6	50.2	85.0
RF	Equally weighted	51.5	89.4	43.8	35.9	23.9	50.2	82.1

Table 3 – a table displaying the overall accuracy of each classifier at all sites and by individual site, grouped by location. Kasivarrenite sites 1 and 2 and Saarikoski are the Northernmost sites, with both Kasivarrentie sites being used for testing. Nikkaluokta and Voanjal are the Southernmost sites, with Voanjal used for training. Storflaket is located between the Northern and Southern sites and was used for testing only.

sensitivity. Meanwhile, the highest sensitivity was achieved in the dry lichen and sedge wetland classes, with dry lichen sensitivity ranging from 42% to 68.1%, and sedge wetland from 56.6% to 71.1%.

The lowest sensitivity was for the willow wetland and moist moss subsiding classes, with sensitivity ranging from 12.3% to 16.8%. Willow wetland was most often misclassified as sphagnum wetland, with mean classification as sphagnum wetland at 41.3% across all models, however it was also commonly misclassified as dwarf shrub and dwarf shrub subsiding, while moist moss subsiding was not misclassified consistently as another class.

5. Discussion

The poor overall accuracies of all models (<45%) when predicting testing sites only, indicate that the ML models trained for this project are not suitable for classifying vegetation types of palsa complexes across a region. Similar studies have achieved similar poor results; Juel et al. (2015) conducted a similar study classifying coastal vegetation in Denmark, obtaining a maximum overall accuracy of 56.6%, and Eischeid et al. (2021) only commented that direct transfer of models was not possible when mapping disturbance in Arctic tundra sites due to macro-F1 scores <50%.

A possible reason for the lower overall accuracy on testing sites found in this study compared to Juel et al. (2015) may be the low number of training sites (3) compared to the 53 sites used for training by Juel et al. (2015). A larger number of training sites would allow for greater capture of regional variation within classes, as the removal of one class at a time from the training data showed that the training sites closest to the testing sites led to the largest decrease in overall accuracy of the testing site. The overall accuracy of Saarikoski decreased by a minimum of 5.9% when data from Kasivarrentie site 2 was removed from the training dataset, and the overall accuracy of Nikkaluokta decreased by a minimum of 11.3% when data from Voanjal was removed from the training data. This may be due to the reduced distance between these sites (approx. 8km between Nikkaluokta and Voanjal; approx. 9km between Kasivarrentie site 2 and Saarikoski). This means that there may be extremely localised trends in vegetation, as a similar decrease was not seen in overall accuracy from Saarikoski when training data from Kasivarrenite site 1 was not included in model training despite the sites being approximately 23km apart and having similar surroundings (i.e. all located in close proximity to well-travelled roads, etc.). This trend indicates that these sites follow Tobler's first Law, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). This is further confirmed as Storflaket did not see a significant decrease across both models with the removal of training data from any single site, and it was >55km from all training sites.

In addition, Storflaket had consistently lower overall accuracy, which may have been due to the large distance separating this site to all training sites. The closest training site to Storflaket was Voanjal, approximately 56km away, while the other training sites were both >100km away, compared to the other testing sites which had a training site <10km away.

Another possible limitation of the training data is the amount and variation of training data collected from each site. Kasivarrentie site 2, despite being a training site, had low overall accuracy (33.9 – 43.8%), but removal of this site from the training data resulted in a decrease in the overall accuracy of Saarikoski by a minimum of 5.9%. This suggests that the training data captured some of the variation of vegetation in both of these sites, but the spread of training data across Kasivarrentie site 2 did not capture all of the variation in vegetation classes within this site. The collection of more data from each of the training sites may allow for more accurate classification, as the models in this study were trained on 6,600 training samples altogether, while the more accurate models in the study from Juel et al. (2015) were trained on 21,000 training samples.

The poor performance of the models may be due to the data itself. Although multispectral data provides a greater number of spectral measurements than RGB imaging, the five broad bands provided by the sensor used in this project may not allow for enough differentiation between vegetation types. A possible solution for this may be the use of hyperspectral data, providing many additional narrow spectral band measurements, meaning that differences between vegetation classes may be captured to be used for classification. Hyperspectral data has been seen to be superior to multispectral data when modelling for a variety of purposes, including for vegetation classification (Lee et al., 2004; Marshall and Thenkabail, 2015; Sluiter and Pebesma, 2010).

In addition, images from UAVs often have different spectral reflectance values across separate sites, due to differences in weather, season, etc., meaning that radiometric calibration is often used to compensate for these differences. As radiometric calibration was not used when the UAV was operated, differences in the reflectance between the sites could be posited as a possible reason for the low overall accuracy achieved on the testing sites. However, Cubero-Castan et al. (2018) investigated the radiometric accuracy that could be achieved without radiometric calibration and found that using Pix4D mapper allowed for the calculation of accurate reflectance values. The weather conditions were highly different during UAV data capture for the Nikkaluokta and Voanjal sites (i.e. overcast weather conditions on Nikkaluokta and sunny conditions on Voanjal), however, the removal of training data from Voanjal resulted in a significant decrease in accuracy for Nikkaluokta, suggesting that the lack of radiometric calibration did not hamper the classification models.

The difference in sensitivity between the biotic and abiotic classes may be due to the uniformity of the abiotic classes. This means that there is little variability in these classes, meaning that the training data used in this project was suitable to classify the water and bare ground classes to a sensitivity >70% for bare ground and >85% for water.

Differences in sensitivity between biotic classes may be due to the similarity in appearance between classes, for example, the three fen classes (sedge wetland, sphagnum wetland and willow wetland) share some species. The willow wetland class was any waterlogged area with the presence of a *Salix* species, meaning that other species present periodically included *Sphagnum* species, also found in the sphagnum wetland class, and *Betula nana* also found in the dwarf shrub and dwarf shrub subsiding classes, and these classes were those that the willow wetland class was most misclassified as. These misclassifications may be a result of both models (RF and SVM) using point data as input, meaning that the combination of multiple species would be missed by the classifier. A model more suited for the identification of classes consisting of a mixture of species may be a deep learning model such as a convolutional neural network (CNN), as the input data for this model can consist of sections of images (Rawat and Wang, 2017), allowing for the model to learn the different reflectance values of the combination of species characteristic of these classes. In comparison, the two modelling methods implemented in this study required input of point data which would encompass only one vegetation species. In addition, neural networks have proven to be superior for vegetation classification, even when trained on point data (Berberoglu et al., 2000).

6. Conclusion

In conclusion, with the small number of sites from which data was collected, the ML methods which were assessed did not prove accurate enough to be applied region wide. The use of more training sites and the collection of more training data from each site were considered as methods to increase accuracy of ML classification of palsas. However, the use of hyperspectral data to better capture differences between vegetation types due to the larger number of spectral bands, and the use of a deep learning model such as a CNN, which may better capture the mixture of species that make up each of the vegetation classes due to input of these models which can be greater than one pixel, may lead to more accurate vegetation classifications.

References

Åkerman, H.J. and Johansson, M., 2008. Thawing permafrost and thicker active layers in sub-arctic Sweden. *Permafrost and periglacial processes*, *19*(3), pp.279-292.

Assmann, J.J., Kerby, J.T., Cunliffe, A.M. and Myers-Smith, I.H., 2018. Vegetation monitoring using multispectral sensors—best practices and lessons learned from high latitudes. *Journal of Unmanned Vehicle Systems*, *7*(1), pp.54-75.

Ballantyne, C.K., 2018. *Periglacial geomorphology*. John Wiley & Sons.

Berberoglu, S., Lloyd, C.D., Atkinson, P.M. and Curran, P.J., 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. *Computers & Geosciences*, *26*(4), pp.385-396.

Bhatnagar, S., Gill, L. and Ghosh, B., 2020. Drone image segmentation using machine and deep learning for mapping raised bog vegetation communities. *Remote Sensing*, *12*(16), p.2602.

Bhatnagar, S., Gill, L., Regan, S., Waldren, S. and Ghosh, B., 2021. A nested drone-satellite approach to monitoring the ecological conditions of wetlands. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*, pp.151-165.

Biskaborn, B.K., Smith, S.L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D.A., Schoeneich, P., Romanovsky, V.E., Lewkowicz, A.G., Abramov, A., Allard, M., Boike, J., Cable, W.L., Christiansen, H.H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Guglielmin, M., Ingeman-Nielsen, T., Isaksen, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kröger, T., Lambiel, C., Lanckman, J.P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A.B.K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M. and Lantuit, H., 2019. Permafrost is warming at a global scale. *Nature communications*, *10*(1), p.264.

Borge, A.F., Westermann, S., Solheim, I. and Etzelmüller, B., 2017. Strong degradation of palsas and peat plateaus in northern Norway during the last 60 years. *The Cryosphere*, *11*(1), pp.1-16.

Bronge, L.B. and Näslund-Landenmark, B., 2002. Wetland classification for Swedish CORINE Land Cover adopting a semi-automatic interactive approach. *Canadian journal of remote sensing*, 28(2), pp.139-155.

Brown, J., Sidlauskas, F.J. and Delinski, G.F., 1997. Circum-arctic map of permafrost and ground ice conditions.

De la Barreda-Bautista, B., Boyd, D.S., Ledger, M., Siewert, M.B., Chandler, C., Bradley, A.V., Gee, D., Large, D.J., Olofsson, J., Sowter, A. and Sjögersten, S., 2022. Towards a Monitoring Approach for Understanding Permafrost Degradation and Linked Subsidence in Arctic Peatlands. *Remote Sensing*, 14(3), p.444.

Desrochers, D.T. and Granberg, H.B., 1988, August. Schefferville snow-ground interface temperatures. In *Fifth International Conference on Permafrost, Trondheim, Norway. Tapir, Trondheim* (Vol. 1, pp. 27-34).

Eischeid, I., Soininen, E.M., Assmann, J.J., Ims, R.A., Madsen, J., Pedersen, Å.Ø., Pirotti, F., Yoccoz, N.G. and Ravolainen, V.T., 2021. Disturbance mapping in Arctic tundra improved by a planning workflow for drone studies: Advancing tools for future ecosystem monitoring. *Remote Sensing*, *13*(21), p.4466.

Feng, L., Palmer, P.I., Parker, R.J., Lunt, M.F. and Bösch, H., 2023. Methane emissions are predominantly responsible for record-breaking atmospheric methane growth rates in 2020 and 2021. *Atmospheric Chemistry and Physics*, *23*(8), pp.4863-4880.

Foody, G.M., Boyd, D.S. and Cutler, M.E., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote sensing of environment*, *85*(4), pp.463-474.

Fraser, R.H., Olthof, I., Lantz, T.C. and Schmitt, C., 2016. UAV photogrammetry for mapping vegetation in the low-Arctic. *Arctic Science*, *2*(3), pp.79-102.

Glagolev, M., Kleptsova, I., Filippov, I., Maksyutov, S. and Machida, T., 2011. Regional methane emission from West Siberia mire landscapes. *Environmental Research Letters*, *6*(4), p.045214.

Hodgkins, S.B., Tfaily, M.M., McCalley, C.K., Logan, T.A., Crill, P.M., Saleska, S.R., Rich, V.I. and Chanton, J.P., 2014. Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proceedings of the National Academy of Sciences*, *111*(16), pp.5819-5824.

Hugelius, G., Loisel, J., Chadburn, S., Jackson, R.B., Jones, M., MacDonald, G., Marushchak, M., Olefeldt, D., Packalen, M., Siewert, M.B. and Treat, C., 2020. Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proceedings of the National Academy of Sciences*, *117*(34), pp.20438-20446.

IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2391 pp. doi:10.1017/9781009157896.

Juel, A., Groom, G.B., Svenning, J.C. and Ejrnaes, R., 2015. Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, *42*, pp.106-114.

Krankina, O.N., Pflugmacher, D., Friedl, M., Cohen, W.B., Nelson, P. and Baccini, A., 2008. Meeting the challenge of mapping peatlands with remotely sensed data. *Biogeosciences*, *5*(6), pp.1809-1820.

Köchy, M., Hiederer, R. and Freibauer, A., 2015. Global distribution of soil organic carbon–Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. *Soil*, *1*(1), pp.351-365.

Koh, L.P. and Wich, S.A., 2012. Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation. *Tropical conservation science*, *5*(2), pp.121-132.

Lee, K.S., Cohen, W.B., Kennedy, R.E., Maiersperger, T.K. and Gower, S.T., 2004. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment*, *91*(3-4), pp.508-520.

Lehmann, J.R., Münchberger, W., Knoth, C., Blodau, C., Nieberding, F., Prinz, T., Pancotto, V.A. and Kleinebecker, T., 2016. High-resolution classification of south patagonian peat bog microforms reveals potential gaps in up-scaled CH4 fluxes by use of Unmanned Aerial System (UAS) and CIR imagery. *Remote Sensing*, *8*(3), p.173.

Luoto, M. and Seppälä, M., 2003. Thermokarst ponds as indicators of the former distribution of palsas in Finnish Lapland. *Permafrost and periglacial processes*, *14*(1), pp.19-27.

Malenovský, Z., Lucieer, A., King, D.H., Turnbull, J.D. and Robinson, S.A., 2017. Unmanned aircraft system advances health mapping of fragile polar vegetation. *Methods in Ecology and Evolution*, *8*(12), pp.1842-1857.

Marshall, M. and Thenkabail, P., 2015. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, pp.205-218.

Mather, P.M. and Koch, M., 2022. Computer processing of remotely-sensed images. Fifth edition. Wiley Blackwell.

McLaughlin, S., Zhalnina, K., Kosina, S., Northen, T.R. and Sasse, J., 2023. The core metabolome and root exudation dynamics of three phylogenetically distinct plant species. *Nature communications*, *14*(1), p.1649.

Miglovets, M.N., Zagirova, S.V., Goncharova, N.N. and Mikhailov, O.A., 2021. Methane emission from palsa mires in northeastern European Russia. *Russian Meteorology and Hydrology*, *46*, pp.52-59.

Mora, C., Vieira, G., Pina, P., Lousada, M. and Christiansen, H.H., 2015. Land cover classification using high-resolution aerial photography in adventdalen, Svalbard. *Geografiska Annaler: Series A, Physical Geography*, *97*(3), pp.473-488.

Olvmo, M., Holmer, B., Thorsson, S., Reese, H. and Lindberg, F., 2020. Sub-arctic palsa degradation and the role of climatic drivers in the largest coherent palsa mire complex in Sweden (Vissátvuopmi), 1955–2016. *Scientific reports*, *10*(1), p.8937.

Peregon, A., Maksyutov, S. and Yamagata, Y., 2009. An image-based inventory of the spatial structure of West Siberian wetlands. *Environmental Research Letters*, *4*(4), p.045014.

Rantanen, M., Karpechko, A.Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T. and Laaksonen, A., 2022. The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, 3(1), p.168.

Räsänen, A., Juutinen, S., Tuittila, E.S., Aurela, M. and Virtanen, T., 2019. Comparing ultra-high spatial resolution remote-sensing methods in mapping peatland vegetation. *Journal of Vegetation Science*, *30*(5), pp.1016-1026.

Rawat, W. and Wang, Z., 2017. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, *29*(9), pp.2352-2449.

Rouse, W.R., 1984. Microclimate of arctic tree line 2. Soil microclimate of tundra and forest. *Water Resources Research*, *20*(1), pp.67-73.

Sanchez-Hernandez, C., Boyd, D.S. and Foody, G.M., 2007: Mapping specific habitats from remotely sensed imagery: support vector machine and support vector data description based classification of coastal saltmarsh habitats. *Ecological informatics*, *2*(2), pp.83-88.

Sannel, A.B.K., Hugelius, G., Jansson, P. and Kuhry, P., 2016. Permafrost warming in a subarctic peatland–which meteorological controls are most important?. *Permafrost and periglacial processes*, *27*(2), pp.177-188.

Sannel, A.B.K. and Kuhry, P., 2011. Warming-induced destabilization of peat plateau/thermokarst lake complexes. *Journal of Geophysical Research: Biogeosciences*, *116*(G3).

Siewert, M.B. and Olofsson, J., 2020. Scale-dependency of Arctic ecosystem properties revealed by UAV. *Environmental Research Letters*, *15*(9), p.094030.

Sjöberg, Y., Marklund, P., Pettersson, R. and Lyon, S.W., 2015. Geophysical mapping of palsa peatland permafrost. *The Cryosphere*, *9*(2), pp.465-478.

Sjögersten, S., Ledger, M., Siewert, M., de la Barreda-Bautista, B., Sowter, A., Gee, D., Foody, G. and Boyd, D.S., 2023. Optical and radar Earth observation data for upscaling methane emissions linked to permafrost degradation in sub-Arctic peatlands in northern Sweden. *Biogeosciences,* 20(20), pp.4221-4239.

Sluiter, R. and Pebesma, E.J., 2010. Comparing techniques for vegetation classification using multi-and hyperspectral images and ancillary environmental data. *International Journal of Remote Sensing*, *31*(23), pp.6143-6161.

Ström, L., Mastepanov, M. and Christensen, T.R., 2005. Species-specific effects of vascular plants on carbon turnover and methane emissions from wetlands. *Biogeochemistry*, *75*, pp.65-82.

Turner, A.J., Frankenberg, C. and Kort, E.A., 2019. Interpreting contemporary trends in atmospheric methane. *Proceedings of the National Academy of Sciences*, *116*(8), pp.2805-2813.

Varner, R.K., Crill, P.M., Frolking, S., McCalley, C.K., Burke, S.A., Chanton, J.P., Holmes, M.E., Isogenie Project Coordinators, Saleska, S. and Palace, M.W., 2022. Permafrost thaw driven changes in hydrology and vegetation cover increase trace gas emissions and climate forcing in Stordalen Mire from 1970 to 2014. *Philosophical Transactions of the Royal Society A*, *380*(2215), p.20210022.

Venter, T.S., Barker, N.P. and le Roux, P.C., 2024. Poor transferability of richness models: Predicting plot-scale plant diversity in the Waterberg, South Africa. *South African Journal of Botany*, *171*, pp.228-236.

Walter, K.M., Zimov, S.A., Chanton, J.P., Verbyla, D. and Chapin III, F.S., 2006. Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming. *Nature*, *443*(7107), pp.71-75.

Walter Anthony, K., Daanen, R., Anthony, P., Schneider von Deimling, T., Ping, C.L., Chanton, J.P. and Grosse, G., 2016. Methane emissions proportional to permafrost carbon thawed in Arctic lakes since the 1950s. *Nature Geoscience*, *9*(9), pp.679-682.

Zhang, Z., Poulter, B., Feldman, A.F., Ying, Q., Ciais, P., Peng, S. and Li, X., 2023. Recent intensification of wetland methane feedback. *Nature Climate Change*, *13*(5), pp.430-433.

Zuidhoff, F.S. and Kolstrup, E., 2000. Changes in palsa distribution in relation to climate change in Laivadalen, northern Sweden, especially 1960–1997. *Permafrost and periglacial processes*, *11*(1), pp.55-69.