Charles univesity

Faculty of science

Department of Applied Geoinformatics and Carthography

Study program: Geography

Study field: Geoinformatics, carthography and remote sensing

Bc. Adam Kulich

ADVANCED REMOTE SENSING METHODS FOR MONITORING OF PEAT BOG VEGETATION IN THE KRKONOŠE MOUNTAINS

POKROČILÉ METODY DÁLKOVÉHO PRŮZKUMU ZEMĚ PRO MONITORING VEGETACE RAŠELINIŠŤ V KRKONOŠÍCH

Master thesis

Thesis supervisor: doc. RNDr. Lucie Kupková, Ph.D.

Prague, 2024

Statement

I declare that I worked on my diploma thesis independently and that I listed all the literature and information sources used. This work or a substantial part of it has not been submitted for the award of another or the same academic degree.

In Prague on April 29, 2024

Bc. Adam Kulich

Acknowlegment

I would like to thank everyone who provided me with materials for the preparation of the thesis, expert consultation, or other support during the work. I especially thank the supervisor of my thesis, Assoc. Prof. RNDr. Lucie Kupková, Ph.D., for her time and expert guidance, her patience and responsiveness during consultations, and useful advice for the creation of the thesis.

I would also like to thank RNDr. Jakub Lysák, Ph.D., and Mgr. Lucie Červená, Ph.D., for their help with data collection and advice on the thesis. I also thank Mgr. Viera Horáková from the management of Krkonoše National Park for her help with data collection, consultation on classification outputs, and other support.

Last but not least, I would also like to thank my fiancée and the rest of my family for their patience and continuous unconditional support.

Katedra aplikované geoinformatiky a kartografie

Přírodovědecká fakulta Univerzity Karlovy

Zadání diplomové práce

Jméno a příjmení: Adam Kulich

Studijní program: Geoinformatika, kartografie a dálkový průzkum Země

Název: Pokročilé metody dálkového průzkumu země pro monitoring vegetace rašelinišť v Krkonoších

Zásady pro vypracování

Téma DP: Diplomová práce se zabývá metodami zpracování multispektrálních obrazových dat z krkonošských rašelinišť a jejich dlouhodobým sledováním. Navazuje na výzkum provedený v autorově bakalářské práci "Monitoring vegetace rašelinišť v Krkonoších s využitím DPZ". Práceprobíhá ve spolupráci s fakultním Týmem laboratorní a obrazové spektroskopie a Mgr. Vierou Horákovou – botaničkou Správy KRNAP a je součástí dlouhodobého projektu KRNAP s cílem revitalizace rašelinišť.

Cíle DP: Vytvořit co nejpřesnější mapu vegetačního pokryvu tří krkonošských rašelinišť (Hraniční louka, Kyselé kouty a Pančavská louka) a navrhnout postup pro další efektivní mapování rašelinišť v Krkonoších ve spolupráci s KRNAP.

Základní popis použitých dat a metod s uvedením výzkumného úkolu v oblasti geoinformatiky: Použita budou obrazová data z bezpilotního letounu DJI Phantom Multispectral, pořízená autorem a Týmem obrazové a laboratorní spektroskopie v různých termínech, dále terénní data nasbíraná tímtéž týmem ve spolupráci s botaničkou Správy KRNAP. Ke zpracování dat budou využity různé metody strojového učení. Bude testován též vliv nespektrálních příznaků jako výška porostu nebo textura obrazu a dále také množství terénních dat na výsledky klasifikace. Hlavním výzkumným úkolem bude zjistit spolehlivost těchto metod pro sledování změn vegetačního pokryvu rašelinišť a standardizovat tento postup pro další práci v KRNAP.

Předpokládané výstupy práce: Konkrétním výstupem práce budou mapy vegetačního pokryvu, hodnocení přesnosti každé z map a také odhad změny vegetačního pokryvu v čase. Dosažená přesnost klasifikací by měla být srovnatelná s výstupem zmíněné bakalářské práce (cca 90 %). Z výzkumného hlediska bude výstupem také doporučení, jaké metody je nejvhodnější použít při mapování/analýze vegetace dalších rašelinišť v KRNAP.

Seznam odborné literatury:

Kupková, L., Červená, L., Potůčková, M., Lysák, J., Roubalová, M., Hrázský, Z., Březina, S., Epstein, H. E., & Müllerová, J. (2023). Towards reliable monitoring of grass species in nature conservation: Evaluation of the potential of UAV and PlanetScope multi-temporal data in the Central European tundra. Remote Sensing of Environment, 294.

Huylenbroeck, L. et al. (2020): Using remote sensing to characterize riparian vegetation: A review of available tools and perspectives for managers, Journal of Environmental Management, 267

Abeysinghe, T. et al. (2019): Mapping Invasive Phragmites australis in the Old Woman Creek Estuary Using UAV Remote Sensing and Machine Learning Classifiers, Remote Sensing, 11(11), 1380

Vedoucí diplomocé práce: doc. RNDr. Lucie Kupková, PhD.

Konzultant diplomové práce: Mgr. Viera Horáková

Datum zadání diplomové práce: 31.8.2023

doc. RNDr. Přemysl Štych, Ph.D.

....

Vedoucí diplomové práce Garant studijního programu

V Praze dne 31.8.2023

Advanced remote sensing methods for monitoring of peat bog vegetation in the Krkonoše Mountains

Abstract

The goal of the thesis was to create maps of vegetation cover for three peat bogs in Krkonoše National Park (KRNAP, Krkonoše NP) and to propose a method for further efficient mapping of peat bogs in the Krkonoše Mountains in collaboration with Krkonoše NP. The research utilized UAV image data and field botanical data measured by GPS. UAV images were preprocessed using Pix4D Mapper software and features such as canopy height and textures derived from GLCM were added to the resulting orthomosaics. Testing confirmed their usefulness in increasing classification accuracy. A separability analysis was performed, and an algorithm was designed to detect errors in the field data. During the analysis, two classification methods, Random Forest and Support Vector Machine (SVM), were compared. The SVM method achieved the most precise results at Kyselé kouty, where average F-1 score reached 0.957, while the F-1 scores for Hraniční louka and Pančavská louka reached only 0.899 and 0.832 respectively. The overlay analysis demonstrated that the results of the individual methods are consistent, and for Pančavská louka, combining classifiers yielded better accuracy than the individual models. The methods which led to the best accuracies in this thesis were recommended in the proposed approach for further mapping of Krkonoše peat bogs.

Pokročilé metody dálkového průzkumu země pro monitoring vegetace rašelinišť v Krkonoších

Abstrakt

Cílem práce bylo vytvořit mapy vegetačního pokryvu tří rašelinišť v Krkonošském Národním parku (KRNAP) a navrhnout postup pro další efektivní mapování rašelinišť v Krkonoších ve spolupráci s KRNAP. Pro výzkum byla využita obrazová data z UAV a terénní botanická data, zaměřeá GPS. Snímky z UAV byly předzpracovány v softwaru Pix4D Mapper a do výsledných ortofot byly přidány příznaky výšky porostu a texury, odvozené z GLCM. Testování potvrdilo jejich užitečnost pro zlepšení přesnosti klasifikací. Byla provedena analýza separability a byl navžen algoritmus, rozpoznávající chyby v terénních datech. Během analýzy byly porovnávány dvě klasifikační metody - Random Forest a Support vector machine (SVM). Nejpřesnějších výsledků dosáhla metoda SVM na Kyselých koutech, kde bylo dosaženo F-1 skóre 0.957, zatímco F-1 skóre Hraniční louky dosáhlo pouze 0.899 a Pančavské louky 0.832. Analýza překryvů ukázala, že výsledky jednotlivých metod jsou konzistentní a v případě Pančavské louky přineslo spojení klasifikátorů I lepší přesnost, než jednotlivé modely. Metody, které vedly k nejlepší přesnosti, byly doporučeny v navrženém postupu pro další krkonošská rašeliniště.

Contents

1 Introduction **9**

1.1 Objectives . . . 9

1.2 Theoretical background . . . 10

1.2.1 Use of non-spectral features in remote sensing . . . 10

1.2.2 Use of UAV remote sensing in wetland species mapping . . . 11

1.2.3 Use of remote sensing in vegetation mapping in Krkonoše NP . . . 12

2 Areas of interest **14**

2.1 Peat bogs in Krkonoše NP . . . 14

2.2 Hraniční louka (Border meadow) . . . 15

2.3 Kyselé kouty (Sour corners) . . . 16

2.4 Pančavské louka (Pančava Meadow) . . . 17

3 Data and methods **18**

3.1 Data collection and classification legend . . . 18

3.2 Image data preprocessing . . . 18

3.3 Ground truth data processing . . . 19

3.4 Canopy height feature creation . . . 21

3.5 Texture feature creation . . . 22

3.6 Ground truth data separability evaluation and cleaning . . . 23

3.7 Reduction of the number of features . . . 25

3.8 Selection and application of a suitable classifier . . . 25

3.9 Testing the classifier's behavior when reducing the number of ground truth data . . . 28

3.10 Creation of assembled vegetation cover maps . . . 28

4 Results **29**

4.1 Classification results . . . 29

4.2 Evaluation of usefulness of the non-spectral features . . . 29

4.3 Separability analysis . . . 32

4.4 Hyperparameter tuning . . . 34

4.5 Results of classifications after data reduction . . . 37

4.6 Overlays and final vegetation cover maps . . . 41

5 Discussion **45**

6 Conclusion **55**

7 References **56**

A Images of classified species **62**

B Classification results for all classes **68**

C Separability tables **75**

D Best maps of vegetation cover for each date and area of interest **81**

List of Figures

1 Orthofotomap of Krkonoše NP with the three areas of interest 14
2 Hraniční louka, 13 July 2023 15
3 Kyselé kouty, 13 July 2023 16
4 Pančavská louka, 14 July 2023 17
5 Illustration of non-spectral features 23
6 Example of the incorrectly labeled polygon on Pančavská louka. 33
7 Simple data reduction on Hraniční louka 37
8 Simple data reduction on Kyselé kouty 38
9 Simple data reduction on Pančavská louka 38
10 Data reduction on Hraniční louka with using the rest of data for validation 39
11 Data reduction on Kyselé kouty with using the rest of data for validation 40
12 Data reduction on Pančavská louka with using the rest of data for validation 40
16 Comparion of the importances of texture bands 47
17 Example of the degree of agreement map 53
18 Separability of pairs of classes for Hraniční louka orthomosaic from July the 13th 2023 . . 75
19 Separability of pairs of classes for Hraniční louka orthomosaic from July the 25th 2023 . . 76
20 Separability of pairs of classes for Kyselé kouty orthomosaic from July the 13th 2023 . . . 77
21 Separability of pairs of classes for Kyselé kouty orthomosaic from July the 25th 2023 . . . 78
22 Separability of pairs of classes for Pančavská louka orthomosaic from July the 14th 2023 . 79
23 Separability of pairs of classes for Pančavská louka orthomosaic from July the 26th 2023 . 80
24 Separability of pairs of classes for Pančavská louka orthomosaic from July the 27th 2023 . 81

List of Tables

1 Characteristics of the collected image data 18
2 Legend and characteristics of ground truth data of Hraniční louka peat bog 20
3 Legend and characteristics of ground truth data of Kyselé kouty peat bog 20
4 Legend and characteristics of ground truth data of Pančavská louka peat bog 21
5 Optimized hyperparameters 26
6 Parameters of genetic optimization 28
7 Overall F-1 score of best classifications on each orthomosaic 29
8 Improvements in F-1 score when implementing canopy height feature 30
9 Improvements in F-1 score when implementing texture features 31

10 Number of bands after reduction of correlated bands 31
11 Summary of separability analysis 32
12 Best hyperparameters according to gridsearch and genetic optimization 34
12 Best hyperparameters according to gridsearch and genetic optimization - continued 35
13 Classifications used for overlay analysis 41
14 Results of overaly analysis 41
15 Computational time requirements 48
16 Classes with less than 3 polygons of the ground truth data 49
17 Comparison of accuracy of similar studies and this thesis 51
18 Comparison of F-1 scores for the most important species 52
19 Classification results for Hraniční louka orthomosaic from July the 13th 2023 (F-1 scores) 68
20 Classification results for Hraniční louka orthomosaic from July the 25th 2023 (F-1 scores) 69
21 Classification results for Kyselé kouty orthomosaic from July the 13th 2023 (F-1 scores) . 70
22 Classification results for Kyselé kouty orthomosaic from July the 25th 2023 (F-1 scores) . 71
23 Classification results for Pančavská louka orthomosaic from July the 14th 2023 (F-1 scores) 72
24 Classification results for Pančavská louka orthomosaic from July the 26th 2023 (F-1 scores) 73
25 Classification results for Pančavská louka orthomosaic from July the 27th 2023 (F-1 scores) 74

1 Introduction

1.1 Objectives

Krkonoše National Park (KRNAP, Krkonoše NP), the oldest national park (established in 1963) in the Czech Republic, protects several diverse ecosystems. These include especially the rare relict arcto-alpine tundra, mountain forests and peat bogs. There are 60 peat bogs on the ridges of the Krkonoše Mountains, of which two - Pančavské and Úpské - are among the wetlands protected by the Ramsar Convention (KRNAP, 2022). However, like other wetlands in the Czech Republic, these sites have been systematically drained in the past and are still facing long-term impacts of climate change and human activity (Volf et al., 2019). They therefore in some cases require revitalisation and in all cases systematic monitoring to show the impact of both global and local changes on their species, composition, and biodiversity. Traditional methods of collecting data from peat bogs and monitoring them are proving unsuitable for this purpose, mainly due to their time-consumption, incompleteness, subjective view of each botanist, and high cost (Manfreda et al., 2018). This thesis aims to provide an approach for monitoring of Krkonoše peat bogs using advanced remote sensing methods, specifically analysis of the multispectral data acquired by Unoccupied Aerial Vehicle (UAV).
For this thesis three peat bogs were selected, each of which is characterized by its specific conditions and species composition - Pančavská louka, Hraniční louka and Kyselé kouty. None of these peat bogs had previously been monitored for a long time or mapped in detail. The thesis is a part of long-term research of the Team of Image and Laboratory Spectroscopy (TILSPEC, www.tilspec.cz) from the Department of Applied Geoinformatics and Cartography at the Faculty of Science, Charles University. It is directly related to the results of mapping the distribution of species in the tundra and on Úpa peat bogs in another part of Krkonoše NP (Kupková et al., 2023). The field botanical data were collected in a close collaboration with the botanists of Krknooše National park.
The main objectives of the thesis are:

1. To use multispectral UAV data, field botanical data and remote sensing methods to create accurate maps of current vegetation cover (at the species level) for Krkonoše peat bogs Pančavská louka, Hraniční louka, and Kyselé kouty, aiming for a total classification accuracy/F1-score of at least 0.9.

2. To propose a methodology for efficient and precise vegetation monitoring of Krkonoše peat bogs using multispectral UAV data.

The following steps/methods were used/tested to achieve the objectives:

1. Acquisition of UAV multispectral data and collection of field botanical sample data for all peat bogs at several dates within the vegetation season 2023

2. Use of various (non-spectral) features - several textural features derived from GLCM (Gray level co-occurrence matrix) and canopy height derived using photogrammetric methods.

3. Separability analysis of the ground truth data

4. The testing of two classification methods - Random Forest and Support vector machine based on optimization of hyperparameters to achieve the highest possible classification accuracy
5. The testing of of accuracy after reducing the number of training data
6. The testing of of the spatial agreement of the best classification outputs to produce final highly accurate vegetation cover map
7. Comparison of classification accuracies on different areas of interest and orthomosaics, focusing on the evaluation of the possibility to transfer/generalize the methods

1.2 Theoretical background

1.2.1 Use of non-spectral features in remote sensing

Classification of remote sensing data is a complex process, and its accuracy can be affected by many key factors. One of the most important of these is the input data, which primarily includes spectral information from each pixel. However, a large number of methods can be used to improve classification accuracy, adding new information to the classification model (Lu & Weng, 2007). These can include, for example, spectral indices, geometric or topographic features, but also texture and canopy height models (J. Cao et al., 2018; Erdem & Bayrak, 2023). The latter two are used in the classification process in this thesis.

The term texture feature denotes a new feature, computed from spectral information in the vicinity of a pixel, which is able to express the texture of this vicinity (Hall-Beyer, 2017). Commonly used texture features are second-order (i.e., derived from the image indirectly) and are most often derived from the Gray level cooccurence matrix (GLCM, a matrix measuring how often different combinations of grayscale occur in the vicinity of a pixel), but also from other metrics involving the image context (GLDM - Gray Level Dependence Matrix, GLSZM - Gray Level Size Zone Matrix). The usefulness of texture features derived from the GLCM for classification was already demonstrated by Baraldi & Parmiggiani and Augusteijn et al. in 1995. Mohammadpour et al. (2022) showed that texture features derived from GLCM can significantly improve the accuracy of the Random forest classifier in complex vegetation classification of Lousã district in Portugal. Similarly, for example, Erdem & Bayrak (2023), compared the abilities of various textural features to classify stands of *Pinus sylvestris*. In addition to textural features derived from GLCM, they also used GLDM and GLSZM.

Another key input attribute for classification can also be canopy height. Although this vegetation characteristic is more often used to estimate biomass volume (C. Cao et al., 2016) or is considered as a final product of analysis by remote sensing methods (Xu et al., 2023; Zhao et al., 2022), as demonstrated for example by J. Cao et al. (2018) in their studies of mangrove vegetation, it can also be an important initial input for distinguishing between vegetation classes. Studies from Liu & Bo (2015) or Zhang et al. (2016) show that vegetation height can significantly improve classification accuracy, especially in the case of tree classification. In most cases, LIDAR or SAR sensors are used to measure vegetation height, but photogrammetric methods can also be used to create a digital surface model (DSM) from data taken with an optical camera on a UAV. This method of creating a DSM is more cost

effective than the others mentioned (J. Cao et al., 2018). J. Cao et al. (2018) found that using a DSM derived from UAV optical data can increase the overall accuracy of mangrove classification by about 5%. Li et al. (2023), when classifying different tree species, were able to increase the overall classification accuracy by 2-3% using a DSM derived from LIDAR data. Similar increase in classification accuracy when using the DSM feature was also observed by Dančejová (2023) in mapping vegetation types in the new wilderness near Kutná Hora (Czech Republic).

1.2.2 Use of UAV remote sensing in wetland species mapping

Vegetation mapping is one of the most common uses of remote sensing and has been gaining popularity in the last few years both in landscape and protected area management and in precision agriculture. Traditional methods of vegetation mapping based on field surveys are less accurate, expensive, and very challenging. Therefore, efforts to use different sources of remote sensing data to classify vegetation at the community or species level are becoming increasingly popular (Xie et al., 2008). Wetlands, as one of the world's most important ecosystems, are one of the areas where advanced monitoring can lead to improved understanding of water dynamics and natural carbon cycle (Fraser & Keddy, 2005). According to Davidson (2014), 57% of wetlands have already been destroyed by human activities and a better understanding of the dynamics of these ecosystems could lead to more effective conservation. However, as Huylenbroeck et al. (2020) report, only 4% of studies use remote sensing technology for vegetation management in wetlands. According to the same article, the majority of these papers focus on either invasive species mapping, wetland ecological monitoring or flood modelling. The small percentage of articles using remote sensing methods is likely related to these methods being mostly unknown in the natural landscape management field, and landscape managers are unable to recognize whether remote sensing technique will be applicable in their case. There is also a lack of longer-term collaboration and effective communication between institutional wetland professionals and remote sensing experts (Huylenbroeck et al., 2020).

The use of UAVs in vegetation monitoring can significantly help to reduce the cost and increase the flexibility of image data acquisition and the intensity of their use in remote sensing has been increasing in recent years (Shakhatreh et al., 2019). In studies of vegetation species composition or similar detailed analyses, satellite or aerial imagery may not provide sufficient spatial resolution, and researchers are also leaning towards the use of UAV because of lower financial costs or the possibility of acquiring their own data within a selected time period (Alvarez-Vanhard et al., 2020; Anderson & Gaston, 2013). UAV imagery can produce both multispectral and hyperspectral image data with a variable spatial resolution ranging from a few millimetres to centimetres. One of the main disadvantages of working with the UAVs is the need for geometric data preprocessing, which a number of software programs help to automate (Colomina & Molina, 2014). Other limitations include small payload, battery life and sensitivity to weather conditions. The use of the UAVs is therefore mostly limited to areas of relatively small extent (Matese et al., 2015; Nex & Remondino, 2014). At the same time, many surveys from around the world point out that wetland vegetation is complex and dynamic, its mapping is very challenging, and it is appropriate to use a large amount of data from different sources including UAVs (Bhatnagar et al., 2021; Kaplan & Avdan, 2018).

In the case of mapping protected wetlands, the advantages of the UAV include the ability to explore sites that are difficult to reach and less disruption to the ecosystem during data collection compared to field surveys. A good example of the use of UAV technology in wetland management is a study by Abeysinghe et al., (2019), seeking to accurately map invasive reed canary grass (*Phragmites australis*) in the Old Woman Creek estuary in Ohio, near Lake Erie (USA). The authors used a multispectral camera on UAV and acquired image data of the area on two dates. Several other attributes such as NDVI (Normalized difference vegetation index) and CHM (Canopy Height Model, vegetation cover height) were calculated as inputs for the classification, and pixel-based and object-based approaches and maximum likelihood methods, SVM and neural network (NN) were used. The pixel-oriented neural network achieved the best results and was able to map the occurrences of the invasive species with 98% accuracy. As the study itself notes, such a result could not have been achieved by field surveys or analysis of aerial or satellite data.
A similar approach was applied by Zaman et al. (2011) in northern Utah, where the occurrences of the same invasive species were mapped. They used UAV data and a multiclass relevance vector machine classifier and achieved 95% accuracy. Another study, by Michez et al. (2016), used the Random forest classifier to classify three invasive species in a wetland area in Wallonia. Despite good results for one of the three species (*Heracleum mantegazzianum*, 97%), the authors of the study concluded that the data was not of sufficient quality for the remaining species (*Impatiens glandulifera* and *Fallopia sachalinensis*) and the results could not be used as a basis for decision-making on practical conservation measures. In contrast, using the same algorithm, studies by Bhatnagar et al. (2021) and Alvarez-Vanhard et al. (2020) achieved accuracy of over 85% for vegetation classes. Both studies used multitemporal data and Bhatnagar et al. (2021) report that the use of remote sensing methods allows for significantly cheaper and easier mapping, especially when UAV technologies are used.

1.2.3 Use of remote sensing in vegetation mapping in Krkonoše NP

Several studies used remote sensing for vegetation mapping in Krkonoše NP recently. The majority of the studies have originated from the collaboration between remote sensing experts and botanists of the national park. For example, Potůčková et al. (2021) aimed to study the changes in vegetation land cover in the territory of the national park relict arctic tundra. According to the results, mainly due to the expansion of grassed areas with taller vegetation, the land cover has changed in 44% of the area of interest in the last 80 years alone. The study was carried out using data from archival aerial imagery. Another study (Kupková et al., 2017) dealt with the classification of vegetation in the Krkonoše tundra at the species level using data from Sentinel-2 and AISA Dual and APEX airborne hyperspectral sensors. The overall accuracy of the best classifications (produced by a SVM classifier) was over 84% and the results showed usefulness of the aerial hyperspectral data.
On the Polish side of the border, where the second part of the national park is located, intensive vegetation research is also conducted using remote sensing tools. Researchers from Warsaw and Wroclaw universities use rather frequently aerial hyperspectral data from the APEX sensor. For example, it was used by Marcinkowska et al. (2014), who mapped types of vegetation communities in the western part of the Krkonoše Mountains. Here again, the SVM method achieved the best results,

namely an overall accuracy of 79%. Among other examples of the use of hyperspectral aerial data in the Polish Krkonoše, are the studies Marcinkowska-Ochtyra et al. (2018) and Marcinkowska-Ochtyra et al. (2017), which both classified vegetation communities above the treeline. In both cases, the SVM classifier was used, and the resulting overall accuracy was 84% and 90%. Although mapping vegetation communities is not identical to mapping at the species level, several of the vegetation classes used in these studies (*Molinia caerulea*, *Pinus mugo*, *Vaccinium myrtillus*) directly correspond to the species classes classified in this thesis. Another wide area in the zone above the treeline was mapped by Sucha et al. (2016), who achieved an overall accuracy of 83.5% when classifying the basic species from satellite and aerial images. Of the several models compared, OBIA (object-based image analysis) achieved best accuracy in this case using the SVM algorithm on aerial imagery.

Another example of remote sensing application in the Krkonoše Mountains for vegetation mapping is a recent study by scientists from the Team of Image and Laboratory Spectroscopy (Kupkova et al., 2023). The main objective was to evaluate the ability of some classifiers to distinguish basic grass species in the Krkonoše tundra using multispectral and hyperspectral UAV data. The study achieved excellent results for common grass species, for which F-1 scores reached over 0.95.

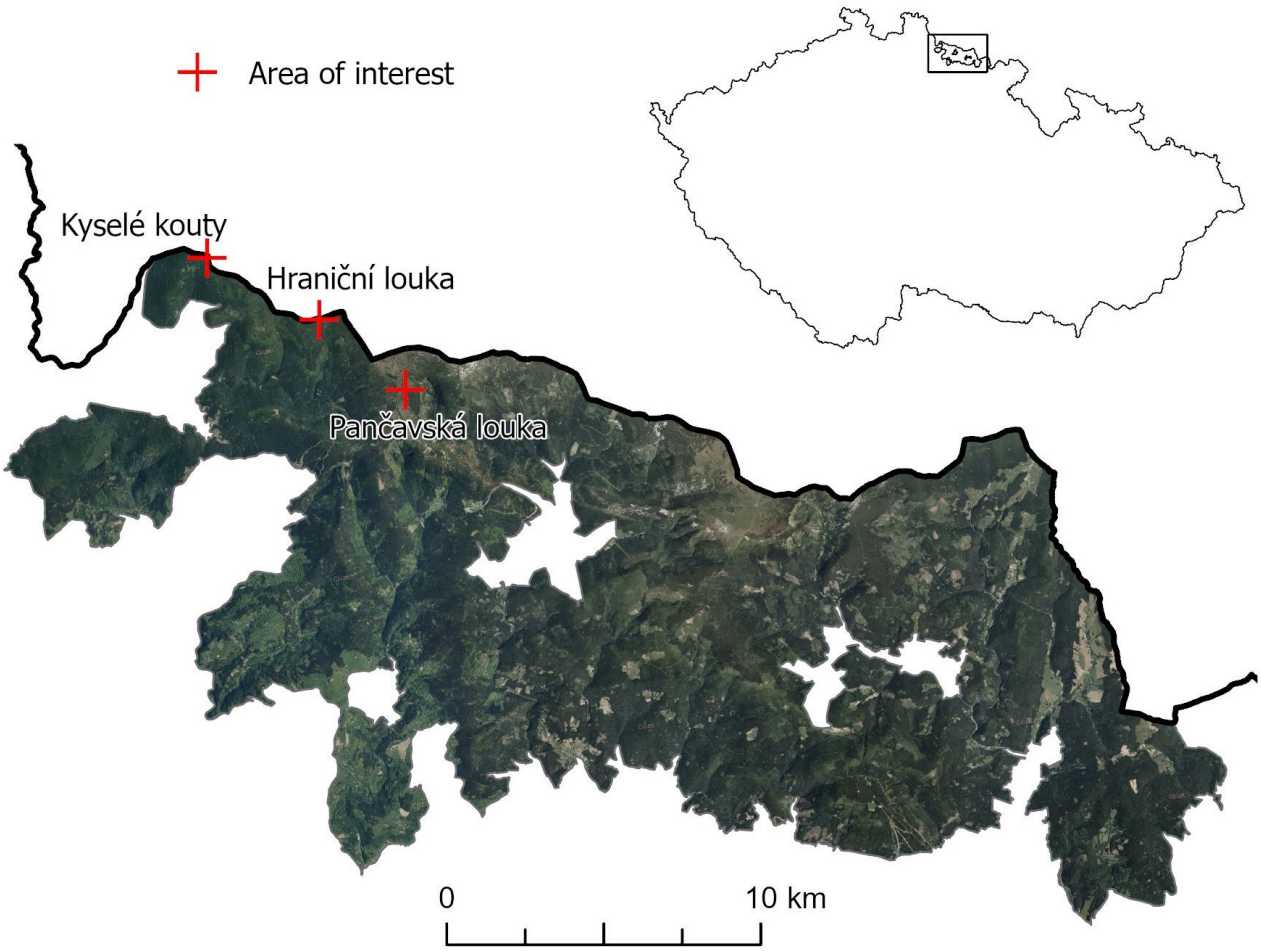
In Krkonoše, several studies have also focused directly on peat bogs. In addition to the bachelor thesis, which this thesis builds on (Kulich, 2022), there is for example an ongoing project on Úpa peat bog in the eastern Krkonoše (Kupková et al., 2020), where a classification accuracy of 86 % was achieved using object-based classification. The report demonstrates the potential of utilizing various sensors (multispectral and hyperspectral) and classification methods, along with different dates of image data acquisition throughout the vegetation season.

2 Areas of interest

2.1 Peat bogs in Krkonoše NP

Peat bog is a specific type of wetland, the main characteristic of which is the gradual deposition of dead plant remains in waterlogged, oxygen-deficient areas (NP Šumava, 2019). In Krkonoše NP, there are about 60 peat bogs, collectively covering an area of 268 ha. These peat bogs are similar to those in Northern Europe in terms of their conditions and are home to a large number of glacial relicts and endangered species (KRNAP, 2022). Peat bogs in Krkonoše NP can be divided into two types, namely forest peat bogs (or peat swamp forests), where peat areas are located among trees, and ordinary peat bogs, forming open areas where trees do not grow. This thesis focuses on three peat bogs in the western Krkonoše: Hraniční louka, Kyselé kouty (Sour Corners), and Pančavská louka (Pančava Meadow). Krkonoše NP and the areas of interest are presented in figure 1.

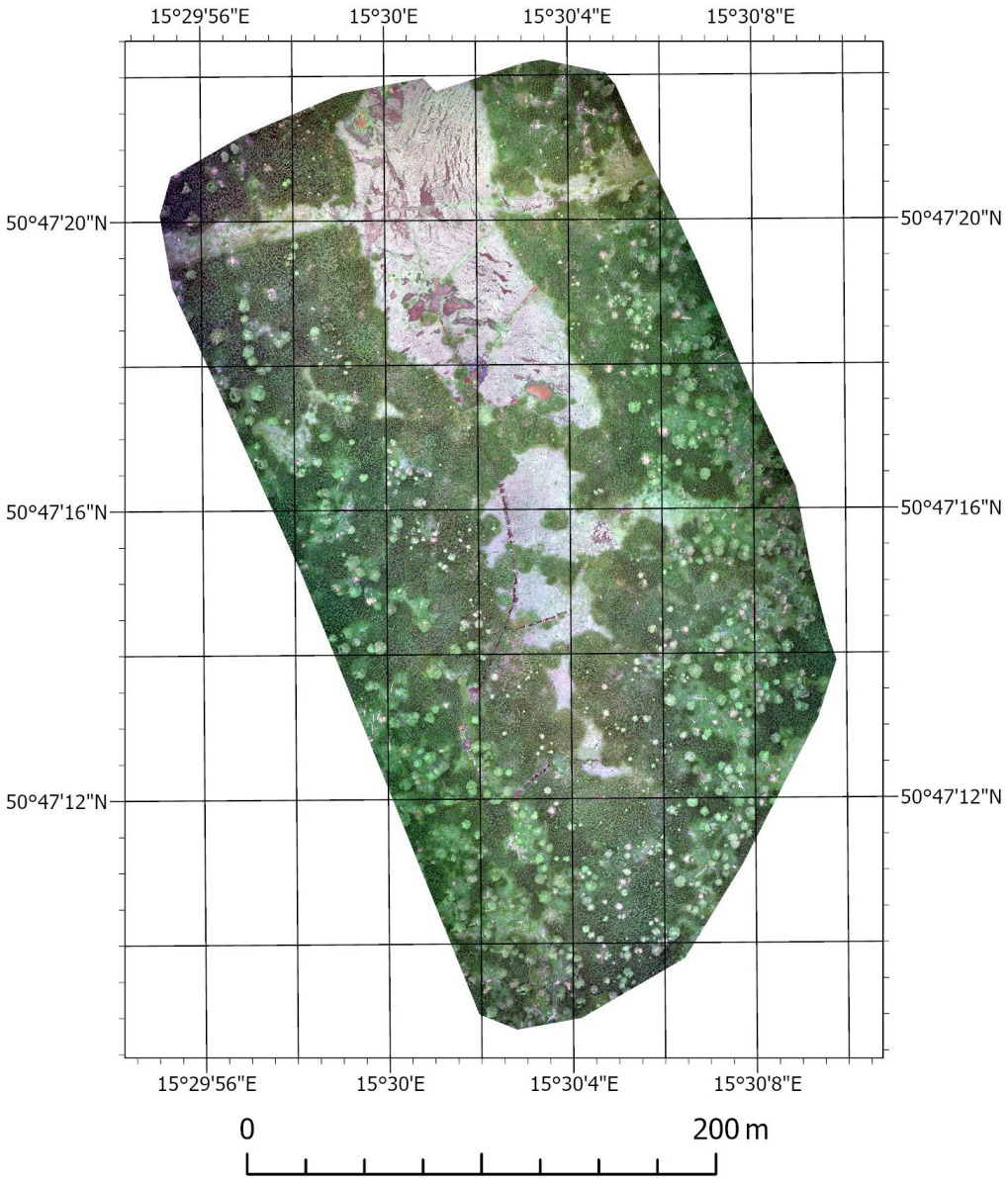
Figure 1: Orthofotomap of Krkonoše NP with the three areas of interest



2.2 Hraniční louka (Border meadow)

The first of the areas of interest is located in the saddle below the summit of Luboch mountain (1296 meters above sea level) at an altitude of about 1250 meters above sea level. It is a very valuable peat bog location, which is an important part of the habitat of the black grouse (*Lyrurus tetrix*) (Volf et al., 2019). The peat bog is surrounded by dense growth of mountain pine (*Pinus mugo*) with occasional solitary spruces (*Picea abies*). On the open area itself, *Trichophorum cespitosum* grows abundantly and in the ponds, *Carex limosa* and *Carex rostrata* are found. *Molinia caerulea*, *Vaccinium uliginosum*, *Vaccinium myrtillus*, and *Nardus stricta* are also found in limited amount on the peat bog. Around the peat lakes, a significant amount of *Drosera rotundifolia* grows. Hraniční louka is ecologically in good condition, and the first phase of revitalization has taken place here.

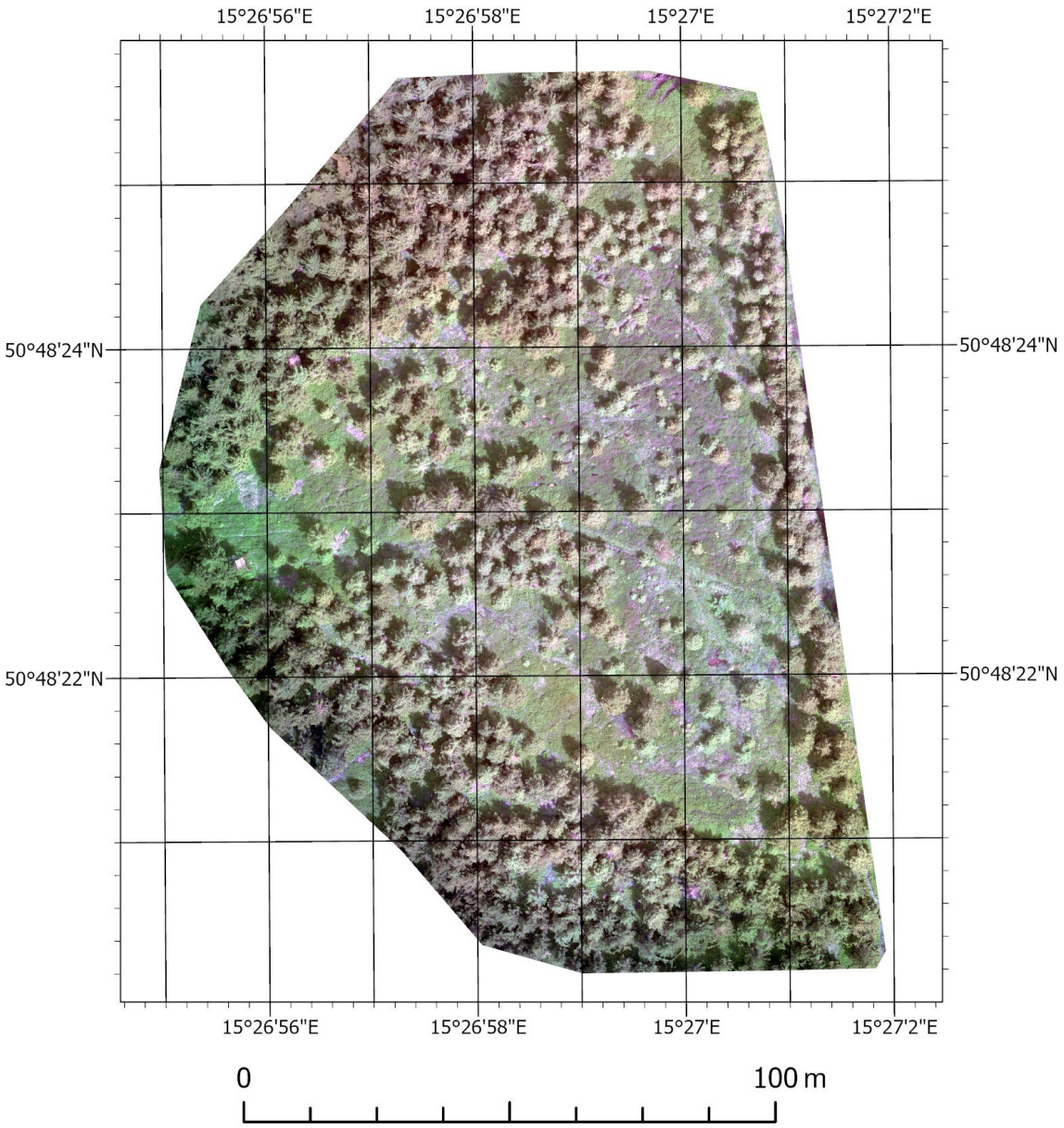
Figure 2: Hraniční louka, 13 July 2023



2.3 Kyselé kouty (Sour corners)

This is the only forest peat bog of the selected sites and is located on the Czech-Polish border on the north side of the westernmost part of the Krkonoše main ridge at an altitude of about 1050 m above sea level. The peat bog was partially drained, and some of the trees on it were cut down, so today the most widespread species found across most of the territory is *Vaccinium myrtillus*. On the remnants of the peat, *Carex sp.* and *Juncus sp.* are found in some places, and beyond the borders of the deforested area, a waterlogged spruce forest (*Picea abies*) grows. During the revitalization of this peat bog, an increase in the area of *Sphagnum* moss is expected, along with an increase in peat bog species (Volf et al., 2019).

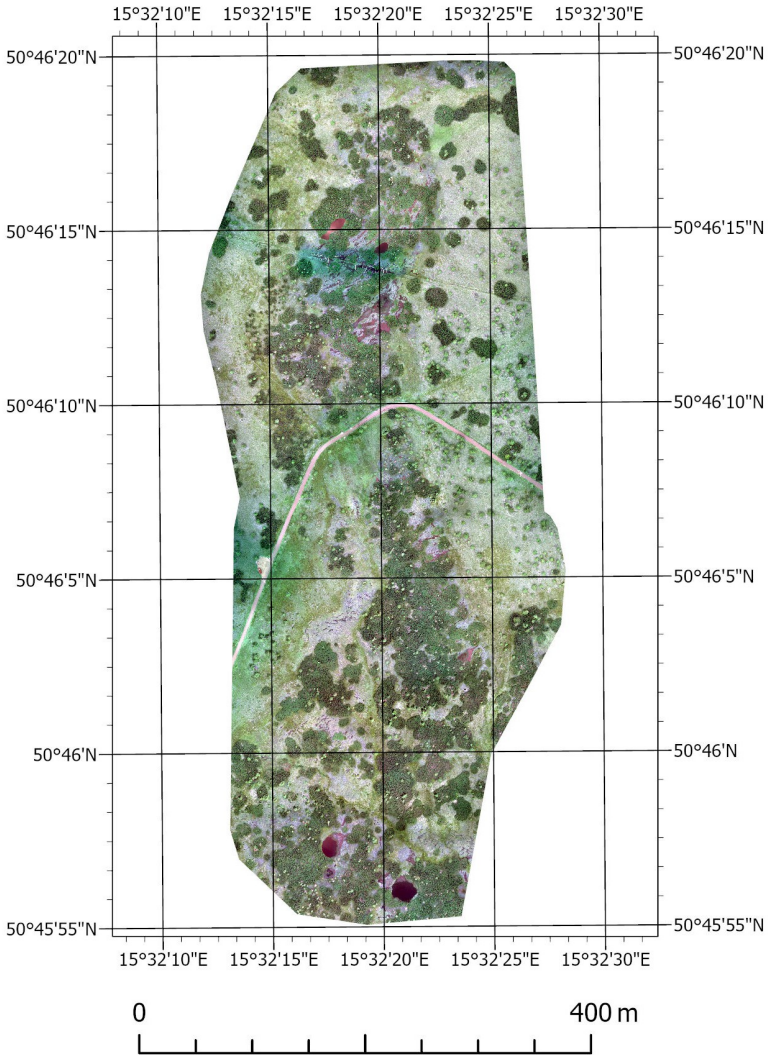
Figure 3: Kyselé kouty, 13 July 2023



2.4 Pančavské louka (Pančava Meadow)

Pančavská louka is the most significant, complex, and largest of all the studied sites. It is located in the highest parts of the western Krkonoše mountains near the source of the Elbe River at an altitude from 1320 to 1370 meters above sea level. The majority of its area is composed of scattered islands of wetland grasses such as *Molinia caerulea*, *Trichophorum cespitosum*, and *Nardus stricta*. Wide stands of dwarf mountain pine (*Pinus mugo*) and occasional solitary spruces (*Picea abies*) also appear around the inner parts of the peat bog. In addition to these species, *Vaccinium uliginosum*, *Vaccinium vitis-idaea*, *Calluna vulgaris*, and several types of grasses (*Carex limosa*, *Carex rostrata*, *Deschampsia cespitosa*, *Eriophorum vaginatum* and *Anthoxanthum odoratum*) also grow on the peat bog. The peat bog has already been revitalized, but global human-induced changes (contamination with nitrogen fertilizers and climate changes) likely lead to the spread of less resistant *Molinia caerulea* at the expense of *Nardus stricta* (Hejcman et al., 2010). These changes will be one of the key points of research focus on Pančavská louka in the future.

Figure 4: Pančavská louka, 14 July 2023



3 Data and methods

3.1 Data collection and classification legend

Two data sets were used in the study, acquired by the author of this thesis, the TILSPEC team from the Faculty of Science, Charles University, and in collaboration with the botanist Viera Horáková from the Krknoše NP Administration on July 13-27, 2024. The first crucial step was acquisition of image data, collected using DJI Phantom 4 Multispectral UAV. The DJI Phantom employs RTK (Real-time kinematic) technology during imaging, enabling precise localization with each shot and simplifying the georeferencing of image data. The camera captures in five bands: Blue (450 nm ± 16 nm), Green (560 nm ± 16 nm), Red (650 nm ± 16 nm), Red Edge (RE, 730 nm ± 16 nm), and Near Infrared (NIR, 840 nm ± 26 nm) (DJI, 2019). Image data collection is summarized in the table 1.

Table 1: Characteristics of the collected image data

Area of interest	Date of acquisition	Overlap	Spatial resolution	Light conditions
Hraniční louka	13 July 2023	Front 77 % / Side 85 %	3.5 cm	Around 4 pm, cloudy
	25 July 2023	Front 77 % / Side 85 %	3.5 cm	Around 5 pm, cloudy
Kyselé kouty	13 July 2023	Front 80 % / Side 85 %	3.5 cm	Around noon, clear sky – visible shadows
	25 July 2023	Front 80 % / Side 85 %	3.5 cm	Around noon, cloudy
Pančavská louka	14 July 2023	Front 70 % / Side 70 %	5 cm	Around noon, cloudy
	26 July 2023	Front 70 % / Side 70 %	5 cm	Around 5 pm, partly cloudy
	27 July 2023	Front 70 % / Side 70 %	5 cm	Around 11 am, clear sky

In the field, at the same time, a large amount of positional data was collected using GPS, intended (after pre-processing) as training and validation data for the image data classifications. Data was collected with Trimble R7 and R10 geodetic GPS receivers, and attributes were added to each record. These included the vegetation species at the targeted location and the radius of the circle around the point in which this species is guaranteed to occur. Spatial accuracy for GPS measurements ranged between 5 and 20 mm. For some easily recognizable vegetation classes (*Pinus mugo*, *Picea abies*) and other non-vegetation classes (water, dry tree trunks), data was manually collected from the orthomosaic.

3.2 Image data preprocessing

Raw data from the UAV is produced in the form of individual images with added geographical information, obtained with the help of the RTK system. To obtain an orthomosaic, it is necessary to

merge the images together using photogrammetric methods with the detection of automatic tie points. This process was carried out in the Pix4D Mapper software. Before creating the orthomosaic, the images were examined in Agisoft Metashape Professional software, which allows for easier work with individual images, and some of them were excluded from the analysis due to their poor RTK data quality. Pix4D Mapper generated an orthomosaic for each of the 5 spectral bands and a DSM, calculated from the photogrammetric 3D point cloud after processing.

3.3 Ground truth data processing

The raw data was obtained from GPS in the form of a text file with individual records. From these records, several incorrectly marked and test records were removed. The GPS receiver Trimble R7 provided points in the S-JTSK coordinate system, which is a local coordinate system used in the former Czechoslovakia, while Trimble R10 in two coordinate systems: WGS-84 and S-JTSK. Since the orthomosaics from the UAV were generated in the UTM-33N coordinate system, which was chosen as suitable for the entire analysis, it was necessary to convert the GPS data into this coordinate system as well. Unlike conversion from WGS-84, the conversion from S-JTSK to UTM-33N is very demanding and not always accurate. The best freely available transformer from ČUZK (Český úřad zeměměřický a katastrální, Czech cadaster office) (ČÚZK, 2010) was therefore used for the conversion. Based on species found in the field, classification legend for each area of interest was elaborated – see tables 2, 3 and 4. When creating the final legend, some species were combined into categories by genus because they could not be distinguished from each other during the analyses (they are listed as species/sp. in the legend). Images of all classified species can be found in appendix A.

Table 2: Legend and characteristics of ground truth data of Hraniční louka peat bog

Class	Number of polygons	Area of ground truth data [m^2]	Corresponding number of pixels
Vaccinium myrtillus	1	0.28	229
Carex limosa	8	0.99	809
Carex rostrata	5	1.41	1152
Unidentified moss	3	0.19	156
Molinia caerulea	5	1.92	1568
Nardus stricta	1	0.28	229
Sphagnum species	9	1.87	1527
Trichophorum cespitosum	13	3.68	3005
Vaccinium uliginosum	5	0.57	466
Pinus mugo	14	11.00	8980
Picea abies	12	9.42	7690
Water	5	3.42	2792
Dry vegetation	5	0.99	809
All combined	86	36.03	29412

Table 3: Legend and characteristics of ground truth data of Kyselé kouty peat bog

Class	Number of polygons	Area of ground truth data [m^2]	Corresponding number of pixels
Bare ground	8	0.97	792
Vaccinium myrtillus	32	51.71	42213
Carex species	2	0.41	335
Deschampsia cespitosa	3	0.27	221
Eriophorum vaginatum	5	1.23	1005
Juncus species	12	1.76	1437
Unidentified moss	13	1.06	866
Nardus stricta	5	1.05	858
Pinus uncinata subsp. uliginosa	1	0.13	107
Picea abies	18	9.22	7527
Sphagnum species	4	0.10	82
Dry vegetation	4	0.28	229
All combined	110	68.18	55672

Table 4: Legend and characteristics of ground truth data of Pančavská louka peat bog

Class	Number of polygons	Area of ground truth data [m2]	Corresponding number of pixels
Vaccinium myrtillus	16	3.28	1312
Vaccinium vitis-idaea	3	0.53	212
Calamagrostis villosa	2	0.91	364
Carex limosa	24	8.84	336
Carex rostrata	29	19.74	7896
Deschampsia cespitosa	4	9.08	3632
Eriophorum angustifolium	10	4,34	1736
Hieracium species	2	0.53	212
Juncus species	2	0.27	108
Unidentified moss	8	0.87	348
Molinia caerulea	106	84.13	33652
Nardus stricta	55	43.15	17260
Salix species	1	0.79	316
Sphagnum species	13	1.46	584
Anthoxanthum odoratum	2	1,57	628
Trichophorum cespitosum	70	54,20	21680
Vaccinium uliginosum	16	3.02	1208
Calluna vulgaris	21	4.18	1672
Water	5	15.71	6284
Pinus mugo	15	47.12	18848
Picea abies	11	34.56	13824
Dry vegetation	7	1.67	668
All combined	422	339.95	132780

3.4 Canopy height feature creation

Although Pix4D Mapper generated a DSM along with the multispectral orthomosaic, this product could not be used as a model training feature due to uneven terrain. To obtain the vegetation height, it was therefore necessary to subtract the terrain's elevation from the DSM. Pix4D Mapper allows for automatic digital terrain model (DTM) creation by filtering the point cloud, but due to complex vegetation in the area of interest (dense stands of *Pinus mugo*, undulating ground surface covered by *Vaccinium myrtillus*), this model was not sufficiently accurate. Therefore, the freely available DMR5G point cloud from ČÚZK (ČUZK, 2016) was used to calculate vegetation height. In the case of Hraniční louka, where part of the classified area is located beyond the state border, the Polish DTM, Numerical Terrain Model (NMT) (Główny Urząd Geodezji i Kartografii, 2023), was also used. The evaluation of the accuracy improvement in classifications when including vegetation height can be found in table 8 on

page 30. For illustration, a canopy height raster is displayed in figure 5 on the next page.

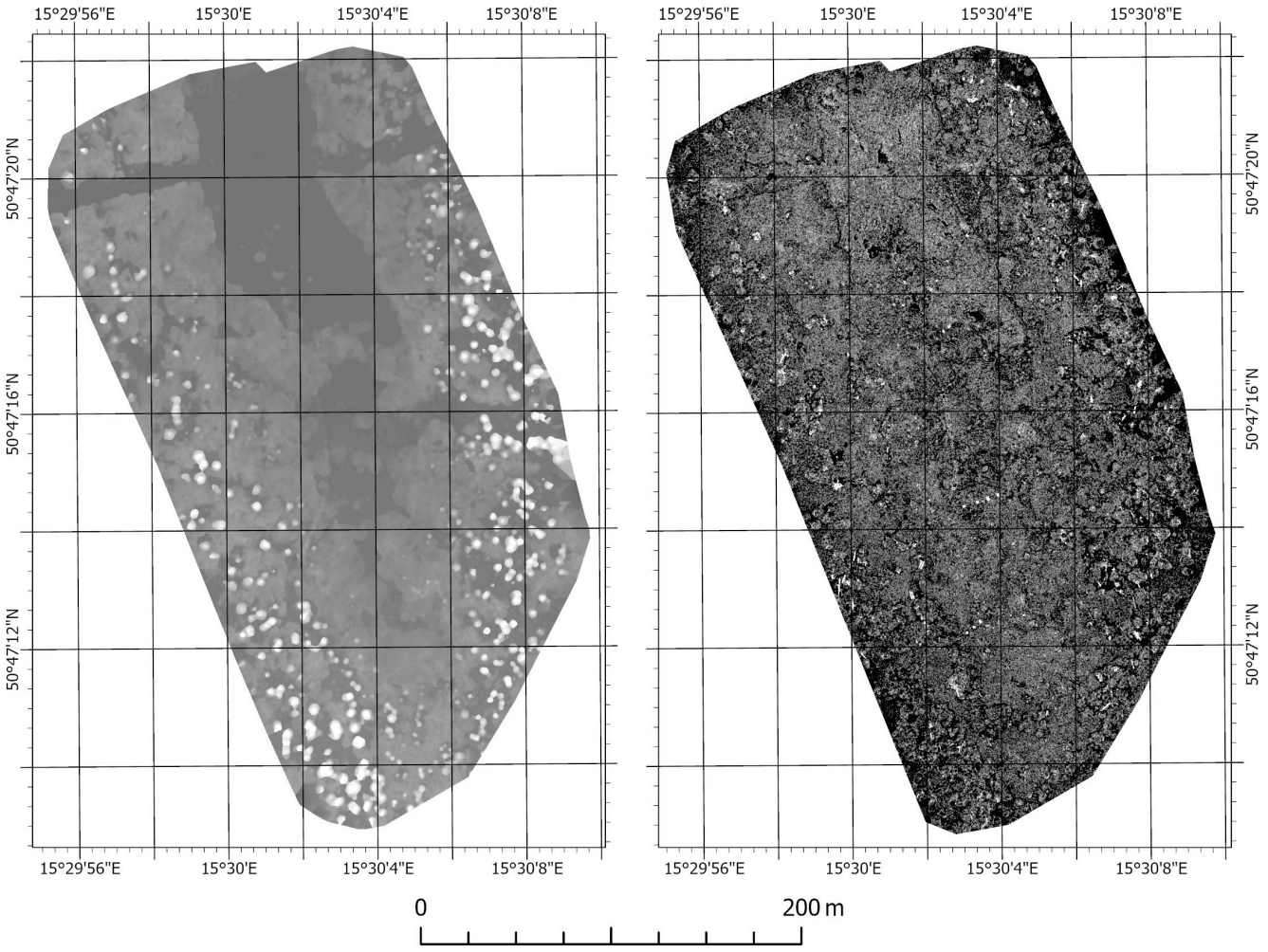
3.5 Texture feature creation

As mentioned, texture features were calculated from the GLCM, which is a matrix representation of frequency of occurrence of pairs of gray values in the neighborhood of each pixel. The size of this neighborhood can be selected and is referred to as the kernel size in the GLCM calculation. Since the GLCM is calculated from grayscale, not from a multispectral image, one GLCM was calculated for the average of the RGB bands, and two separate ones for each of the NIR and RE bands. For each of these new greyscale images, 7 GLCMs with kernel sizes of 3, 5, 7, 9, 11, 13, and 15 were calculated. For each of these GLCMs, 9 texture bands were subsequently calculated, following (Hall-Beyer, 2017):

1. GLCM Mean, calculated as $\mu = \sum_{i,j=0}^{n-1} iP_{i,j}$
2. Standard deviation, calculated as $\sigma = \sum_{i,j=0}^{n-1} P_{i,j}(i-\mu)^2$
3. Correlation, calculated as $\sum_{i,j=0}^{n-1} P_{i,j} \left[\frac{(i-\mu)(j-\mu)}{\sigma^2}\right]$
4. Contrast, calculated as $\sum_{i,j=0}^{n-1} P_{i,j}(i-j)^2$
5. Dissimilarity, calculated as $\sum_{i,j=0}^{n-1} P_{i,j}|i-j|$
6. Homogeneity, calculated as $\sum_{i,j=0}^{n-1} \frac{P_{i,j}}{1+(i-j)^2}$
7. ASM, calculated as $\sum_{i,j=0}^{n-1} P_{i,j}^2$
8. Maximum, which is the maximum value in GLCM
9. Entropy, calculated as $\sum_{i,j=0}^{n-1} P_{i,j}(-\log P_{i,j})$

The evaluation of the improvement in classification accuracy when including texture features can be found in table 9 on page 31. All calculations in this part of the analysis were performed using a custom script in the Python programming language with the help of the NumPy, ArcPy, GDAL, OGR, scikit-image and opencv libraries. For illustration, a texture feature raster is displayed in figure 5 on the next page.

Figure 5: Illustration of non-spectral features. The canopy height model for Hraniční louka (left) and the standard deviation from GLCM for Hraniční louka (right), July 13th



3.6 Ground truth data separability evaluation and cleaning

Like any other step in scientific research, the collection and processing of field data can be subject to inaccuracies or direct errors. One of the goals of this thesis is therefore to propose an algorithm that will detect these errors as efficiently as possible and thus improves the accuracy of the classification. Based on experiences with data collection in the field and its processing, it appears that the quality of ground truth data can decrease in the following cases:

1. GPS receiver error - for example, inaccurate measuring of coordinates.
2. Human error during data acquisition - for example, incorrect determination of vegetation species, incorrect recording of vegetation class, incorrect estimation or measuring of the polygon's diameter.
3. Technical error during processing - for example, incorrect coordinate transformation.

4. Human error during processing - overwriting or deleting of important data, incorrect targeting of coordinates from the orthomosaic (in the case of classes that are collected over the orthomosaic), or incorrect estimation of the polygon size in the case of these classes.

5. The targeted polygon does not correspond with the data on the orthomosaic - for example, the area defined by the polygon is partially or completely in shadow on the orthomosaic, or, when collecting data at different times, some places may be recently flooded or changed according to the seasonal development.

Some of the possible errors can be easily detected, for example, in the case of a GPS receiver error, information about the accuracy of the measurement is usually available. In the case of an incorrect coordinate transformation, it is a systematic error that can be easily observed. However, if one of the human errors occurs, it is not easy to detect, and when searching for these errors, it is necessary to rely on the ability to measure the quality of the training data. One such option is the analysis of separability of individual classes. In this thesis, it the Jeffries-Matusita index (J-M index) was used, which is one of the most common metrics for these purposes in remote sensing (H. Liu et al., 2020; Richards & Jia, 2006; Schmidt & Skidmore, 2003; Sothe et al., 2019). The J-M index is a statistical measure indicating the distance between two probability distributions ranging from 0 (identical) to 2 (completely separable). The J-M index for two classes with many features can be calculated as

$$J_i, j = 2(1 - e^{-B})$$

in which

$$B = \frac{1}{8}(m_i - m_j)^t \left[\frac{\Sigma_i + \Sigma_j}{2}\right]^{-1} (m_i - m_j) + \frac{1}{2}\log\left[\frac{|(\Sigma_i + \Sigma_j)/2|}{|\Sigma_i|^{1/2}|\Sigma_j|^{1/2}}\right]$$

in which m_i, m_j reffer to matrices of means and Σ_i, Σ_j reffer to covariance matrices.
The J-M index is always calculated only for a pair of vegetation classes and should have lower values with poor separability of these classes (Richards & Jia, 2006). For the separability analysis performed in this thesis, the reasoning was that if classes in some polygons of ground truth data were switched or mixed (based on the human errors), the J-M index should decrease compared to when the ground truth data are truly of high quality (without such errors).
The proposed algorithm first extracts the complete set of ground data and measures the separability of every two classes using the J-M index. It then focuses on all classes that do not achieve sufficient separability (the threshold for the sufficient J-M index can be set within the algorithm and was named threshold 1) and checks if the J-M index increases if each one of the polygons is removed from the dataset. If the J-M index increases, the result is saved, and the polygon is suggested for inspection or removal from the ground data (how significant increase in J-M index is needed can also be set within the algorithm using threshold 2). Since separability can change depending on the number of input features, it is not easy to set either of the two thresholds straightforwardly. For basic testing in this thesis, 10 input features were chosen for this analysis (all spectral bands, vegetation height, and several texture bands), and the thresholds were set sequentially at 1.9 and 0.1. The value of 1.9 was thus designated as a minimal satisfactory level of separability and 0.1 as a significant improvement in

separability. The algorithm can be run without these threshold restrictions, but removing almost any of the polygons may increase separability, probably due to the elimination of some outlier pixels. This algorithm, designed by the author, will be referred to in the thesis as the Faulty polygon detector (FPD). All calculations in this part of the analysis were performed using a custom script in the Python programming language with the help of the NumPy, ArcPy, GDAL, OGR and SciPy libraries.

3.7 Reduction of the number of features

Because a total of 189 texture bands were calculated for each orthomosaic, it was deemed necessary to reduce their number for faster analysis. The most common reduction technique is Principal Component Analysis (PCA) (Kupková et al., 2017; Mohammadpour et al., 2022; Richards & Jia, 2006), but PCA would not allow examining which bands have what effect on the outcome. Therefore, the bands were reduced based on mutual correlation. To accelerate the process, mutual correlation of bands was calculated only in for the ground truth data, as these are most important parts of the orthomosaic for the analysis. From the resulting raster, bands that reached Pearson correlation coefficient greater than 0.95 with another band were removed. The evaluation of classification accuracy after this reduction and when including multiple bands simultaneously can be found in 4.2. All calculations in this part of the analysis were performed using a custom script in the Python programming language with the help of the NumPy, ArcPy, GDAL, OGR and SciPy libraries.

3.8 Selection and application of a suitable classifier

Choosing the right classifier is one of the most important steps in classification itself. There are significant differences in the implementation requirements and accuracy among commonly used algorithms. Since this thesis aims to establish an easily replicable path usable for the management of Krkonoše NP, the demands are high in both aspects. The ideal algorithm should be easy to implement and, at the same time, achieve good accuracy. The simplest classifiers, such as the maximum likelihood classifier or simple logistic regression, did not achieve good accuracy in comparable studies (Kulich, 2022; Kupková et al., 2017, 2023) or in initial tests on data from Krkonoše peat bogs. Therefore, classifiers that achieved good results in similar conditions (J. Cao et al., 2018; Du et al., 2021; Kupková et al., 2017, 2023; Windle et al., 2023; Zagajewski et al., 2021) were chosen: Random Forest (RF) and Support Vector Machine (SVM). Random Forest was chosen as the primary classifier because it was expected to achieve better accuracy, mainly because this was demonstrated in the bachelor thesis (Kulich, 2022). Support Vector Machine was then chosen as a secondary algorithm to achieve the best possible accuracy with it and compare its usability with Random Forest. After selecting the algorithm, it was also necessary to choose the best model hyperparameters for each, i.e., external parameters affecting its structure. This process is often referred to as hyperparameter tuning (Simon et al., 2023). The RF algorithm falls into the category of ensemble algorithms, its principle being the combination of many randomly created decision trees, and the final categorization of a pixel is decided based on the voting of subordinate models (Breiman, 2001). It was chosen as the main classifier primarily for its ability to classify nonlinear data, its capability to assess the importance of individual features, and its

relatively fast training and prediction. The most commonly tuned hyperparameters of the RF algorithm are ntree (the number of trees in the model), mtry (the number of possible splits in each node of a tree), and the maximum depth of a tree (Probst et al., 2019).
The SVM algorithm, on the other hand, seeks an optimal separating hyperplane that maximizes its distance from the nearest data in hyperspace. Its hyperparameters include, firstly, the regularization parameter C, determining the degree of tolerance for misclassifications during training. For classifying nonlinear data, a kernel function is also used, recalculating relationships between all pairs of points (this transformation is sometimes called the "kernel trick"). The kernel function can be polynomial, RBF (Radial Basis Function), or sigmoid. For the last three functions, the gamma parameter, describing the size of the function curve, needs to be set, and the polynomial function has a degree parameter (Scikit-learn developers, 2024; Vapnik & Cortes, 1995). The classification results for SVM varied significantly for each kernel used, and therefore, the results for each kernel will always be mentioned separately in the result tables 7.
All calculations in this part of the analysis were performed using a custom script in the Python programming language with the help of the Scikit-learn, NumPy, ArcPy, GDAL, OGR and SciPy libraries. All tuned hyperparameters of both models are listed in table 5.

Table 5: Optimized hyperparameters. In square brackets, the ranges within which the hyperparameter values were tested are indicated.

<table>
<tr><th>algorithm</th><th colspan="3">hyperparameters</th></tr>
<tr><td>Random forest</td><td>ntree [30-1000]</td><td>mtry [3-20]</td><td>Maximum tree depth [3-25]</td></tr>
<tr><td>SVM – linear function</td><td rowspan="3">Regularization parametr C [0-1000]</td><td></td><td></td></tr>
<tr><td>SVM – polynomial function</td><td rowspan="3">Gamma parameter [0-1000]</td><td>Polynomial degree [2,3]</td></tr>
<tr><td>SVM – RBF function</td><td></td></tr>
<tr><td>SVM – Sigmoid function</td><td></td></tr>
</table>

When tuning these algorithms and evaluating their accuracy, it is first necessary to define a measure of their accuracy. Commonly used measures of accuracy include overall accuracy, defined as the ratio of correct predictions to all predictions, and the F-1 score, defined for individual classes as the harmonic mean of their precision and recall. As a suitable measure of accuracy for this thesis, the average F-1 score of individual classes, weighted by their area in the chosen area of interest, was selected. If the term "overall F-1 score" is mentioned later in the thesis, it refers to this weighted average. The area of the classes was determined by signal classification using the RF classifier with hyperparameters set as ntree - 500, mtry - square root of number of features, max depth - None. These default parameters were selected because the literature suggests that RF should achieve good accuracy even in this typical default setting (Probst et al., 2019). For each orthomosaic, the entire tuning process was conducted separately, and achieving the best overall F-1 score, ideally over 0.9, was most crucial. The final F-1 score was measured using a test dataset, which was carefully selected to be as representative as possible and was excluded from training and validation during hyperparameter tuning. The test dataset

contained approximately 25 % of the ground truth data. For the tuning, k-fold cross-validation was used for validating each new model, meaning the validation was done several times with several validation subsets, and training was conducted with the remaining data. With given hyperparameters, the accuracy of the algorithm was evaluated as the average overall F-1 score of all validations. The overall accuracy and final overall F-1 score were then assessed on the test dataset with the model with optimal hyperparameters, trained on the entire remainder of the data.

Data division into cross-validation subsets was done sequentially, so each validation was conducted with an entirely different dataset. At Hraniční louka and Kyselé kouty, where less ground truth data was available, the training and validation dataset was divided into 3 cross-validation folds of the same sizes, while data from Pančavská louka was divided into 5 cross-validation folds. Since some classes had low number of polygons collected (sometimes even 1 or 2), it was necessary in some cases to divide them for validation and testing. The polygons were divided so that there was no overlap between the individual datasets. However, it is necessary to note that, as was discovered in initial testing, the accuracy measurement of classification for these classes is significantly affected by this, due to the spatial dependency training, validation, and testing datasets.

For selecting optimal hyperparameters of models, a grid search was always first conducted. This involves repeatedly validating the model for different hyperparameters in a regular grid of pre-set values. The results of the grid search allow understanding the behavior of the algorithm across a wide range of settings and understanding which changes in hyperparameters lead to a change in accuracy and which do not. However, grid search may not find the best solution, especially if the accuracy of the model depends on small changes in hyperparameters or fails to capture the range in which optimal values lie. For these reasons, a second approach was also tried, optimization with a genetic algorithm (Holland, 1975; Lambora et al., 2019). This optimization algorithm is a machine learning method, mimicking natural evolution in nature. It first initializes a population of models with random hyperparameters in given range, evaluates them for cross-validation average overall F-1 score, and then in each iteration randomly combines their parameters and adds random mutations. From the resulting models, only those with the best overall F-1 score proceed to the next iteration. The best model hyperparameters are then saved. In practice, the hyperparameters of a given model are recorded into a bit chain. A new population is then created by combining these chains, and mutations are created by flipping a random bit. The mutations generate models with entirely different hyperparameters in some populations, thus reducing the chance of the genetic algorithm reaching a local optimum. After studying of the relevant literature (Á. E. Eiben et al., 1999; A. E. Eiben & Smit, 2011) and initial testing, the parameters in table 6 on the next page of the genetic algorithm were used for hyperparameter tuning.

Table 6: Parameters of genetic optimization

Parameter	Value
Population size	50
Number of bits	16
Mutation rate	0.02
Crossover rate	0.9
Number of iterations	20

3.9 Testing the classifier's behavior when reducing the number of ground truth data

For monitoring vegetation in Krkonoše NP, a key question is also how much ground truth data needs to be collected for the classification to have a good accuracy, which in the case of this thesis corresponds to an overall F-1 score of at least 0.9 (in case of lower scores for the best classifier, a comparable result with the best achieved). Therefore, an algorithm was created to examine the accuracy of classifications after removing a predetermined amount of field data. The proportion of data (labeled in the algorithm as the data removal index) that were not included in training or validation was gradually increased, and random polygons, the number of which corresponded to the data removal index, were always removed. The dataset was then randomly divided into two parts: approximately two-thirds constituted the training data, and the remaining third made up the validation data. Thus, training and validation were conducted 20 times for each index. It was assumed that with worse data representativeness (in this case, also a lower number of polygons) in the training dataset, the data would have high variability, leading to the inability to generalize the model and lower accuracy. Subsequently in a second analysis, each of the removed polygons was included in the validation dataset, and the rest of the process remained the same. This made it possible to measure the accuracy the model could achieve on more independent data. This procedure was again tried for all orthomosaics. All calculations in this part of the analysis were performed using a custom script in the Python programming language with the help of the Scikit-learn, NumPy, ArcPy, GDAL, OGR and SciPy libraries.

3.10 Creation of assembled vegetation cover maps

In addition to vegetation cover maps predicted by already prepared models, a further accuracy check was carried out by assembling the resulting maps together. On each area, all classifications achieving an overall F-1 score of at least 0.9 (or comparable to the best algorithm's overall F-1 score in case of lower accuracy) were selected. Which classifications were selected for overlays can be seen in table 13 on page 41. These classifications outputs were then overlaid on each other, and a new (ensemble) classification was created, selecting for each pixel the class on which most of the individual classifiers agreed. The new classification accuracy was then measured on the test dataset, and it was determined how many of the classifiers agreed on what portion of the area of interest. Therefore, a second map (see

figure 17 on page 53) was created for the resulting vegetation cover map, describing how many of the maps agree on the species prediction at a given location. Besides serving as an additional accuracy check of the classifications, this analysis can distinguish places that were likely accurately mapped from places where some models may have been confused, and it might be necessary to check them in the field. It may also reveal inaccuracies in the classification of individual classes throughout the whole area of interest, independently on test dataset. All calculations in this part of the analysis were performed using a custom script in the Python programming language with the help of the Scikit-learn, NumPy, ArcPy, GDAL, OGR and SciPy libraries.

4 Results

4.1 Classification results

The best hyperparameters, the final F-1 scores for the entire area and individual classes for both orthomosaics of all areas of interest are summarized in table 7. Complete classification results for all orthomosaics and classes are available in appendix B.

Table 7: Overall F-1 score of best classifications on each orthomosaic

Area of interest, date	RF	Linear SVM	Polynomial SVM	SVM with RBF	SVM with sigmoid
Hraniční louka, 13 July 2023	0.899	0.875	0.879	0.895	0.779
Hraniční louka, 25 July 2023	0.893	0.805	0.710	0.805	0.807
Kyselé kouty, 13 July 2023	0.924	0.931	0.930	0.924	0.918
Kyselé kouty, 25 July 2023	0.939	0.957	0.950	0.949	0.878
Pančavská louka, 14 July 2023	0.796	0.816	0.831	0.832	0.773
Pančavská louka, 26 July 2023	0.776	0.798	0.806	0.795	0.747
Pančavská louka, 27 July 2023	0.753	0.701	0.769	0.758	0.694

Among all three areas of interest, all classifiers achieved the best accuracy at Kyselé kouty, slightly worse accuracy was measured at Hraniční louka, and the worst at Pančavská louka. There were also differences in accuracies across the orthomosaics at different times. The RF classifier achieved the best accuracies at Hraniční louka, while the SVM performed better at the two remaining areas of interest.

4.2 Evaluation of usefulness of the non-spectral features

The influence of various features on the accuracy of classifications was then investigated. In all the tests, cross-validation was always used, and the Random Forest classifier with hyperparameters set as ntree - 500, mtry - square root of number of features, max depth - None was employed. The mentioned results are the average of all cross-validations. The RF classifier with these default parameters was selected because the literature suggests that it should achieve good accuracy even in this typical default setting

(Probst et al., 2019). This approach of the feature testing was chosen because it was necessary to decide whether to use these features before proceeding with the actual model tuning and evaluation of its accuracy.

Before the tuning of the classifier, it was first tested whether the canopy height feature would increase the accuracy of the classification. An increase in the F-1 score occurred in all cases, and its level is summarized in the table 8.

Table 8: Improvements in F-1 score when implementing canopy height feature

Area of interest, date	Overall F-1 score increase	The greatest increase in F-1 score for individual classes
Hraniční louka, 13 July 2023	0.108	*Picea abies* (0.188), water (0.128), *Carex limosa* (0.103), Pinus mugo (0.075)
Hraniční louka, 25 July 2023	0.071	*Molinia caerulea* (0.335), *Picea abies* (0.136), Pinus mugo (0.093), *Sphagnum sp.* (0.080)
Kyselé kouty, 13 July 2023	0.069	*Picea abies* (0.333), bare ground (0.268), *Eriophorum vaginatum* (0.093)
Kyselé kouty, 25 July 2023	0.056	*Picea abies* (0.260), *Carex sp.* (0.229), *Nardus stricta* (0.157), *Juncus sp.* (0.116)
Pančavská louka, 14 July 2023	0.072	*Carex rostrata* (0.166), *Nardus stricta* (0.160), *Picea abies* (0.158), *Pinus mugo* (0.110), *Calluna vulgaris* (0.106)
Pančavská louka, 26 July 2023	0.113	*Picea abies* (0.224), *Pinus mugo* (0.195), *Carex rostrata* (0.147), *Carex limosa* (0.138)
Pančavská louka, 27 July 2023	0.097	*Pinus mugo* (0.191), *Picea abies* (0.166), *Calluna vulgaris* (0.156), *Vaccinium myrtillus* (0.094)

The testing results show that accuracy most frequently improved for the conifers *Picea abies* and *Pinus mugo*, although the greatest increase in accuracy was observed in the class *Molinia caerulea* within Hraniční louka. The overall increase in accuracy and the most affected classes significantly depended on the specific orthomosaic.

Before the tuning of the classifier, it was also tested whether GLCM texture features would increase the accuracy of the classification. In this case, only those texture features that remained after the reduction of correlated data, described in 3.7, were used. An increase in accuracy occurred in all cases, and its level is summarized in the table 9 on the following page.

Table 9: Improvements in F-1 score when implementing texture features

Area of interest, date	Overall F-1 score increase	The greatest increase in F-1 score for individual classes
Hraniční louka, 13 July 2023	0.025	*Sphagnum sp.* (0.147), *Trichophorum cespitosum* (0.129), *Pinus mugo* (0.064)
Hraniční louka, 25 July 2023	0.073	*Carex limosa* (0.299), *Carex rostrata* (0.161), *Trichophorum cespitosum* (0.103), *Sphagnum species* (0.087)
Kyselé kouty, 13 July 2023	0.025	Dry vegetation (0.645), *Carex sp.* (0.311)
Kyselé kouty, 25 July 2023	0.040	*Carex sp.* (0.161), *Eriophorum vaginatum* (0.147), bare ground (0.066)
Pančavská louka, 14 July 2023	0.046	*Pinus mugo* (0.138), *Picea abies* (0.136), *Eriophorum angustifolium* (0.053)
Pančavská louka, 26 July 2023	0.038	*Vaccinium myrtillus* (0.115), *Eriophorum angustifolium* (0.053), *Picea abies* (0.092)
Pančavská louka, 27 July 2023	0.043	*Vaccinium myrtillus* (0.094), *Eriophorum angustifolium* (0.048)

The increase in accuracy after including the texture feature was smaller than that those after including canopy height, but the classification accuracy of some classes still improved significantly. The most substantial increase in classification accuracy occurred for Hraniční louka on July 25th.
It was also explored whether the accuracy of classifications changes when including all 189 bands and after their reduction (as described in 3.7). In all cases, the classification's overall F-1 score did not change by more than 0.005, indicating that important information was preserved even after deleting correlated bands. The final numbers of bands are summarized in the table 10.

Table 10: Number of bands after reduction of correlated bands

Area of interest, date	Final number of bands
Hraniční louka, 13 July 2023	50
Hraniční louka, 25 July 2023	65
Kyselé kouty, 13 July 2023	42
Kyselé kouty, 25 July 2023	44
Pančavská louka, 14 July 2023	38
Pančavská louka, 26 July 2023	38
Pančavská louka, 27 July 2023	38

As seen from the table 10, most bands were always removed from the dataset. The number of correlated bands varied significantly for each dataset, but in most cases, it was similar or the same for each

orthomosaic of the same area of interest. This was not the case only for Hraniční louka, where there was a difference of 15 bands.

4.3 Separability analysis

In the first step, separability was calculated for all pairs of classes. The three least separable pairs of classes and the average separability of all pairs for each orthomosaic are listed in the table 11. Complete J-M index matrices can be seen in the appendix C.

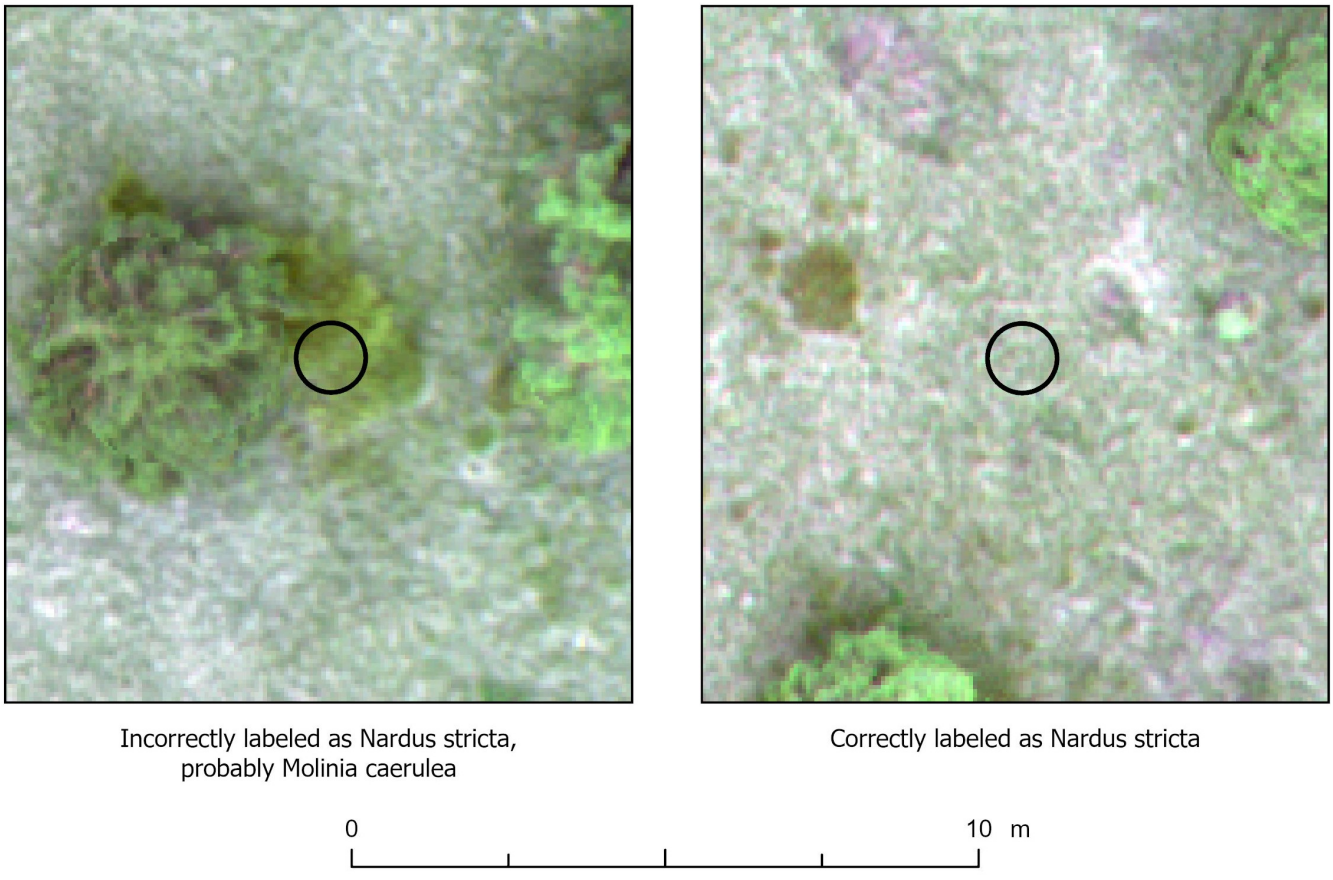
Table 11: Summary of separability analysis

Area of interest, date	Average separability	Pairs of classes with low J-M index
Hraniční louka, 13 July 2023	1.98	*Sphagnum sp.* and *Trichophorum cespitosum* (1.66) *Sphagnum sp.* and *Carex limosa* (1.82) *Sphagnum sp.* and *Picea abies* (1.84)
Hraniční louka, 25 July 2023	1.99	*Sphagnum sp.* and *Trichophorum cespitosum* (1.74) *Sphagnum sp.* and *Picea abies* (1.8) *Sphagnum sp.* and *Carex limosa* (1.86)
Kyselé kouty, 13 July 2023	1.92	*Eriophorum vaginatum* and *Juncus sp.* (1.28) *Vaccinium myrtillus* and unidentified moss (1.55) *Eriophorum vaginatum* and *Carex sp.* (1.56)
Kyselé kouty, 25 July 2023	1.95	*Eriophorum vaginatum* and *Juncus sp.* (1.41) *Juncus sp.* and *Nardus stricta* (1.58) *Eriophorum vaginatum* and *Deschampsia cespitosa* (1.75)
Pančavská louka, 14 July 2023	1.92	*Carex limosa* and *Eriophorum angustifolium* (1.06) *Vaccinium myrtillus* and *Calluna vulgaris* (1.27) *Sphagnum sp.* and *Trichophorum cespitosum* (1.33)
Pančavská louka, 26 July 2023	1.92	*Carex limosa* and *Eriophorum angustifolium* (0.99) *Sphagnum sp.* and *Trichophorum cespitosum* (1.22) *Sphagnum sp.* and *Calluna vulgaris* (1.29)
Pančavská louka, 27 July 2023	1.88	*Carex limosa* and *Eriophorum angustifolium* (0.91) *Nardus stricta* and *Trichophorum cespitosum* (1.23) *Sphagnum sp.* and *Trichophorum cespitosum* (1.26)

The results indicate that the separability of most classes is high, but there are several poorly separable pairs of classes at Kyselé kouty and Pančavská louka. The best separability is seen in the data from Hraniční louka, and the worst is at Pančavská louka, which also corresponds to poorer classification results, especially regarding the orthomosaic from July 27th.
The next step was to test the FPD's (Faulty polygon detector) performance. It was tested on the ground truth data of Hraniční louka, which was otherwise proved to be very well separable (most J-M

indices over 1.9). In a total of 200 tests, the class of one of the polygons was randomly changed each time and it was tested whether it would be detected by the FPD. It was measured that in 40.7% of cases, the faulty polygon was correctly identified, but in roughly half of these cases, FPD identified more than one polygon as faulty. In 59.2% of cases, the change was not detected at all.
After this test, the FPD was applied to each of the orthomosaics. Neither Hraniční louka nor Kyselé kouty showed any results, but when the FPD was applied to the ground truth data of Pančavská louka, two demonstrably erroneous polygons were found. The first was apparently incorrectly recorded in the field, and it was a polygon marked as *Nardus stricta* even though the orthomosaic clearly shows a significant color difference between other *Nardus stricta* stands and this location (see figure 6). In the second case, it was a water body polygon that was too large and encroached on dry land. After removing these polygons, the F-1 score for the *Nardus stricta* class increased by 0.037, but the F-1 score for water did not increase (it decreased by 0.001). The first was completely removed from the ground truth data, and the second was slightly shifted so that it did not encroach on the lake shore.

Figure 6: Example of the incorrectly labeled polygon on Pančavská louka.



4.4 Hyperparameter tuning

One of the first important questions was whether the two used tuning algorithms - gridsearch and optimization with a genetic algorithm - identified similar combinations of hyperparameters as the best. Both approaches were used for tuning the model on each orthomosaic. It was found that the genetic algorithm, especially with a higher number of hyperparameters, can be significantly faster in achieving results, but gridsearch provides an opportunity to delve deeper and determine which parameters are truly important for the accuracy of the model. The table 12 on the next page lists the best parameters of individual models, derived from gridsearch and then from optimization with a genetic algorithm (GS = gridsearch, GO = genetic optimization).

Table 12: Best hyperparameters according to gridsearch and genetic optimization (GS = gridsearch, GO = genetic optimization)

Area of interest, date	Method	RF	Linear SVM	Polynomial SVM	SVM with RBF	SVM with sigmoid
Hraniční louka, 13 July 2023	GS	ntree = 300 mtry = 6-8 or 13-16 depth = 15	C = 1	C = 100 Degree = 2 Gamma = 0.004	C = 1000 Gamma = 1e-07	C = 1000 Gamma = 1e-07
	GO	ntree = 464 mtry = 15 depth = 18	C = 2.57	C = 433.09 Degree = 3 Gamma = 0.0082	C = 945.37 Gamma = 2.86e-07	C = 394.22 Gamma = 3.22e-09
Hraniční louka, 25 July 2023	GS	ntree = 200 mtry = 16 depth = 16	C = 0.01	C = 0.001 Degree = 3 Gamma = 0.024	C = 1000 Gamma = 0.004	C = 1 Gamma = 0.004
	GO	ntree = 138 mtry = 18 depth = 17	C = 9.94	C = 1.64e-07 Degree = 2 Gamma = 0.034	C = 593.02 Gamma = 8.84e-07	C = 4.928 Gamma = 3.53e-07
Kyselé kouty, 13 July 2023	GS	ntree = 350 mtry = 5 depth = 18	0.001	C = 0.1 Degree = 2 Gamma = 0.0001	C = 1000 Gamma = 1e-06	C = 1000 Gamma = 1e-07
	GO	ntree = 885 mtry = 6 depth = 19	C = 0.00024	C = 0.15 Degree = 2 Gamma = 4.31e-05	C = 10.97 Gamma = 5.38e-06	C = 2.95 Gamma = 2.82e-07

Table 12: Best hyperparameters according to gridsearch and genetic optimization (GS = gridsearch, GO = genetic optimization) - continued

Area of interest, date	Method	RF	Linear SVM	Polynomial SVM	SVM with RBF	SVM with sigmoid
Kyselé kouty, 25 July 2023	GS	ntree = 250 mtry = 7 depth = 16	C = 0.001	C = 0.001 Degree = 2 Gamma = 0.0001	C = 10 Gamma = 1 e-05	C = 1000 Gamma = 1e-07
	GO	ntree = 970 mtry = 8 depth = 19	C = 0.00012	C = 0.0199 Degree = 2 Gamma = 0.00029	C = 235.06 Gamma = 1.899e-06	C = 107.20 Gamma = 5.07e-09
Pančavská louka, 14 July 2023	GS	ntree = 350 mtry = 7 depth = 18	C = 0.001	C = 0.01 Degree = 3 Gamma = 1e-04	C = 1 Gamma = 1e-04	C = 1 Gamma = 1e-07
	GO	ntree = 984 mtry = 8 depth = 20	C = 0.00036	C = 0.024 Degree = 3 Gamma = 4.36e-05	C = 436.3 Gamma = 2.9e-06	C = 218.2 Gamma = 9.6e-08
Pančavská louka, 26 July 2023	GS	ntree = 200 mtry = 7 depth = 18	C = 0.001	C = 1 Degree = 3 Gamma = 1e-04	C = 1 Gamma = 1e-04	C = 10 Gamma = 1e-07
	GO	ntree = 51 mtry = 8 depth = 19	C = 0.0001	C = 14.7 Degree = 3 Gamma = 1.19e-05	C = 72.02 Gamma = 4.51e-07	C = 10.23 Gamma = 2.35e-07
Pančavská louka, 27 July 2023	GS	ntree = 200 mtry = 8 depth = 18	C = 0.001	C = 0.1 Degree = 3 Gamma = 1e-04	C = 1 Gamma = 1e-05	C = 10 Gamma = 1e-07
	GO	ntree = 830 mtry = 9 depth = 20	C = 2.07e-05	C = 0.1 Degree = 3 Gamma = 6.46e-05	C = 98.2 Gamma = 1.40e-06	C = 912.7 Gamma = 1.40e-08

The final best parameters obtained by both methods do not differ significantly in most cases. However, in some cases (especially ntree in RF and parameter C in SVM), significant differences can be observed. In all these cases, however, both grid search and comparison of results have shown that these hyperparameters do not significantly change the overall F-1 score of the model, and as described below, no direct dependence on the value of the parameter was found. Therefore, the final accuracies after cross-validation are nearly identical. In the case of SVM classifiers, the average overall F-1 score was always lower when using hyperparameters derived from grid search, but with a difference of always less than 0.01. For the RF algorithm, the difference in accuracy between the two models was always within 0.001. In the end, the hyperparameters achieved through genetic optimization were used for the final classification, although the opposite decision would likely not have led to a significantly different result. The results from the gridsearch reveal the following characteristics for individual hyperparameters:

- Random Forest - Each of these parameters significantly influences the duration of training and prediction:
  - ntree - Did not affect the final accuracy of classification once a certain value was exceeded (in this case, around 200).
  - mtry - The best mtry depended on the number of features. The best results were usually achieved by a value similar to square root of number of features or around double that, which aligns with the literature (Simon et al., 2023).
  - max depth - Mainly depends on the complexity of the dataset (Probst et al., 2019).
- SVM - Parameters seem to be interdependent, so tuning all of them is necessary for good classification results:
  - C - The appropriate setting of this parameter depends on the kernel function. In the case of a linear function, accuracy does not significantly change, and no simple dependency could be observed with other functions, although changes in performance were sometimes observed.
  - degree - No clear correlation with classification accuracy was observed.
  - gamma - Only set for non-linear SVM, and the optimal value for each function was usually different. It's worth mentioning that it usually had to be very low (10e-5 or lower) to achieve good accuracy with the RBF and sigmoid functions.

The results from the gridsearch also indicate that Random Forest and linear SVM almost always achieve similar accuracy, while tuning the SVM algorithm with a different kernel function requires much more testing and tuning. It was also observed that in some combinations of SVM hyperparameters, the computation runs significantly longer. Since the computation time was not directly measured, data that would help identify these combinations are missing. In some cases, it appeared that a high value of the regularization parameter C could extend the training time, but this dependence does not always hold. This phenomenon considerably prolonged both the gridsearch and the optimization process.

4.5 Results of classifications after data reduction

As mentioned in the 3.9, this testing was conducted in two different settings. Initially, some polygons were completely removed from the ground truth data. The results for this analysis are presented in figures 7, 8 and 9. These figures contain validations from all orthomosaics of the given area of interest.

Figure 7: Simple data reduction on Hraniční louka

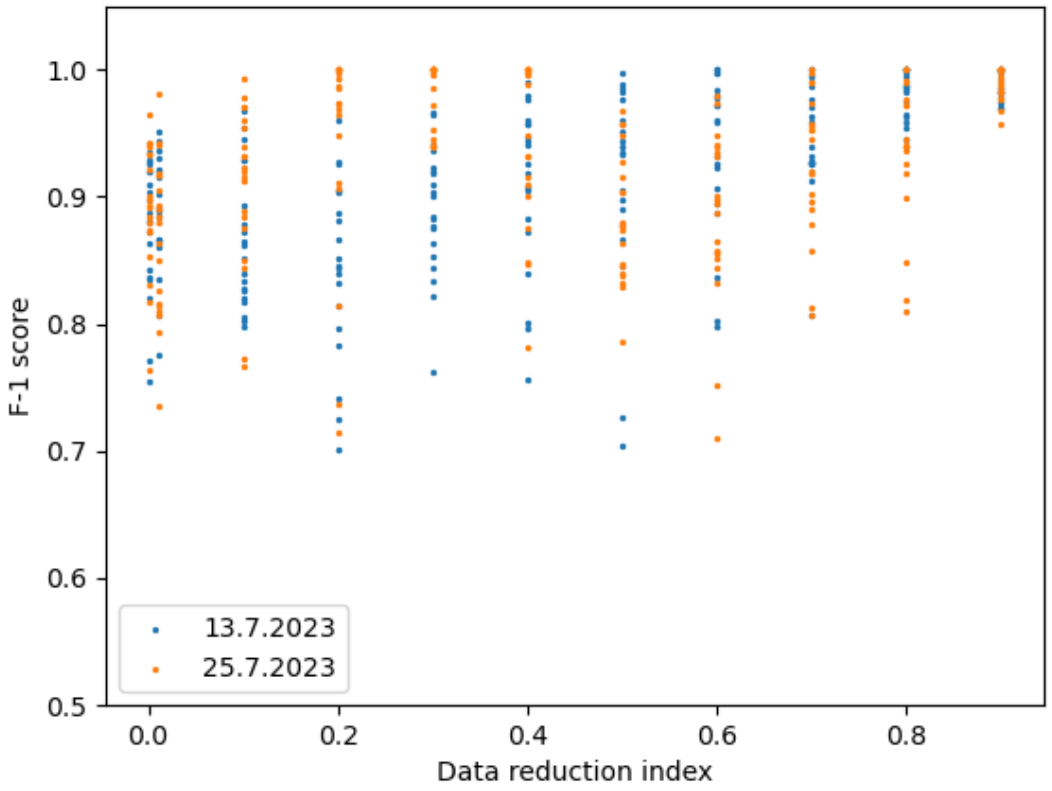


Figure 8: Simple data reduction on Kyselé kouty

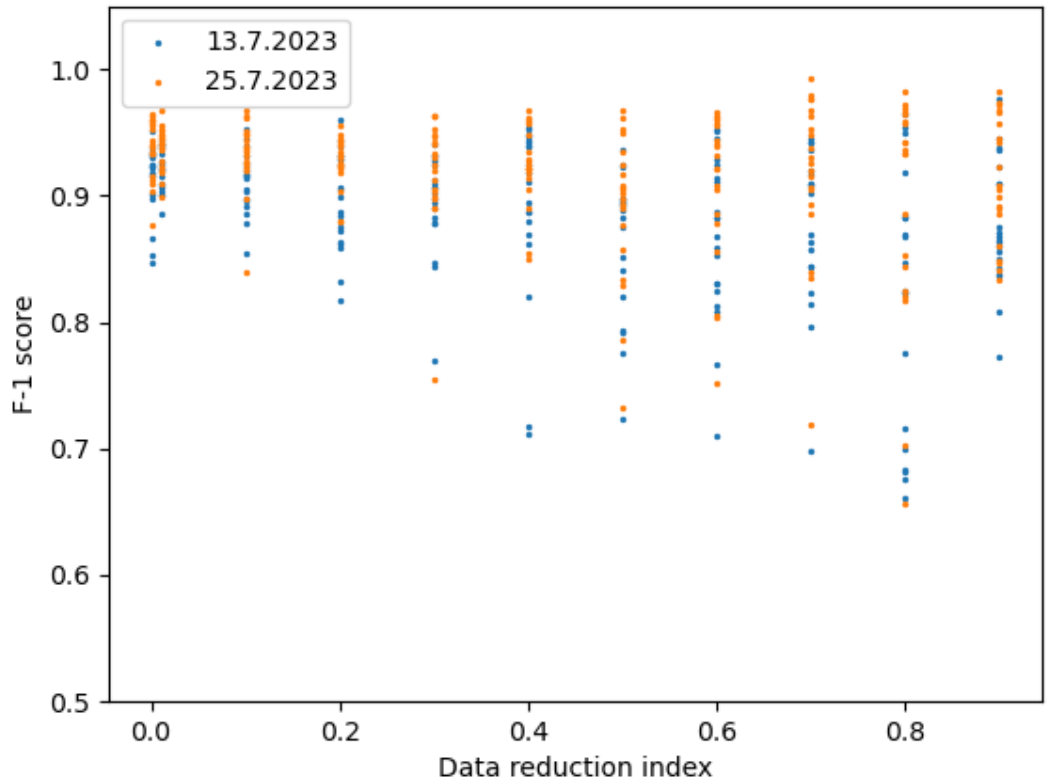
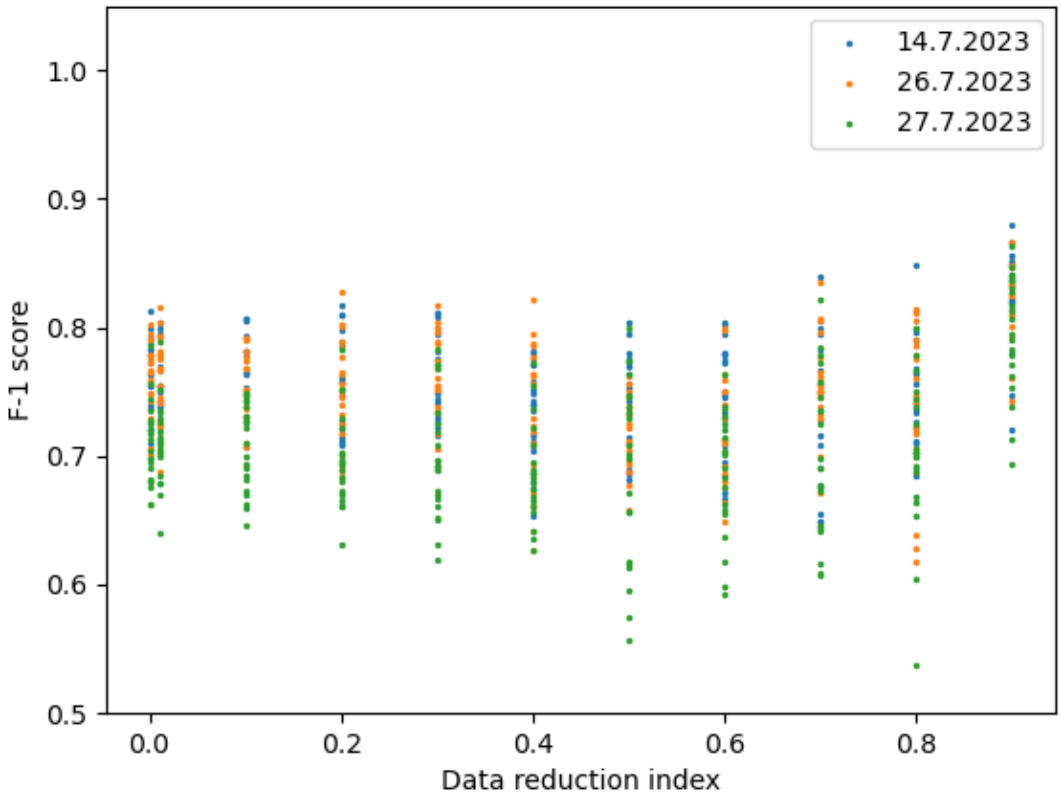


Figure 9: Simple data reduction on Pančavská louka



From figures 7, 8 and 9, it is evident that as data were removed, the variance in accuracies increased, and accuracies generally improved as well. The latter phenomenon is likely caused by the division of individual polygons to train and validation parts, which is automatically performed for classes with a small amount of training and validation data available.
Furthermore, it was tested what the accuracy of classifications would be when the removed polygons were added to the validation data. The results for this second analysis are presented in figures 10, 11 and 12. These graphs contain validations from all orthophotos of the given area of interest.

Figure 10: Data reduction on Hraniční louka with using the rest of data for validation

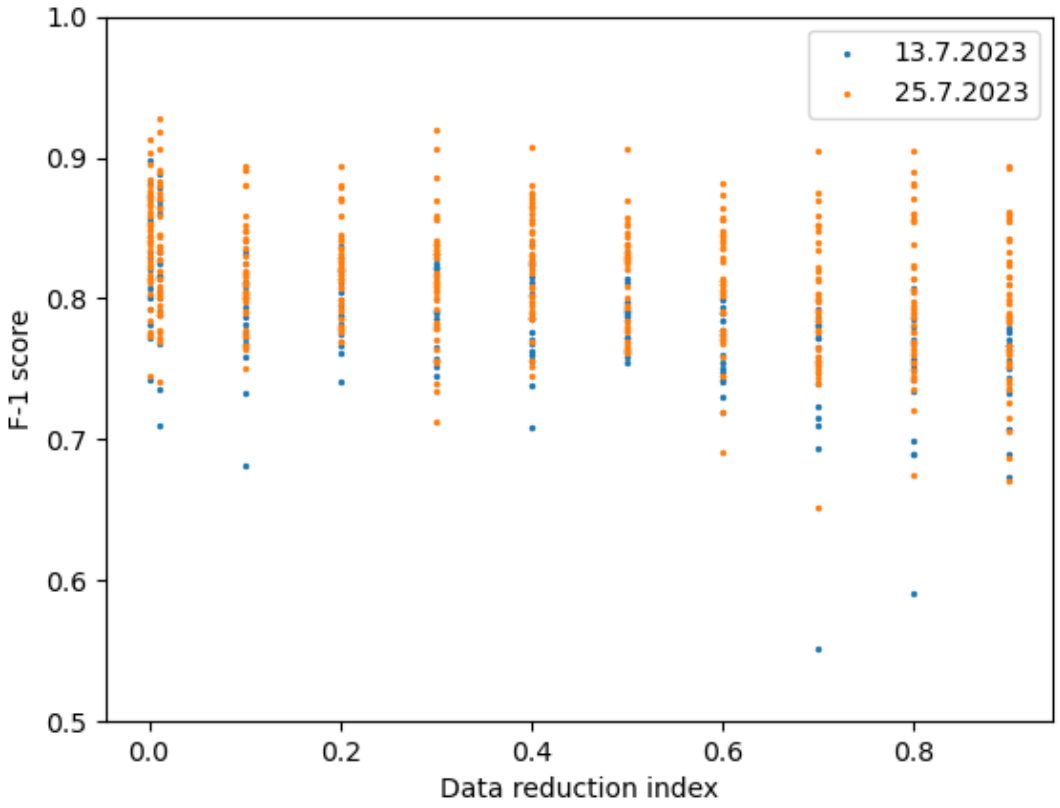


Figure 11: Data reduction on Kyselé kouty with using the rest of data for validation

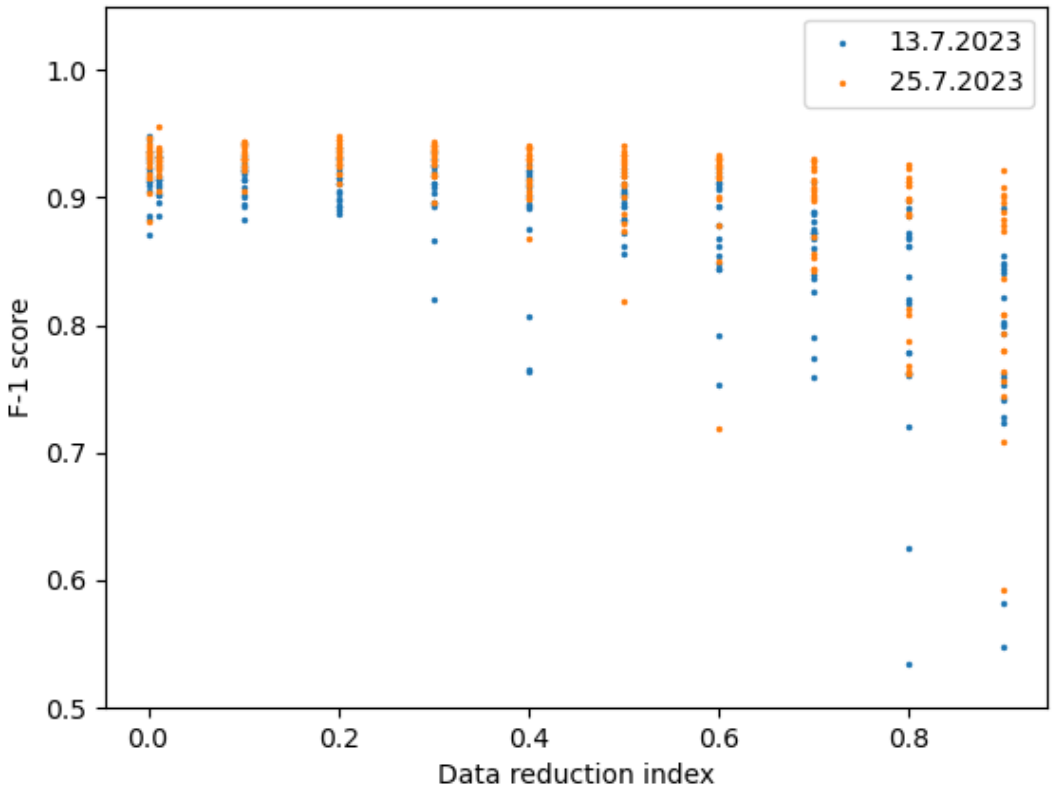
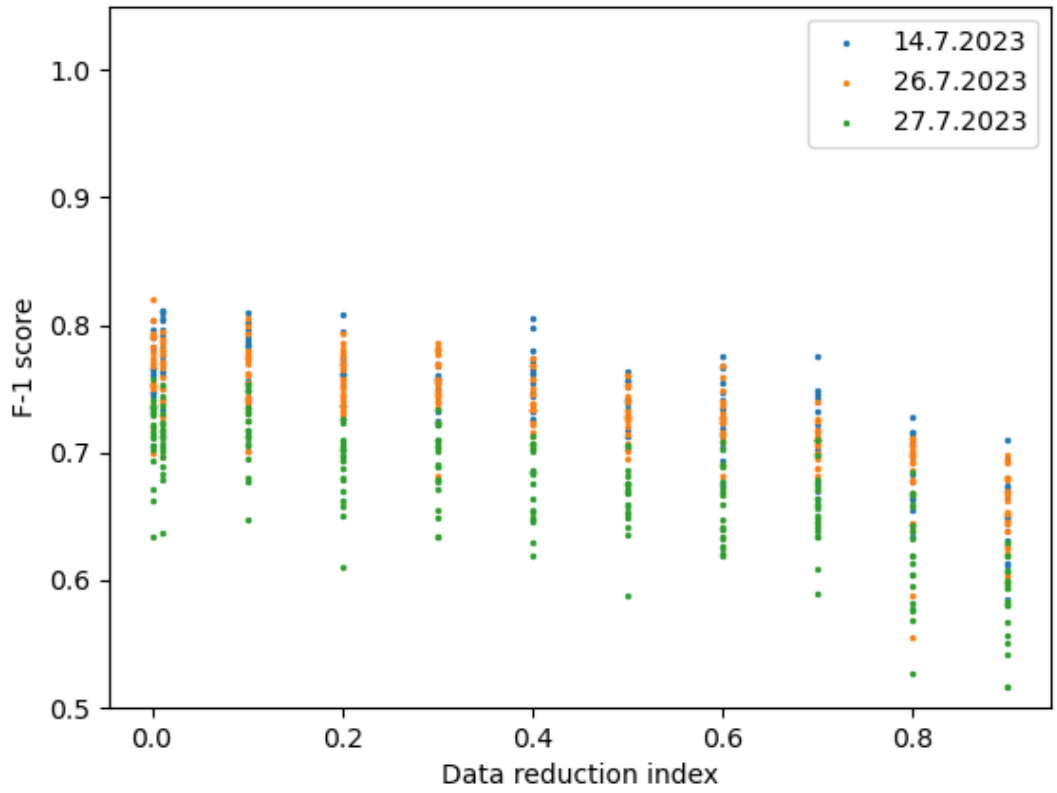


Figure 12: Data reduction on Pančavská louka with using the rest of data for validation



In figures 10, 11 and 12, it is apparent that, besides a slight increase in the variance of accuracies, accuracies also decreased because there was no division of individual polygons, making the validation more precise.

4.6 Overlays and final vegetation cover maps

Classifications that achieved overall F-1 score close (within a 0.05 difference) to the best classification on a given orthomosaic were included in the overlay. The table 13 shows the classifications which were included into the calculation of the overlays.

Table 13: Classifications used for overlay analysis (bolded if used)

Area of interest, date	RF	Linear SVM	Polynomial SVM	SVM with RBF	SVM with sigmoid
Hraniční louka, 13 July 2023	**0.899**	**0.875**	**0.879**	**0.895**	0.779
Hraniční louka, 25 July 2023	**0.893**	0.805	0.710	0.805	0.807
Kyselé kouty, 13 July 2023	**0.924**	**0.931**	**0.930**	**0.924**	**0.918**
Kyselé kouty, 25 July 2023	**0.939**	**0.957**	**0.950**	**0.949**	0.878
Pančavská louka, 14 July 2023	**0.796**	**0.816**	**0.831**	**0.832**	0.773
Pančavská louka, 26 July 2023	**0.776**	**0.798**	**0.806**	**0.795**	0.747
Pančavská louka, 27 July 2023	**0.753**	0.701	**0.769**	**0.758**	0.694

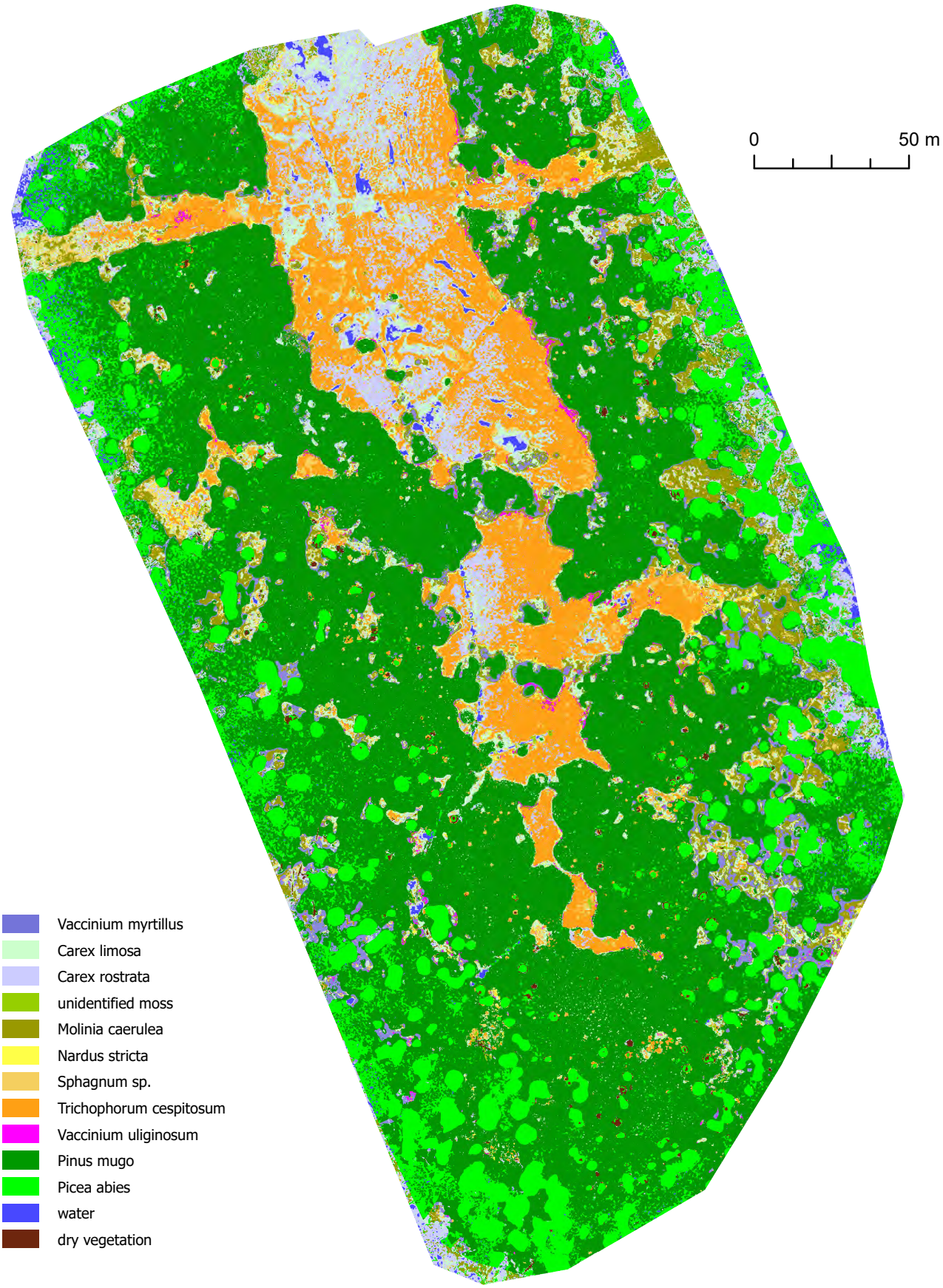
As the table 13 shows, for the second orthomosaic of Hraniční louka, the differences in accuracies were too large, and overlaying was not performed.
The most accurate final maps for each area of interest are displayed in figures 13, 14 and 15. Other important maps are displayed in appendix D. The table 14 shows measures of accuracy and consistency for all overlays. The degree of agreement of classifications is measured by the proportions of the orthomosaic on which an exact number of classifications agreed during the "voting" process for the final class.

Table 14: Results of overaly analysis

Area of interest, date	Best single classifier F-1 score	Overlay F-1 score	Proportion of the orthomosaic area by the number of agreeing classifications				
			0	2	3	4	5
Hraniční louka, 13 July 2023	89.9	88.8	0.20%	6.2%	16.5%	77.1%	
Kyselé kouty, 13 July 2023	93.1	91.2	0.01%	1.3%	4.5%	4.8%	89.4%
Kyselé kouty, 25 July 2023	95.7	94.3	0.10%	4.0%	7.4%	88.5%	
Pančavská louka, 14 July 2023	83.1	83.5	0.12%	3.4%	6.5%	90.0%	
Pančavská louka, 26 July 2023	80.6	80.7	0.24%	5.2%	7.6%	87.0%	
Pančavská louka, 27 July 2023	76.9	75.5	2.5%	11.2%	86.3%		

figure 13: HRANIČNÍ LOUKA

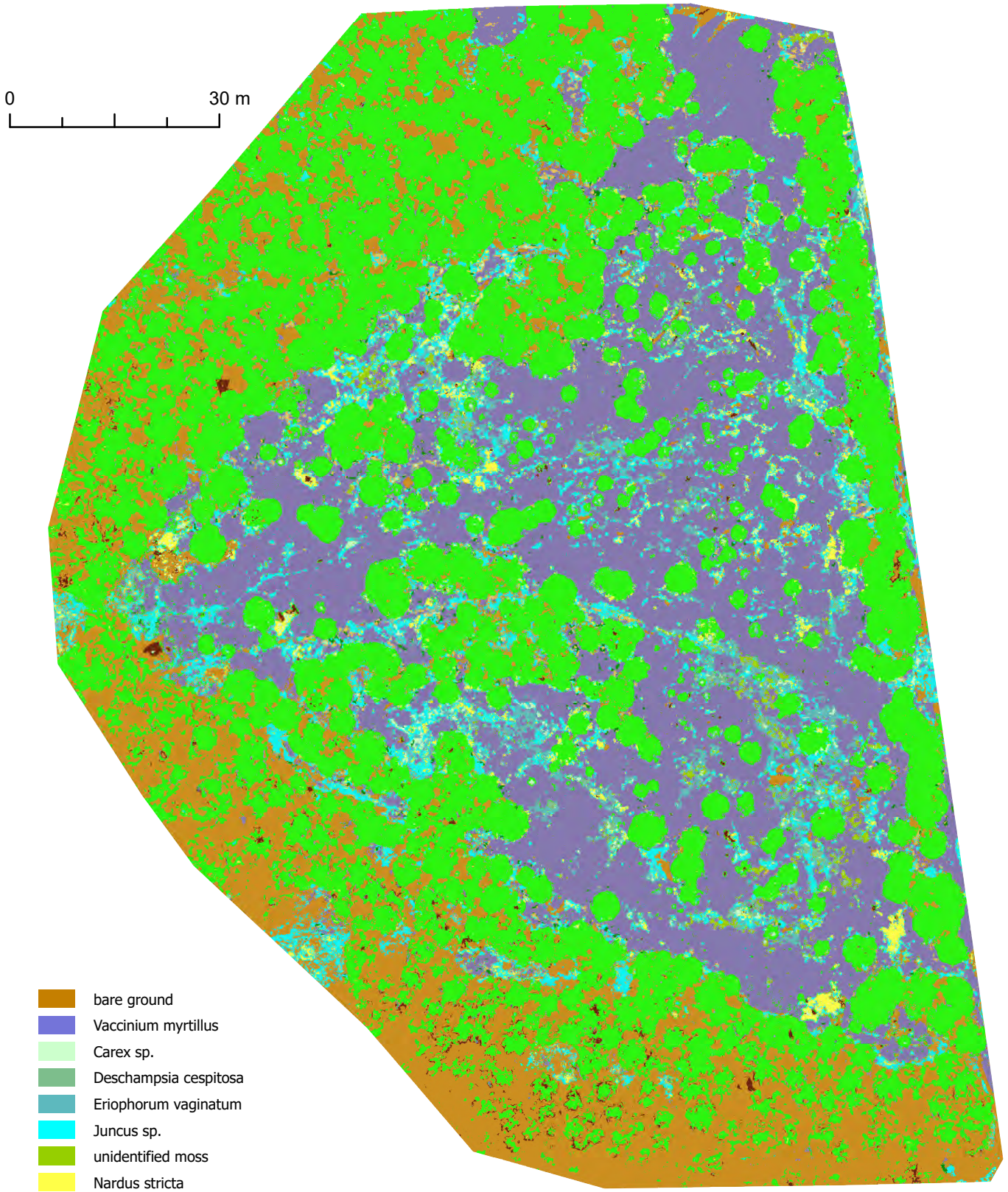
vegetation cover of peat bog in the Krkonoše NP (July 13th, 2023)
classified by Random forest, F-1 score 0.899



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

figure 14: KYSELÉ KOUTY

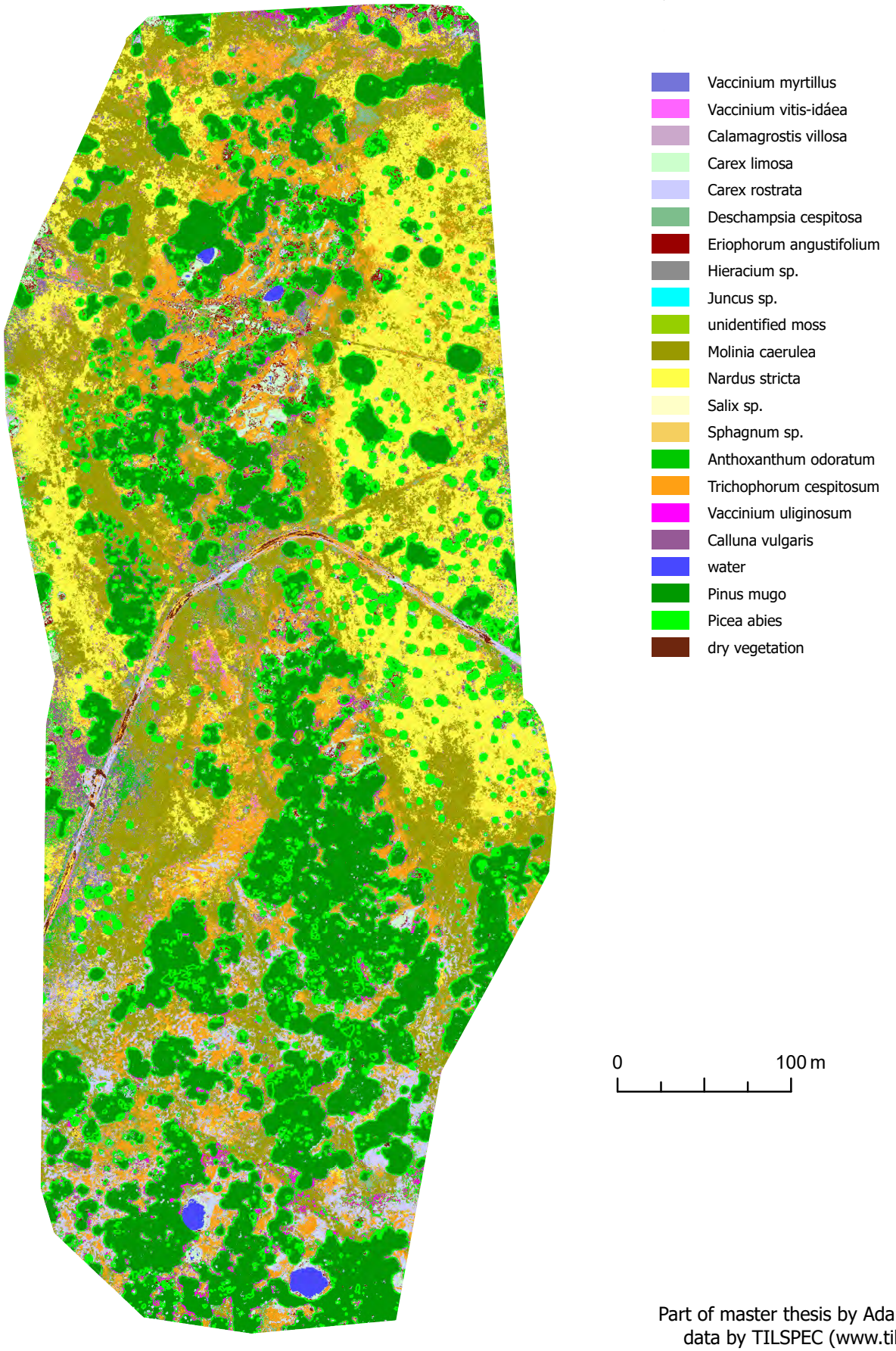
vegetation cover of peat bog in the Krkonoše NP (July 25th, 2023)
classified by Support vector machine, F-1 score 0.957



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

figure 15: PANČAVSKÁ LOUKA

vegetation cover of peat bog in the Krkonoše NP (July 14th, 2023)
ensemble of Random forest and Support vector machine classifications, F-1 score 0.835



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

5 Discussion

The first goal of this thesis was to classify each of the peat bogs with a final overall F-1 score of at least 0.9. This accuracy was only surpassed on one of the three studied peat bogs (Kyselé kouty) and on two of the seven classified orthomosaics (both classifications of Kyselé kouty). Despite not achieving complete success, this result is still very good because comparable studies generally did not achieve better results (Beyer et al., 2019; Kupková et al., 2020), and 0.9 is a very ambitious goal. An overall F-1 score over 0.9 was achieved only at Kyselé kouty, but the best classifications at Hraniční louka also came very close to this goal. At the larger and more complex Pančavská louka, the best result achieved was only 0.832. The best overall F-1 score at Hraniční louka was 0.899, and at Kyselé kouty, it was 0.957. At Hraniční louka, the RF classifier achieved better accuracy, but on the two remaining areas of interest, it was the SVM classifier that performed better. Overall accuracy of the classifications was also high (see appendix B). However, despite the rather good overall classification results, the analysis showed that to achieve consistently satisfactory results for individual classes, higher amount of ground truth data should be collected for many of the classified species.

The low amount of the ground truth data (number of polygons) for several categories caused some problems in the analysis. Despite great efforts and several days spent in the field, it was not possible to collect enough ground truth data. The main problem is not just in the small number of data entering the classifier training, but also in the small representativeness of the data. Spectral information for each polygon can vary due to variability within the species and different lighting conditions at various places on the orthomosaic. Ideally, all spectral diversity for a given class should be included in the training, validation, and testing data, which cannot be guaranteed in the case of a small amount of data. This lack of data can be theoretically resolved by using part of each polygon for training the classifier and other part for validation, but in such a case (as will be explained in more detail below), the validation loses its informative power due to the spatial autocorrelation of information from training and validation data. If it is considered important not to divide individual polygons into training and validation parts, each class should contain enough polygons/pixels to safely perform cross-validation. However, for many classes at Hraniční louka and Kyselé kouty, there were 4 or fewer polygons, making it impossible to include more than one of them in the validation or testing data. This leads to an uneven results of classification accuracy evaluation because the representativeness is not sufficient in either the training or the validation data. For classes where there was more data, distributed across the entire area of interest, the accuracies of individual validations were relatively consistent.

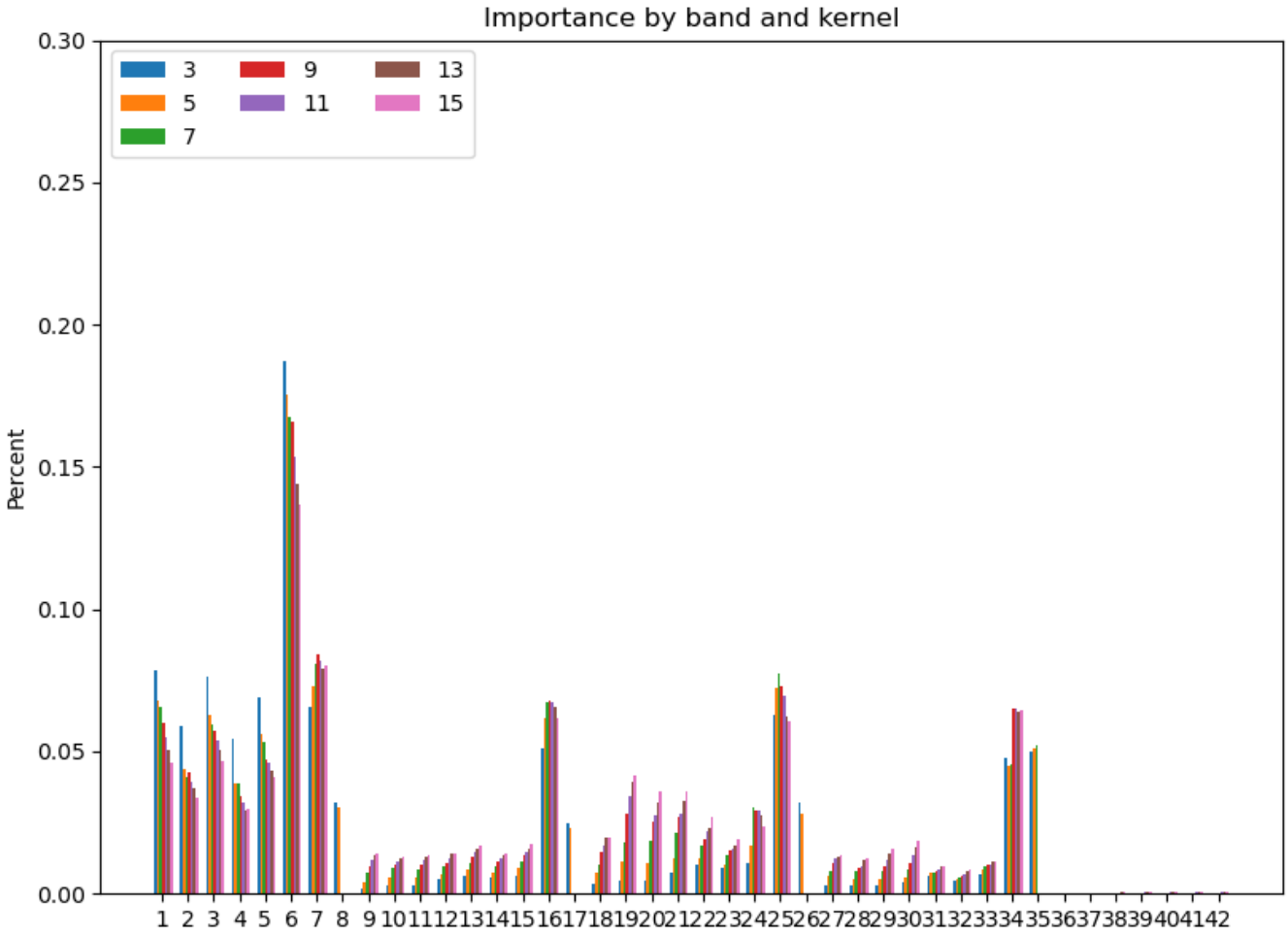
Collecting GPS data in the field is one of the most challenging parts of the entire research. Besides the poor accessibility of the multiple areas of interest and the physical demands of measuring with heavy equipment, in some cases, it is very difficult to find a piece of homogeneous vegetation suitable for inclusion in the ground truth data on a peat bog of such large size. Although it can be very challenging, it turns out that a truly good ground truth dataset must consist of a larger number of representatively distributed measurements. For representative training and validation, it would be good to have enough polygons for all classes, in my opinion ideally at least 6-10 for each and also reliable number of pixels inside these polygons.

The low number of field data for some categories is also the main reason why the stratified sampling (based on the real share of the categories in the field) could not be fully employed. As the literature (Mohri et al., 2018) indicates, the amount of data entering the training phase of the classifier in individual classes can influence the model and its performance on the resulting data. Therefore, it can be considered a great success that, despite the lower amount and not stratified distribution of training/validation data, the classifiers demonstrated such good accuracy on independent data from the test set. In the next work over this topic (during PhD project) further data will be collected in and it will be tested if the theoretically sufficient and stratified training/validation dataset can bring further classification accuracy improvements.

As already noted, non-spectral features proved to be very useful across all areas of interest, which is consistent with the available literature (J. Cao et al., 2018; Mohammadpour et al., 2022; Zhang et al., 2016). The canopy height feature increased the classification accuracy of the *Picea abies* class, which is the only tree species present in the peat bogs, on all three areas of interest. On Hraniční louka and Pančavská louka, this feature also improved the classification accuracy of *Pinus mugo* (F-1 score increased by 0.075 - 0.191), which also forms stands significantly protruding above the terrain. On Hraniční louka, the accuracy of water classification also increased (by 0.128) and *Carex limosa* (by 0.103), which predominantly grows in shallow ponds, suggesting that the photogrammetric DSM is capable of recognizing even the height difference of *Carex* stalks above the flat water surface. The same phenomenon occurred on Pančavská louka for *Carex rostrata* and *Carex limosa*. In other cases, the canopy height feature contributed to better classification of tall grasses like *Molinia caerulea* (Hraniční louka) or *Eriophorum vaginatum* (Kyselé kouty). On Pančavská louka, the classification accuracy of *Vaccinium myrtillus* which grows across the site in smaller clumps protruding above the surrounding terrain also improved. These results show that the canopy height feature is a very important input for peat bog classification and should not be overlooked in further research.

Textural features increased classification accuracy less, but also significantly. Often, this increase in accuracy occurred in classes whose classification was insufficiently accurate, for example, *Sphagnum sp.* in the case of Hraniční louka or *Eriophorum angustifolium* in the case of Pančavská louka. For instance, the significant increase in classification accuracy of *Trichophorum cespitosum* (F-1 score higher by 0.103 and 0.129) on both orthomosaics of Hraniční louka, where this herb is a very important species covering most of the open peat bog area, is very noteworthy. However, regarding textural features, it remains to be questioned which of them truly make sense to include in classification. When testing classifications, it was found that when all textural features were included, some almost did not contribute new information (see figure 16 on the next page). This problem was ultimately addressed by simply removing correlated bands, but it certainly would be possible to proceed more conservatively, as the computation of textural features can take a very long time. Literature also suggests (Erdem & Bayrak, 2023) that there are other types of textural features than those derived from GLCM, and in some cases, they may increase classification accuracy even more.

Figure 16: Comparion of the importances of texture bands. Bands 1-5 are spectral, band 6 is canopy height. Derived from RF model from Hraniční louka



Another significant success was the analysis of separability. The classical method of calculating separability indices (Schmidt & Skidmore, 2003) has been used for a long time, but utilizing separability to identify faulty ground truth polygons is a new concept that addresses a problem not commonly tackled. However, the testing results show that (at least with the used settings) the FPD (Faulty Polygon Detector) is, despite its success on Pančavská louka, inaccurate in general, as it correctly identified problematic polygons only in 40% of cases. The inaccuracy of the FPD during testing could be due to the fact that not every pair of classes had high separability from the beginning, and not every class had enough ground truth data. Therefore, the FPD primarily detected faulty polygons of *Nardus stricta* (a common class for which a large amount of data was collected) and water (a class with spectral features very easily distinguished from surrounding vegetation). It is also possible that better results could have been achieved with a different setting of input parameters. In this thesis, the availability of several orthomosaics, which could be analyzed separately (FPD achieved the same results on all of them), as well as field experience, without which it would not be possible to determine from the

orthomosaic what species is growing at a given location, were utilized in evaluating the algorithm's outputs. Ideally the results of the FPD should be controlled in the field to be sure about the identification of the class of the problematic polygon.
Another issue, encountered in the testing was that the FPD has a low specificity and sometimes marks not only the truly faulty polygon but also several others. This phenomenon likely arises because any amount of pixels removed from an area not easily separable in feature space increases the J-M distance between these two classes. In this case, it may help that the outcome of the FPD also includes the rate of improvement in the resulting J-M index when the potentially faulty polygons are removed. The results from Pančavská louka certainly point to the potential use of FPD or a similar algorithm and its necessity, because even with the automation of much of the classification process, human error can still occur.
As for the used classifiers, both have several advantages and disadvantages for classifying peat bogs. The RF classifier trains and predicts very quickly, and its tuning only slightly improves accuracy compared to the default settings. On the other hand, the accuracy of SVM is strongly dependent on input parameters, and classification with this algorithm also takes much longer. Table 15 shows the time of the classification for each orthomosaic by each method.

Table 15: Computational time requirements [hours : minutes : seconds]

Area of interest, date	RF	Linear SVM	Polynomial SVM	SVM with RBF	SVM with sigmoid
Hraniční louka, 13 July 2023	00:11:27	00:23:19	00:36:51	01:21:39	
Hraniční louka, 25 July 2023	00:11:15				
Kyselé kouty, 13 July 2023	00:10:16	1:36:46	1:30:13	4:23:03	03:55:07
Kyselé kouty, 25 July 2023	00:10:27	1:19:38	0:41:16	1:59:11	
Pančavská louka, 14 July 2023	00:27:56	15:24:04	12:07:58	26:18:48	
Pančavská louka, 26 July 2023	00:29:42	14:57:52	7:43:44	35:37:10	
Pančavská louka, 27 July 2023	00:33:03		11:29:26	39:55:14	

The only advantage of the SVM algorithm is its sometimes slightly better accuracy. However, given such a demanding process of tuning hyperparameters (several different kernels and many parameters) and long classification time, it raises the question of whether it might be better to use more sophisticated algorithms, such as convolutional neural networks (CNNs), in an effort to maximize classification accuracy. Nevertheless, the process of optimization using a genetic algorithm has proven to be a very effective and fast way to achieve results similar to the gridsearch, as shown in table 12 on page 35.
The results also show a large difference in the accuracy of the classification of individual areas of interest. In addition to varying data quality, these differences correspond to the species diversity on the peat bogs (Pančavská louka is the most diverse, while Kyselé kouty is the least) and also to the size of the areas of interest. With the UAV data acquisition, the probability of homogeneous lighting conditions is significantly lower for larger area, which may be one of the reasons why Pančavská louka was

classified with the worst accuracy. These results suggest that it may not be possible to expect similar accuracy across different types of peat bogs. It would also be appropriate to study more the effect of complexity on the classification results. Merely three very distinct peat bogs are too small a sample for a more detailed analysis, as the largest peat bog in this case was also the most complex. Additionally, the comparison of several orthomosaics from similar periods and the same area of interest shows that accuracies can vary significantly during the vegetation season within the same area. Interestingly, it is not easy to observe a clear dependency of the final accuracy on lighting conditions from the results. It seems that orthomosaics taken under cloudy conditions achieve better results, but there are too few orthomosaics available to confirm this hypothesis. Differences may be more probably caused by a slight change in the phase of the vegetation season, which can alter the distinguishability of some classes and perhaps also by different wetness and water content in different days of acquisition (example of Kyselé kouty peat bog).

In this thesis, a multitemporal approach has not been tested. Ideally, for such an approach, larger time intervals between data acquisitions would be needed, but it was also deemed more important to compare the accuracy of individual orthomosaics' classifications. The Krkonoše NP administration would prefer to reduce the costs of monitoring peat bogs as much as possible, and therefore wants to focus more on classification without a multitemporal approach. However, the results of this thesis and relevant literature (Huylenbroeck et al., 2020; Kupková et al., 2023), suggest that multitemporal imaging could provide important information for the classification of peat bogs (vegetation in general) on the species level and should be compared with the classification of a single orthomosaic.

As mentioned in the results, when reducing the number of training data, a gradual decrease in accuracy is evident, but the variance in accuracies during random selection for training or validation does increase only little in the most cases (this doesn't apply to Kyselé kouty). From this, it can be inferred that the ideal amount of training data would be higher (or at least more representatively distributed) than all available data. If the decrease in training data is accompanied by an increase in validation data, the drop in accuracy is evident. In cases simulating a real lack of data, however, the validation accuracy gradually increases over time. This phenomenon is caused by the algorithm partitioning ground truth data into training and validation parts when there are too few polygons for a given class (less than 3), leading to a division that so significantly impacts the results that at some point it completely reverses the validation outcomes, making them entirely unusable. For some classes, this method of training and validation was used even during the main classification analysis because too little ground truth data was available. These classes are summarized in the table 16.

Table 16: Classes with less than 3 polygons of the ground truth data

Area of interest	Classes
Hraniční louka	*Vaccinium myrtillus*, *Nardus stricta*
Kyselé kouty	*Carex sp.*, *Pinus uncinata*,
Pančavská louka	*Callamagrostis villosa*, *Hieracium sp.*, *Juncus sp.*,
	Salix sp., *Anthoxanthum odoratum*

All these classes either have very high or very low accuracies in classifications, depending on the algorithm and the timing of the imagery. The issue is that most of these classes are very rare in the areas of interest, and it was not easy to find additional locations of occurrence. According to experience derived from past research of the TILSPEC team, it is not a good idea to completely exclude these classes from the classification, as the classifier would not consider the option of an unknown species growing on these places. How to approach these rare species could thus be one of the important issues for the further research.

Regarding the accuracy of other species (results are summarized in appendix B), one of the main successes of this thesis is the high classification accuracy of main peat bog species such as *Trichophorum cespitosum* (0.830), or *Carex rostrata* (0.994) on Hraniční louka, and *Molinia caerulea* (0.836) and *Nardus stricta* (0.886) on Pančavská louka. For *Trichophorum cespitosum* on Hraniční louka, the best F-1 score achieved was 0.866, and no comparable classification of this species was found in the literature for comparison. On Hraniční louka, out of the two *Carex* species, only *Carex rostrata* was consistently well classified. *Carex limosa*, another important peat bog species, according to the confusion matrix, was often classified as *Molinia caerulea*. Kyselé kouty, although it has the best overall accuracy, shows the worst accuracy for specialized grass species, which generally could not be classified with satisfactory accuracy. The high overall F-1 score is mainly due to a larger amount and area of non-specialized species like *Vaccinium myrtillus* or *Picea abies*, which were very accurately classified. On Pančavská louka, the greatest success is the high F-1 scores of *Molinia caerulea* (highest 0.919) and *Nardus stricta* (highest 0.899). In the case of *Nardus stricta*, although not as good a result as Kupková et al. (2023) was achieved, the accuracy is very similar. For the class *Trichophorum cespitosum*, significantly different accuracies were achieved on each orthomosaic (F-1 scores ranging from 0.51 to 0.85). As already discussed earlier, similar inconsistencies can be observed in the overall F-1 scores, which, given that the test dataset polygons were always the same, show that even small differences in orthomosaics can significantly affect classification accuracy.

Table 17 contain compares overall accuracy (OA), achieved by classification in this thesis and relevant literature. Although the main measure of accuracy in this thesis was the weighted F-1 score, most literature uses OA, thus the table compares only OAs. Table 18 compares F-1 score for the most important species, achieved by classification in this thesis and relevant literature.

Table 17: Comparison of accuracy of similar studies and this thesis. All of the compared studies used UAV to acquire image data

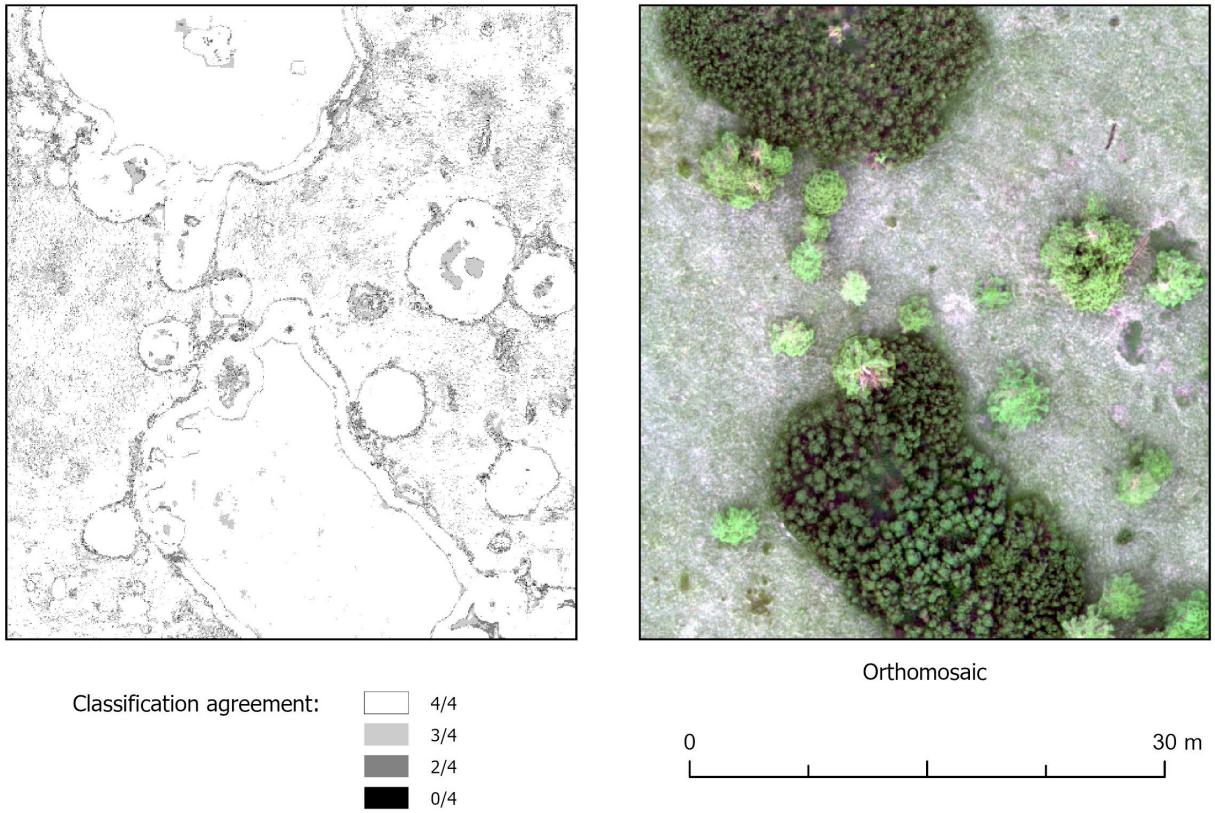
Area of interest	Study	Type of vegetation	Number of classes	Classifier	OA (%)
Uchter moor, Germany	Knoth et al., 2013	Species on the peat bog	4	Object based SVM	91.0
NE Germany	Beyer et al., 2019	Species on the peat bog	11	RF	89.0
South Florida, USA	Zweig et al., 2015	Species on the wetland	9	Object based SVM	69.0
Eagle Bay, Florida, USA	Pande-Chhetri et al., 2017	Species on the wetland	12	Object based SVM	71.0
Northern China	Du et al., 2021	Species on the wetland	6	RF	87.8
Qi'ao island, China	J. Cao et al., 2018	Species of mangroves	10	SVM	88.7
Poptar island, Maryland, USA	Windle et al., 2023	Species on the wetland	5	RF	98
Krkonoše, Úpské rašeliniště	Kupková et al., 2020	Species on the peat bog	14	Object based SVM	86.0
Krkonoše, Bílá louka	Kupková et al., 2023	Species on the grassland	7	Object based SVM	95.9
Krkonoše, Hraniční louka	This thesis	Species on the peat bog	13	RF	91.0
Krkonoše, Kyselé kouty	This thesis	Species on the peat bog	12	SVM	93.4
Krkonoše, Pančavská louka	This thesis	Species on the peat bog	22	SVM	82.7

Table 18: Comparison of F-1 scores for the most important species in this thesis with F-1 scores for the same species in the relevant literature

Species	Study	Classifier	F-1 score
Nardus stricta	This thesis (Pančavská louka)	SVM	0.886
Nardus stricta	Kupková et al., 2017	SVM	0.792
Nardus stricta	Kupková et al., 2023	Object based SVM	0.947
Molinia caerulea	This thesis (Pančavská louka)	SVM	0.836
Molinia caerulea	Kupková et al., 2017	SVM	0.707
Molinia caerulea	Kupková et al., 2023	Object based SVM	0.992
Trichophorum cespitosum	This thesis (Hraniční louka)	RF	0.830
Trichophorum cespitosum	Kupková et al., 2017	Object based SVM	0.770
Picea abies	This thesis (Hraniční louka)	RF	0.965
Picea abies	Zagajewski et al., 2021	SVM	0.917
Pinus mugo	This thesis (Hraniční louka)	RF	0.998
Pinus mugo	Kupková et al., 2017	Object based SVM	0.997

This comparison shows that the overall accuracy of the classifications is comparable to similar studies. The classifications of each important species also usually show comparable accuracy to previous studies. This was particularly true for the classification of *Picea abies* and *Pinus mugo*, which, thanks to the canopy height feature, were classified with nearly 100% accuracy. Conversely, for important species on Pančavská louka (*Molinia caerulea* and *Nardus stricta*), it appears that accuracy could still be improved. The results of the study Kupková et al. (2023) suggest that it would be appropriate to include object-oriented methods among the algorithms tested in further research.
Another important output of this thesis is the result of the overlap analysis. It shows that, as seen from the validation accuracies of individual classes, most classes with sufficient area within each area of interest were classified with similarly by all algorithms. On every orthomosaic, all classifications agreed on 75 – 90 % of the classified area. From the maps describing the degree of agreement in classifications (figure 17 on the following page), it is also evident that the greatest disagreements occur at the edges of stands of different species, where mixing of these species can occur and thus the inability of the classifier to correctly predict the vegetation class. Additionally, in the area of interest with the worst classification accuracy, Pančavská louka, the overlay even yielded better results than the individual classifications.

Figure 17: Example of the degree of agreement map



When addressing how to further improve the accuracy of peat bogs classifications, it is advisable to draw inspiration from the available literature. Many studies on wetland classification utilize hyperspectral data (Du et al., 2021; Zlinszky et al., 2012), which are very suitable for identifying minor differences in water content in plant tissues (Marcinkowska-Ochtyra et al., 2018). Thus, using hyperspectral data could lead to increased accuracy. In contrast, other studies in the Krkonoše Mountains have shown that the use of hyperspectral data may not lead to increased accuracy (Kupková et al., 2023). That was one of the important reasons why it was not used in this thesis.
As already mentioned, it would be beneficial to acquire image data with the UAV several times during the vegetation season and either use this data for multitemporal classification or at least compare them with each other to decide when is the best time for collecting image data from peat bogs. These multitemporal approaches are also common in vegetation monitoring (Kupková et al., 2023; Windle et al., 2023). Another possible suggestion might be to use a different machine learning algorithm, such as Gradient Boost or CNNs. As the author of this thesis, however, I would primarily suggest focusing more on input data than on the algorithm. As some analyses have shown, there was still too little ground truth data in some classes for accurate prediction. In the future, the most important question will be which peat bogs in Krkonoše NP are essential to map and which of the species are most important to monitor for peat bog management. Some of this information is already available (e.g., the need to

monitor the spread of *Nardus stricta* and *Molinia caerulea* on Pančavská louka), but a more detailed discussion with the national park management about the results of the thesis will likely be necessary to determine the next steps. Another key follow-up question will be whether it is possible to classify less common vegetation cover classes with similar accuracy as classes found in large areas/non mixed dense patches of the territory.
From the results of the thesis, it is evident that peat bogs are a very challenging ecosystem for species-level classification. Their dynamics, diversity, and poor accessibility are significant obstacles to quality long-term research. That is also why Krkonoše NP is looking for ways to be as efficient as possible in management using remote sensing methods and to achieve good results at the lowest possible cost. From the results and discussion in this thesis, the following procedure (proposed methodological steps) for creating a peat bog map for Krkonoše NP, emerges:

- Data acquisition phase:
  - Image data should be captured using UAVs under homogeneous lighting conditions with a spatial resolution of 3-5 cm and front and side overlaps of 70-85%.
  - Ground truth data coordinates should be measured with geodetic GPS equipment with accuracy to within a few centimeters. When collecting data, it is crucial not to underestimate their amount. Every small piece of homogeneous vegetation can be used for training and can contribute to greater complexity in the dataset. Ideally, it would be appropriate to have 6 or more sufficiently large polygons for each class. There should be an even greater number of polygons for more common classes, and in all cases, they should be evenly distributed across the area of interest.
- Data preprocessing phase
  - Image data can be processed using automated software workflow (Pix4D Mapper, Agisoft Metashape), but the resulting orthomosaic should be carefully checked to ensure if it is well georeferenced and does not contain undesirable artifacts.
  - Before classification, it is strongly recommended to add additional (besides spectral) features to the dataset that improve classification accuracy. The first of these is the canopy height model, which can be derived photogrammetrically from image data and DTM, and the others are texture features, which help recognize the properties of the classified classes that are not identifiable on a level of a single pixel.
  - Before classification, the separability of the data should be analyzed and the data should be checked for errors. This can be done using an automated algorithm like the FPD proposed in this thesis, or manually by monitoring polygons over the corresponding orthomosaic.
- Data analysis phase
  - The F-1 score, weighted by class area, is a suitable metric for validation, but stratified selection based on class share in the area of interest should also be included in the data selection for training algorithms due to better data representation.

- From the tested classifiers it is recommended to use an RF classifier (in general a classifier that allows for a simple and fast learning and prediction, does not require long hyperparameter tuning, provides good accuracy).
- When testing multiple classification methods, it is beneficial to compare the resulting classifications using an overlap analysis. This can serve as an additional accuracy metric and may even produce a more precise result than the individual classifications.

6 Conclusion

The goal of the thesis was to provide accurate vegetation maps on the species level for three peat bogs in the western part of the Krkonoše Mountains using UAV multispecral data, field botanical data and remote sensing methods for analysis. The other goal was to propose a methodology for efficient and precise vegetation monitoring of Krkonoše peat bogs using multispectral UAV data. The highest F-1 score was achieved at Kyselé kouty (0.957), slightly lower at Hraniční louka (0.899), and the lowest at Pančavská louka (0.832). The pre-set goal of an F-1 score of at least 0.9 was only achieved at Kyselé kouty, but the classification results of the two remaining areas of interest are still very good, as the F-1 score of the most important classes reached high values and the overlay analysis showed that the outputs of both classifiers agree over a large part of the areas of interest (75 – 90%). The varying accuracy of the classifications of each peat bog could have been caused by their complexity, lighting conditions, size and the quality of ground truth data. The classification of each area of interest was tested on two or three orthomosaics acquired on different dates, and it was found that the resulting accuracies also varied significantly. This is likely due to minor differences in lighting conditions of the area of interest during data acquisition and small shifts in vegetation season. Common species, especially conifers - *Picea abies* (0.965), *Pinus mugo* (0.998), as well as mountain or peat bog species - *Molinia caerulea* (0.836), *Nardus stricta* (0.886), *Trichophorum cespitosum* (0.830) were classified with the highest accuracy. Rarer classes, for which sufficient ground truth data was not collected, were classified with the lowest accuracy. The thesis demonstrated that added non-spectral features, in this case canopy height and textural features derived from GLCM, can significantly improve the F-1 score of classification (an increase by 0.113 in the case of canopy height and 0.073 in the case of GLCM textures). Also, through separability analysis and a proposed algorithm for detecting faulty polygons, the need for more detailed investigation of the input ground truth data was demonstrated, as it might be prone to errors during data acquisition or processing.

Among the tested classifiers, the RF algorithm achieved the best accuracy at Hraniční louka, while the best results on the remaining two areas of interest were achieved using the SVM algorithm. Despite slightly better accuracy, the SVM algorithm was assessed as less suitable for peat bog classification, mainly due to long computation times and demanding tuning. The small amount of ground truth data for some categories did not prevent good overall classification results on the test dataset but led to large variability in accuracies of some categories during cross-validation.

The thesis demonstrated the suitability of UAV multispectral data and remote sensing methods for mapping the vegetation cover of peat bogs and proposes a methodological approach, outlined in the last

paragraph of the discussion, which, when implemented, will ensure sufficient accuracy of mapping Krkonoše peat bogs in the future. The map outputs of this thesis will also be used in peat bog management practice in Krkonoše NP. In the coming years, the plan is to continue in this research and to test other approaches, such as different classifiers, to experiment with the amount and distribution of training/validation data or to find out if the classification of multitemporal composites can provide higher classification accuracy. Further field data will be collected and also the classification results will be validated in the field together with botanists of the national park. Also, interdisciplinary discussion about the significance of various species and conservation aspects of the peat bogs and about the possibilities of remote sensing for these aspects and future monitoring of peat bogs will be important.

7 References

When creating the text of the diploma thesis, translations using DeepL.com and ChatGPT 4 were utilized. The text was originally written in Czech and even after translation, it was further modified. No part of the text was directly generated by ChatGPT (or any other language model); it was used only for translation. Abeysinghe, T., Milas, A. S., Arend, K., Hohman, B., Reil, P., Gregory, A., &

Vázquez-Ortega, A. (2019). Mapping invasive Phragmites australis in the Old Woman Creek estuary using UAV remote sensing and machine learning classifiers. Remote Sensing, 11(11). https://doi.org/10.3390/rs11111380

Alvarez-Vanhard, E., Houet, T., Mony, C., Lecoq, L., & Corpetti, T. (2020). Can UAVs fill the gap between in situ surveys and satellites for habitat mapping? Remote Sensing of Environment, 243. https://doi.org/10.1016/j.rse.2020.111780

Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. In Frontiers in Ecology and the Environment (Vol. 11, Issue 3, pp. 138–146). https://doi.org/10.1890/120150

Augusteijn, M. F., Clemens, L. E., & Shaw, K. A. (1995). Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. IEEE Transactions on Geoscience and Remote Sensing, 33(3), 616–626. https://doi.org/10.1109/36.387577

Baraldi, A., & Parmiggiani, F. (1995). An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. IEEE Transactions on Geoscience and Remote Sensing, 33(2), 293–304. https://doi.org/10.1109/36.377929

Beyer, F., Jurasinski, G., Couwenberg, J., & Grenzdörffer, G. (2019). Multisensor data to derive peatland vegetation communities using a fixed-wing unmanned aerial vehicle. International Journal of Remote Sensing, 40(24), 9103–9125. https://doi.org/10.1080/01431161.2019.1580825

Bhatnagar, S., Gill, L., Regan, S., Waldren, S., & Ghosh, B. (2021). A nested drone-satellite approach to monitoring the ecological conditions of wetlands. ISPRS Journal of Photogrammetry and Remote Sensing, 174, 151–165. https://doi.org/10.1016/j.isprsjprs.2021.01.012

Breiman, L. (2001). Random Forests. Machine Learning, Vol. 45, 5-32.

Cao, C., Ni, X., Wang, X., Lu, S., Zhang, Y., Dang, Y., & Singh, R. P. (2016). Allometric scaling theory-based maximum forest tree height and biomass estimation in the Three Gorges reservoir region using multi-source remote-sensing data. International Journal of Remote Sensing, 37(5), 1210–1222. https://doi.org/10.1080/01431161.2015.1117682

Cao, J., Leng, W., Liu, K., Liu, L., He, Z., & Zhu, Y. (2018). Object-Based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models. Remote Sensing, 10(1). https://doi.org/10.3390/rs10010089

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. In ISPRS Journal of Photogrammetry and Remote Sensing (Vol. 92, pp. 79–97). Elsevier B.V. https://doi.org/10.1016/j.isprsjprs.2014.02.013

ČÚZK. (2010). Transformace souřadnic. https://geoportal.cuzk.cz

ČUZK. (2016). ZABAGED® - Výškopis - DMR 5G. https://geoportal.cuzk.cz

Dančejová, D. (2023). Využitie diaľkového prieskumu Zeme pre klasifikáciu vegetácie novej divočiny v zázemí Kutnej Hory Remote sensing for classification of new wilderness vegetation in the hinterland of Kutná Hora [Bachelor thesis]. Charles University.

Davidson, N. C. (2014). How much wetland has the world lost? Long-term and recent trends in global wetland area. Marine and Freshwater Research, 65(10), 934. https://doi.org/10.1071/MF14173

DJI. (2019). Phantom 4 Muntispectral user manual. https://www.dji.com/p4-multispectral/video

Du, B., Mao, D., Wang, Z., Qiu, Z., Yan, H., Feng, K., & Zhang, Z. (2021). Mapping Wetland Plant Communities Using Unmanned Aerial Vehicle Hyperspectral Imagery by Comparing Object/Pixel-Based Classifications Combining Multiple Machine-Learning Algorithms. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 8249–8258. https://doi.org/10.1109/JSTARS.2021.3100923

Eiben, Á. E., Hinterding, R., & Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. IEEE Transactions on Evolutionary Computation, 3(2), 124–141. https://doi.org/10.1109/4235.771166

Eiben, A. E., & Smit, S. K. (2011). Parameter tuning for configuring and analyzing evolutionary algorithms. In Swarm and Evolutionary Computation (Vol. 1, Issue 1, pp. 19–31). Elsevier B.V. https://doi.org/10.1016/j.swevo.2011.02.001

Erdem, F., & Bayrak, O. C. (2023). Evaluating the effects of texture features on Pinus sylvestris classification using high-resolution aerial imagery. Ecological Informatics, 78. https://doi.org/10.1016/j.ecoinf.2023.102389

Fraser, L. H., & Keddy, P. A. (2005). The World's Largest Wetlands: Ecology and conservation. Cambridge University Press.

Główny Urząd Geodezji i Kartografii. (2023). Numeryczny model terenu (NMT). https://www.geoportal.gov.pl/pl/dane/numeryczny-model-terenu-nmt/

Hall-Beyer, M. (2017). GLCM TEXTURE: A TUTORIAL. Department of Geography, University of Calgary. Hejcman, M., Češková, M., & Pavlů, V. (2010). Control of Molinia caerulea by cutting management on sub-alpine grassland. Flora: Morphology, Distribution, Functional Ecology of Plants, 205(9), 577–582. https://doi.org/10.1016/j.flora.2010.04.019

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

Huylenbroeck, L., Laslier, M., Dufour, S., Georges, B., Lejeune, P., & Michez, A. (2020). Using remote sensing to characterize riparian vegetation: A review of available tools and perspectives for managers. In Journal of Environmental Management (Vol. 267). Academic Press. https://doi.org/10.1016/j.jenvman.2020.110652

Kaplan, G., & Avdan, U. (2018). Monthly analysis of wetlands dynamics using remote sensing data. ISPRS International Journal of Geo-Information, 7(10). https://doi.org/10.3390/ijgi7100411

Knoth, C., Klein, B., Prinz, T., & Kleinebecker, T. (2013). Unmanned aerial vehicles as innovative remote sensing platforms for high-resolution infrared imagery to support restoration monitoring in cut-over bogs. Applied Vegetation Science, 16(3), 509–517. https://doi.org/10.1111/avsc.12024

Kulich, A. (2022). Monitoring vegetace rašelinišť v Krkonoších s využitím dálkového průzkumu Země [Bachelor thesis]. Charles University.

Kupková, L., Červená, L., Potůčková, M., Lysák, J., Roubalová, M., Hrázský, Z., Březina, S., Epstein, H. E., & Müllerová, J. (2023). Towards reliable monitoring of grass species in nature conservation: Evaluation of the potential of UAV and PlanetScope multi-temporal data in the Central European tundra. Remote Sensing of Environment, 294. https://doi.org/10.1016/j.rse.2023.113645

Kupková, L., Červená, L., Potůčková, M., Lysák, J., Šašková, M., Šrollerů, A., Klinerová, T., Bobek, P., & Müllerová, J. (2020). Vegetation of the tundra in the Krkonoše Mts.-past, present and future, Annual Report 2020.

Kupková, L., Červená, L., Suchá, R., Jakešová, L., Zagajewski, B., Březina, S., & Albrechtová, J. (2017). Classification of tundra vegetation in the Krkonoše Mts. National park using APEX, AISA dual and

sentinel-2A data. European Journal of Remote Sensing, 50(1), 29–46. https://doi.org/10.1080/22797254.2017.1274573

Lambora, A., Gupta, K., & Chopra, K. (2019). Genetic Algorithm- A Literature Review. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 380–384. https://doi.org/10.1109/COMITCon.2019.8862255

Li, Q., Hu, B., Shang, J., & Li, H. (2023). Fusion Approaches to Individual Tree Species Classification Using Multisource Remote Sensing Data †. Forests, 14(7). https://doi.org/10.3390/f14071392

Liu, X., & Bo, Y. (2015). Object-Based Crop Species Classification Based on the Combination of Airborne Hyperspectral Images and LiDAR Data. Remote Sensing, 7(1), 922–950. https://doi.org/10.3390/rs70100922

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, 28(5), 823–870. https://doi.org/10.1080/01431160600746456

Manfreda, S., McCabe, M. F., Miller, P. E., Lucas, R., Madrigal, V. P., Mallinis, G., Dor, E. Ben, Helman, D., Estes, L., Ciraolo, G., Müllerová, J., Tauro, F., de Lima, M. I., de Lima, J. L. M. P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., … Toth, B. (2018). On the use of unmanned aerial systems for environmental monitoring. In Remote Sensing (Vol. 10, Issue 4). MDPI AG. https://doi.org/10.3390/rs10040641

Marcinkowska-Ochtyra, A., Zagajewski, B., Ochtyra, A., Jarocińska, A., Wojtuń, B., Rogass, C., Mielke, C., & Lavender, S. (2017). Subalpine and alpine vegetation classification based on hyperspectral APEX and simulated EnMAP images. International Journal of Remote Sensing, 38(7), 1839–1864. https://doi.org/10.1080/01431161.2016.1274447

Marcinkowska-Ochtyra, A., Zagajewski, B., Raczko, E., Ochtyra, A., & Jarocińska, A. (2018). Classification of high-mountain vegetation communities within a diverse Giant Mountains ecosystem using airborne APEX hyperspectral imagery. Remote Sensing, 10(4). https://doi.org/10.3390/rs10040570

Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., & Gioli, B. (2015). Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. Remote Sensing, 7(3), 2971–2990. https://doi.org/10.3390/rs70302971

Michez, A., Piégay, H., Jonathan, L., Claessens, H., & Lejeune, P. (2016). Mapping of riparian invasive species with supervised classification of Unmanned Aerial System (UAS) imagery. International Journal of Applied Earth Observation and Geoinformation, 44, 88–94. https://doi.org/10.1016/j.jag.2015.06.014

Mohammadpour, P., Viegas, D. X., & Viegas, C. (2022). Vegetation Mapping with Random Forest Using Sentinel 2 and GLCM Texture Feature—A Case Study for Lousã Region, Portugal. Remote Sensing, 14(18). https://doi.org/10.3390/rs14184585

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of Machine Learning (Second edition). The MIT Press.

Nex, F., & Remondino, F. (2014). UAV for 3D mapping applications: A review. In Applied Geomatics (Vol. 6, Issue 1, pp. 1–15). Springer Verlag. https://doi.org/10.1007/s12518-013-0120-x

NP Šumava. (2019). O rašeliništích. https://life.npsumava.cz

Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., & Wilhelm, V. L. (2017). Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. European Journal of Remote Sensing, 50(1), 564–576. https://doi.org/10.1080/22797254.2017.1373602

Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. WIREs Data Mining and Knowledge Discovery, 9(3). https://doi.org/10.1002/widm.1301

Richards, J. A. (John A., & Jia, Xiuping. (2006). Remote sensing digital image analysis: an introduction. Springer.

Scikit-learn developers. (2024). sklearn.svm.SVC. https://scikit-learn.org

Shakhatreh, H., Sawalmeh, A. H., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., Othman, N. S., Khreishah, A., & Guizani, M. (2019). Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. In IEEE Access (Vol. 7, pp. 48572–48634). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2019.2909530

Simon, S., Kolyada, N., Akiki, C., Potthast, M., Stein, B., & Siegmund, N. (2023). Exploring Hyperparameter Usage and Tuning in Machine Learning Research. 2023 IEEE/ACM 2nd International Conference on AI Engineering – Software Engineering for AI (CAIN), 68–79. https://doi.org/10.1109/CAIN58948.2023.00016

Správa KRNAP. (2022). Severská rašeliniště. https://Www.Krnap.Cz/Severska-Raseliniste/

Vapnik, V., & Cortes, C. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1023/A:1022627411411

Volf, O., Koptík, J., Obstová, L., Holá, E., Manukjanová, A., & Štechová, T. (2019). Revitalizace mokřadů – zpracování odborné části, Závěrečná zpráva za rok 2019.

Windle, A. E., Staver, L. W., Elmore, A. J., Scherer, S., Keller, S., Malmgren, B., & Silsbe, G. M. (2023). Multi-temporal high-resolution marsh vegetation mapping using unoccupied aircraft system

remote sensing and machine learning. Frontiers in Remote Sensing, 4.
https://doi.org/10.3389/frsen.2023.1140999

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology, 1(1), 9–23. https://doi.org/10.1093/jpe/rtm005

Xu, C., Zhao, D., Zheng, Z., Zhao, P., Chen, J., Li, X., Zhao, X., Zhao, Y., Liu, W., Wu, B., & Zeng, Y. (2023). Correction of UAV LiDAR-derived grassland canopy height based on scan angle. Frontiers in Plant Science, 14. https://doi.org/10.3389/fpls.2023.1108109

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A., & Kycko, M. (2021). Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the krkonoše/karkonosze transboundary biosphere reserve. Remote Sensing, 13(13).
https://doi.org/10.3390/rs13132581

Zaman, B., Jensen, A. M., & McKee, M. (2011). Use of high-resolution multispectral imagery acquired with an autonomous unmanned aerial vehicle to quantify the spread of an invasive wetlands species. 2011 IEEE International Geoscience and Remote Sensing Symposium, 803–806.
https://doi.org/10.1109/IGARSS.2011.6049252

Zhang, Z., Kazakova, A., Moskal, L., & Styers, D. (2016). Object-Based Tree Species Classification in Urban Ecosystems Using LiDAR and Hyperspectral Data. Forests, 7(12), 122.
https://doi.org/10.3390/f7060122

Zhao, Y., Chen, R. H., Bakian-Dogaheh, K., Whitcomb, J., Yi, Y., Kimball, J. S., & Moghaddam, M. (2022). Mapping Boreal Forest Species and Canopy Height using Airborne SAR and Lidar Data in Interior Alaska. International Geoscience and Remote Sensing Symposium (IGARSS), 2022-July, 4955–4958. https://doi.org/10.1109/IGARSS46834.2022.9883311

Zlinszky, A., Mücke, W., Lehner, H., Briese, C., & Pfeifer, N. (2012). Categorizing wetland vegetation by airborne laser scanning on Lake Balaton and Kis-Balaton, Hungary. Remote Sensing, 4(6), 1617–1650. https://doi.org/10.3390/rs4061617

Zweig, C. L., Burgess, M. A., Percival, H. F., & Kitchens, W. M. (2015). Use of Unmanned Aircraft Systems to Delineate Fine-Scale Wetland Vegetation Communities. Wetlands, 35(2), 303–309.
https://doi.org/10.1007/s13157-014-0612-4

A Images of classified species

Anthotraxum odoratum
(source: TILSPEC)

Callamagrostis villosa
(source: TILSPEC)

Calluna vulgaris
(source: TILSPEC)

Carex limosa
(source: TILSPEC)

Carex rostrata
(source: TILSPEC)

Deschampsia cespitosa
(source: TILSPEC)

Eriophorum angustifolium
(source: TILSPEC)

Eriophorum vaginatum
(source: TILSPEC)

Hieracium sp.
(source: TILSPEC)

Juncus sp.
(source: TILSPEC)

Example of unidentified moss
(source: TILSPEC)

Molinia caerulea
(source: TILSPEC)

Nardus stricta
(source: TILSPEC)

Picea abies
(source: TILSPEC)

Pinus mugo
(source: Wikimedia Commons)

Pinus unicata subs. uliginosa
(source: TILSPEC)

Salix sp.
(source: TILSPEC)

Sphagnum sp.
(source: TILSPEC)

Trichophorum cespitosum
(source: TILSPEC)

Vaccinium myrtillus
(source: TILSPEC)

Vaccinium vitis-idaea
(source: TILSPEC)

Vaccinium uliginosum
(source: TILSPEC)

B Classification results for all classes

Table 19: Classification results for Hraniční louka orthomosaic from July the 13th 2023 (F-1 scores)

	RF	SVM (linear)	SVM (polynomial)	SVM (RBF)	SVM (Sigmoid)
Best hyperparameters	ntree = 464 mtry = 15 depth = 18	C = 2.57	C = 433.09 Degree = 3 Gamma = 0.0082	C = 945.37 Gamma = 2.86e-07	C = 394.22 Gamma = 3.22e-09
Overall F-1 score	0.899	0.875	0.879	0.895	0.779
Overall accuracy	91.0%	89.0%	90.2%	90.7%	80.3%
Vaccinium myrtillus	1.000	0.838	0.867	0.899	0.000
Carex limosa	0.339	0.456	0.324	0.506	0.000
Carex rostrata	0.994	0.963	0.966	0.961	0.816
Unidentified moss	0.956	0.918	0.955	0.969	0.517
Molinia caerulea	0.610	0.400	0.443	0.555	0.474
Nardus stricta	1.000	0.977	0.956	0.970	0.00
Sphagnum species	0.776	0.583	0.634	0.627	0.506
Trichophorum cespitosum	0.830	0.706	0.807	0.777	0.740
Vaccinium uliginosum	0.000	0.256	0.318	0.333	0.000
Pinus mugo	0.998	0.995	0.987	0.989	0.923
Picea abies	0.965	0.994	0.995	0.995	0.954
Water	1.000	1.000	1.000	1.000	0.985
Dry vegetation	0.000	0.924	0.942	0.906	0.908

Table 20: Classification results for Hraniční louka orthomosaic from July the 25th 2023 (F-1 scores)

	RF	SVM (linear)	SVM (polynomial)	SVM (RBF)	SVM (Sigmoid)
Best hyperparameters	ntree = 138 mtry = 18 depth = 17	C = 9.94	C = 1.64e-07 Degree = 2 Gamma = 0.034	C = 593.02 Gamma = 8.84e-07	C = 4.928 Gamma = 3.53e-07
Overall F-1 score	0.893	0.805	0.710	0.805	0.807
Overall accuracy	90.0%	81.8%	79.0%	81.8%	80.3%
Vaccinium myrtillus	1.000	0.526	0.000	0.526	0.526
Carex limosa	0.518	0.805	0.304	0.797	0.821
Carex rostrata	1.000	0.977	0.870	0.977	0.977
Unidentified moss	0.000	0.000	0.000	0.000	0.000
Molinia caerulea	0.825	0.783	0.781	0.783	0.783
Nardus stricta	1.000	1.000	0.000	1.000	1.000
Sphagnum species	0.021	0.037	0.000	0.046	0.056
Trichophorum cespitosum	0.866	0.762	0.733	0.765	0.766
Vaccinium uliginosum	0.9375	0.271	0.000	0.271	0.271
Pinus mugo	1.000	0.834	0.870	0.834	0.834
Picea abies	0.965	0.992	0.904	0.992	0.992
Water	1.000	1.000	0.946	1.000	1.000
Dry vegetation	0.000	0.931	0.722	0.931	0.931

Table 21: Classification results for Kyselé kouty orthomosaic from July the 13th 2023 (F-1 scores)

	RF	SVM (linear)	SVM (polynomial)	SVM (RBF)	SVM (Sigmoid)
Best hyperparameters	ntree = 885 mtry = 6 depth = 19	C = 0.00024	C = 0.15 Degree = 2 Gamma = 4.31e-05	C = 10.97 Gamma = 5.38e-06	C = 2.95 Gamma = 2.82e-07
Overall F-1 score	0.924	0.931	0.930	0.924	0.918
Overall accuracy	93.4%	92.6%	92.5%	91.6%	91.9%
Bare ground	0.954	0.802	0.688	0.740	0.930
Vaccinium myrtillus	0.969	0.969	0.966	0.960	0.964
Carex species	0.760	0.760	0.836	0.838	0.000
Deschampsia cespitosa	0.000	0.000	0.000	0.000	0.000
Eriophorum vaginatum	0.143	0.143	0.074	0.039	0.015
Juncus species	0.284	0.284	0.334	0.321	0.350
Unidentified moss	0.324	0.324	0.264	0.055	0.000
Nardus stricta	0.578	0.578	0.638	0.480	0.000
Pinus uncinata subsp. uliginosa	0.029	0.000	0.253	0.056	0.000
Picea abies	0.973	0.988	0.992	0.991	0.986
Sphagnum species	0.000	0.000	0.000	0.000	0.000
Dry vegetation	0.802	0.795	0.914	0.914	0.646

Table 22: Classification results for Kyselé kouty orthomosaic from July the 25th 2023 (F-1 scores)

	RF	SVM (linear)	SVM (polynomial)	SVM (RBF)	SVM (Sigmoid)
Best hyperparameters	ntree = 970 mtry = 8 depth = 19	C = 0.00012	C = 0.0199 Degree = 2 Gamma = 0.00029	C = 235.06 Gamma = 1.899e-06	C = 107.20 Gamma = 5.07e-09
Overall F-1 score	0.939	0.957	0.950	0.949	0.878
Overall accuracy	90.3%	90.4%	89.7%	89.9%	82.3%
Bare ground	1.000	0.997	0.994	1.000	0.945
Vaccinium myrtillus	0.932	0.977	0.981	0.969	0.859
Carex species	0.951	0.878	0.906	0.914	0.791
Deschampsia cespitosa	0.000	0.010	0.009	0.000	0.000
Eriophorum vaginatum	0.349	0.618	0.646	0.496	0.091
Juncus species	0.770	0.704	0.681	0.690	0.617
Unidentified moss	0.626	0.701	0.696	0.727	0.000
Nardus stricta	0.165	0.503	0.142	0.154	0.057
Pinus uncinata subsp. uliginosa	0.686	0.500	0.930	0.947	0.000
Picea abies	1.000	0.985	0.974	0.985	0.974
Sphagnum species	0.000	0.000	0.000	0.000	0.000
Dry vegetation	1.000	0.775	0.519	0.936	0.583

Table 23: Classification results for Pančavská louka orthomosaic from July the 14th 2023 (F-1 scores)

	RF	SVM (linear)	SVM (polynomial)	SVM (RBF)	SVM (Sigmoid)
Best hyperparameters	ntree = 984 mtry = 8 depth = 20	C = 0.00036	C = 0.024 Degree = 3 Gamma = 4.36e-05	C = 436.3 Gamma = 2.9e-06	C = 218.2 Gamma = 9.6e-08
Overall F-1 score	0.796	0.816	0.831	0.832	0.773
Overall accuracy	80.8%	80.7%	82.6%	82.7%	75.7%
Vaccinium myrtillus	0.425	0.487	0.556	0.568	0.360
Vaccinium vitis-idaea	0.000	0.000	0.000	0.000	0.000
Calamagrostis villosa	0.975	0.811	0.905	0.889	0.393
Carex limosa	0.532	0.335	0.302	0.301	0.366
Carex rostrata	0.591	0.504	0.492	0.491	0.294
Deschampsia cespitosa	0.173	0.359	0.436	0.408	0.323
Eriophorum angustifolium	0.063	0.238	0.168	0.173	0.238
Hieracium species	1.000	0.997	1.000	1.000	0.859
Juncus species	0.975	0.307	0.919	0.882	0.000
Unidentified moss	0.000	0.000	0.000	0.000	0.000
Molinia caerulea	0.790	0.815	0.836	0.835	0.789
Nardus stricta	0.865	0.865	0.886	0.884	0.830
Salix species	0.936	0.292	0.826	0.766	0.000
Sphagnum species	0.355	0.438	0.421	0.423	0.000
Anthoxanthum odoratum	1.000	0.992	1.000	1.000	0.934
Trichophorum cespitosum	0.819	0.840	0.847	0.853	0.799
Vaccinium uliginosum	0.030	0.620	0.665	0.667	0.380
Calluna vulgaris	0.262	0.554	0.430	0.445	0.392
Water	1.000	1.000	1.000	1.000	1.000
Pinus mugo	0.960	0.981	0.992	0.994	0.971
Picea abies	0.933	0.906	0.924	0.929	0.886
Dry vegetation	0.000	0.995	0.983	0.980	100.0

Table 24: Classification results for Pančavská louka orthomosaic from July the 26th 2023 (F-1 scores)

	RF	SVM (linear)	SVM (polynomial)	SVM (RBF)	SVM (Sigmoid)
Best hyperparameters	ntree = 51 mtry = 8 depth = 19	C = 0.000103	C = 14.7 Degree = 3 Gamma = 1.19e-05	C = 72.02 Gamma = 4.51e-07	C = 10.23 Gamma = 2.35e-07
Overall F-1 score	0.776	0.794	0.806	0.796	747
Overall accuracy	80.0%	80.9%	81.7%	81.2%	76.3%
Vaccinium myrtillus	297	0.558	0.345	0.489	0.000
Vaccinium vitis-idaea	0.000	0.000	0.192	0.000	0.000
Calamagrostis villosa	0.945	0.600	0.808	0.688	0.000
Carex limosa	0.652	0.640	0.623	0.633	0.681
Carex rostrata	0.302	0.322	0.396	0.336	0.168
Deschampsia cespitosa	0.163	0.193	0.044	0.082	0.137
Eriophorum angustifolium	0.074	0.046	0.261	0.137	0.000
Hieracium species	0.992	0.958	0.997	0.956	0.842
Juncus species	1.000	0.382	0.986	0.715	0.000
Unidentified moss	0.000	0.000	0.108	0.000	0.000
Molinia caerulea	0.886	0.898	0.910	0.900	0.815
Nardus stricta	0.814	0.899	0.869	0.884	0.827
Salix species	0.827	0.533	0.846	0.620	0.000
Sphagnum species	0.000	0.114	0.161	0.209	0.000
Anthoxanthum odoratum	0.999	0.985	0.993	0.977	0.892
Trichophorum cespitosum	0.623	0.697	0.710	0.721	0.637
Vaccinium uliginosum	0.205	0.621	0.629	0.624	0.199
Calluna vulgaris	0.111	0.347	0.316	0.323	0.080
Water	1.000	1.000	1.000	1.000	1.000
Pinus mugo	1.000	0.960	0.988	0.958	0.997
Picea abies	0.895	0.848	0.867	0.839	0.902
Dry vegetation	0.634	0.902	0.569	0.912	0.762

Table 25: Classification results for Pančavská louka orthomosaic from July the 27th 2023 (F-1 scores)

	RF	SVM (linear)	SVM (polynomial)	SVM (RBF)	SVM (Sigmoid)
Best hyperparameters	ntree = 830 mtry = 9 depth = 20	C = 2.07e-05	C = 0.1 Degree = 3 Gamma = 6.46e-05	C = 98.2 Gamma = 1.40e-06	C = 912.7 Gamma = 1.40e-08
Overall F-1 score	0.753	0.701	0.769	0.758	0.694
Overall accuracy	80.0%	74.6%	80.2%	79.4%	73.9%
Vaccinium myrtillus	0.189	0.016	0.183	0.149	0.000
Vaccinium vitis-idaea	0.977	0.000	0.937	0.489	0.000
Calamagrostis villosa	0.953	0.355	0.857	0.808	0.275
Carex limosa	0.439	0.499	0.413	0.484	0.495
Carex rostrata	0.389	0.180	0.377	0.329	0.175
Deschampsia cespitosa	0.014	0.009	0.038	0.030	0.000
Eriophorum angustifolium	0.060	0.035	0.147	0.179	0.023
Hieracium species	1.000	0.646	0.989	0.984	0.596
Juncus species	1.000	0.000	0.910	0.403	0.000
Unidentified moss	0.000	0.000	0.187	0.198	0.000
Molinia caerulea	0.902	0.897	0.919	0.906	0.892
Nardus stricta	0.868	0.770	0.879	0.847	0.767
Salix species	0.795	0.000	0.861	0.618	0.000
Sphagnum species	0.506	0.000	0.000	0.000	0.000
Anthoxanthum odoratum	1.000	0.854	0.985	0.967	0.835
Trichophorum cespitosum	0.546	0.534	0.596	0.582	0.531
Vaccinium uliginosum	0.190	0.709	0.708	0.762	0.595
Calluna vulgaris	0.196	0.395	0.382	0.461	0.308
Water	1.000	1.000	0.986	1.000	1.000
Pinus mugo	0.890	0.810	0.883	0.877	0.802
Picea abies	0.848	0.828	0.842	0.855	0.827
Dry vegetation	0.667	0.643	0.703	0.690	0.629

C Separability tables

Figure 18: Separability of pairs of classes for Hraniční louka orthomosaic from July the 13th 2023

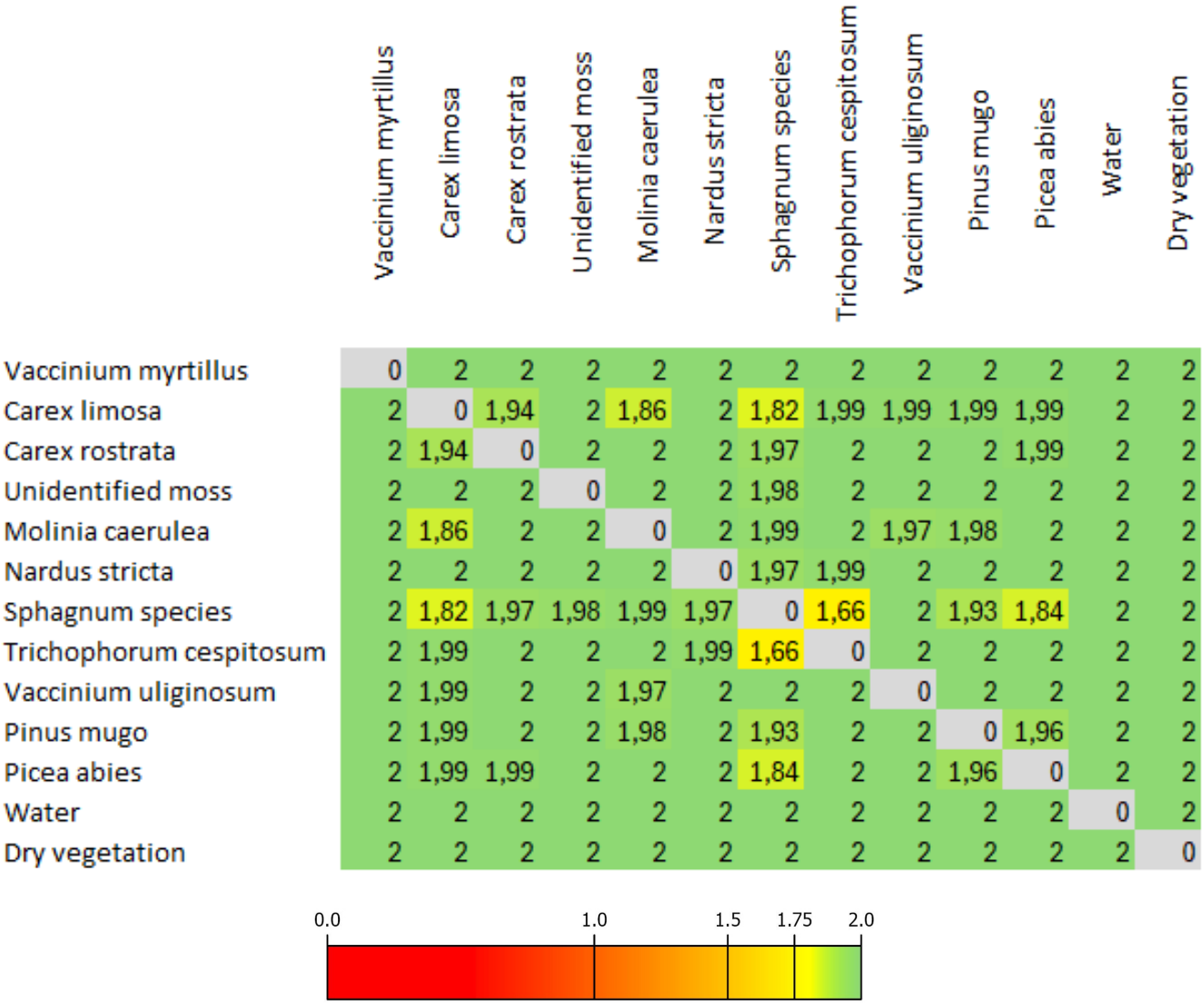
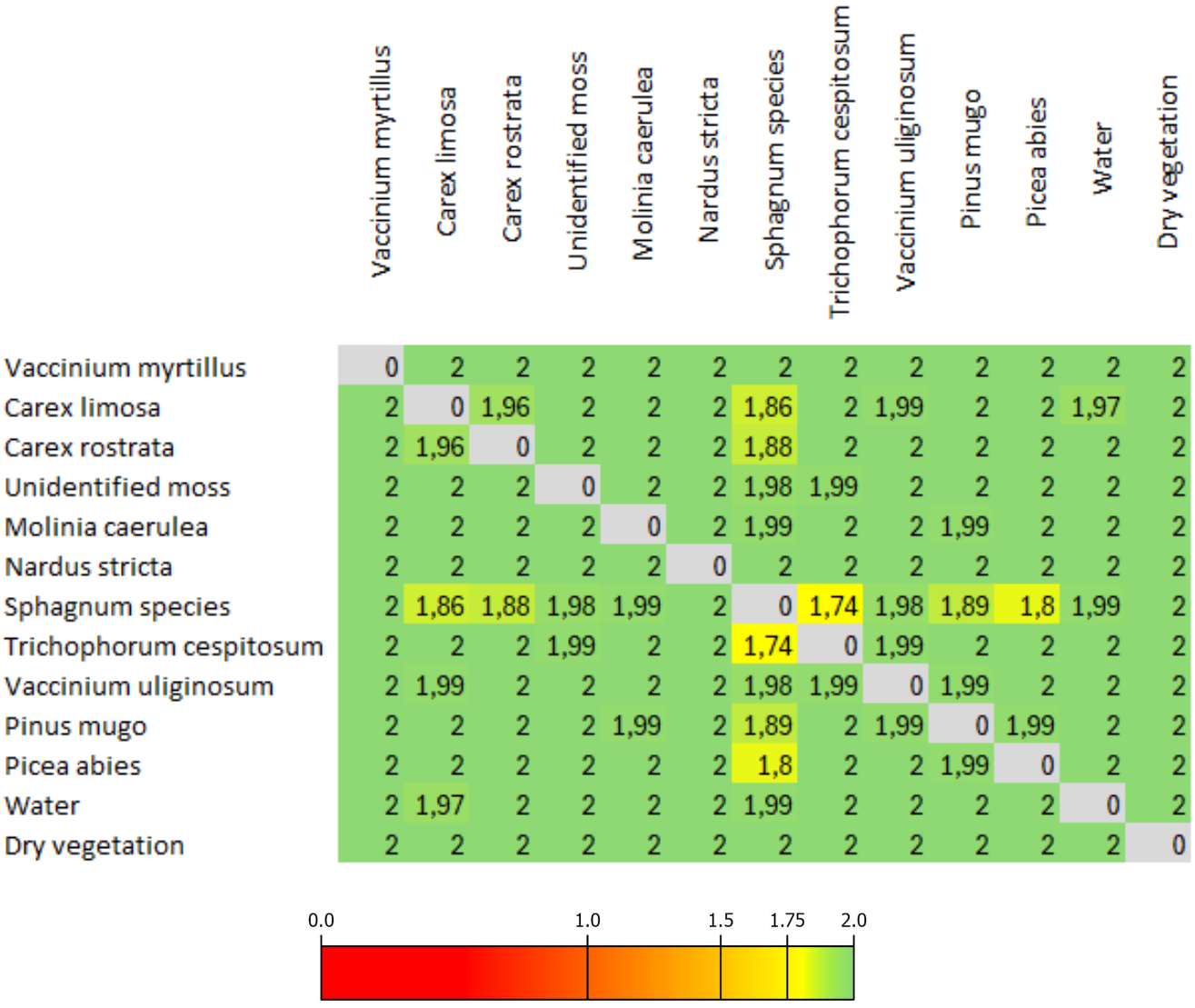


Figure 19: Separability of pairs of classes for Hraniční louka orthomosaic from July the 25th 2023



	Vaccinium myrtillus	Carex limosa	Carex rostrata	Unidentified moss	Molinia caerulea	Nardus stricta	Sphagnum species	Trichophorum cespitosum	Vaccinium uliginosum	Pinus mugo	Picea abies	Water	Dry vegetation
Vaccinium myrtillus	0	2	2	2	2	2	2	2	2	2	2	2	2
Carex limosa	2	0	1,96	2	2	2	1,86	2	1,99	2	2	1,97	2
Carex rostrata	2	1,96	0	2	2	2	1,88	2	2	2	2	2	2
Unidentified moss	2	2	2	0	2	2	1,98	1,99	2	2	2	2	2
Molinia caerulea	2	2	2	2	0	2	1,99	2	2	1,99	2	2	2
Nardus stricta	2	2	2	2	2	0	2	2	2	2	2	2	2
Sphagnum species	2	1,86	1,88	1,98	1,99	2	0	1,74	1,98	1,89	1,8	1,99	2
Trichophorum cespitosum	2	2	2	1,99	2	2	1,74	0	1,99	2	2	2	2
Vaccinium uliginosum	2	1,99	2	2	2	2	1,98	1,99	0	1,99	2	2	2
Pinus mugo	2	2	2	2	1,99	2	1,89	2	1,99	0	1,99	2	2
Picea abies	2	2	2	2	2	2	1,8	2	2	1,99	0	2	2
Water	2	1,97	2	2	2	2	1,99	2	2	2	2	0	2
Dry vegetation	2	2	2	2	2	2	2	2	2	2	2	2	0

Figure 20: Separability of pairs of classes for Kyselé kouty orthomosaic from July the 13th 2023

	Bare ground	Vaccinium myrtillus	Carex species	Deschampsia cespitosa	Eriophorum vaginatum	Juncus species	Unidentified moss	Nardus stricta	Pinus uncinata subsp. uliginosa	Picea abies	Sphagnum species	Dry vegetation
Bare ground	0	1,99	2	2	2	2	1,99	2	2	2	2	2
Vaccinium myrtillus	1,99	0	1,93	1,97	1,77	1,9	1,55	1,94	1,93	1,96	1,83	2
Carex species	2	1,93	0	1,79	1,56	1,66	1,92	1,89	2	2	1,99	2
Deschampsia cespitosa	2	1,97	1,79	0	1,58	1,75	1,99	1,94	2	1,99	2	2
Eriophorum vaginatum	2	1,77	1,56	1,58	0	1,28	1,9	1,74	2	1,99	1,94	2
Juncus species	2	1,9	1,66	1,75	1,28	0	1,91	1,64	2	2	1,97	2
Unidentified moss	1,99	1,55	1,92	1,99	1,9	1,91	0	1,99	2	1,99	1,88	2
Nardus stricta	2	1,94	1,89	1,94	1,74	1,64	1,99	0	1,99	1,99	1,99	2
Pinus uncinata subsp. uliginosa	2	1,93	2	2	2	2	2	1,99	0	2	1,98	2
Picea abies	2	1,96	2	1,99	1,99	2	1,99	1,99	2	0	2	2
Sphagnum species	2	1,83	1,99	2	1,94	1,97	1,88	1,99	1,98	2	0	2
Dry vegetation	2	2	2	2	2	2	2	2	2	2	2	0

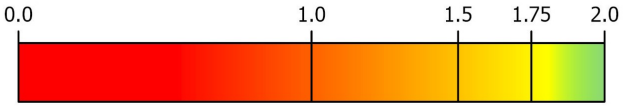
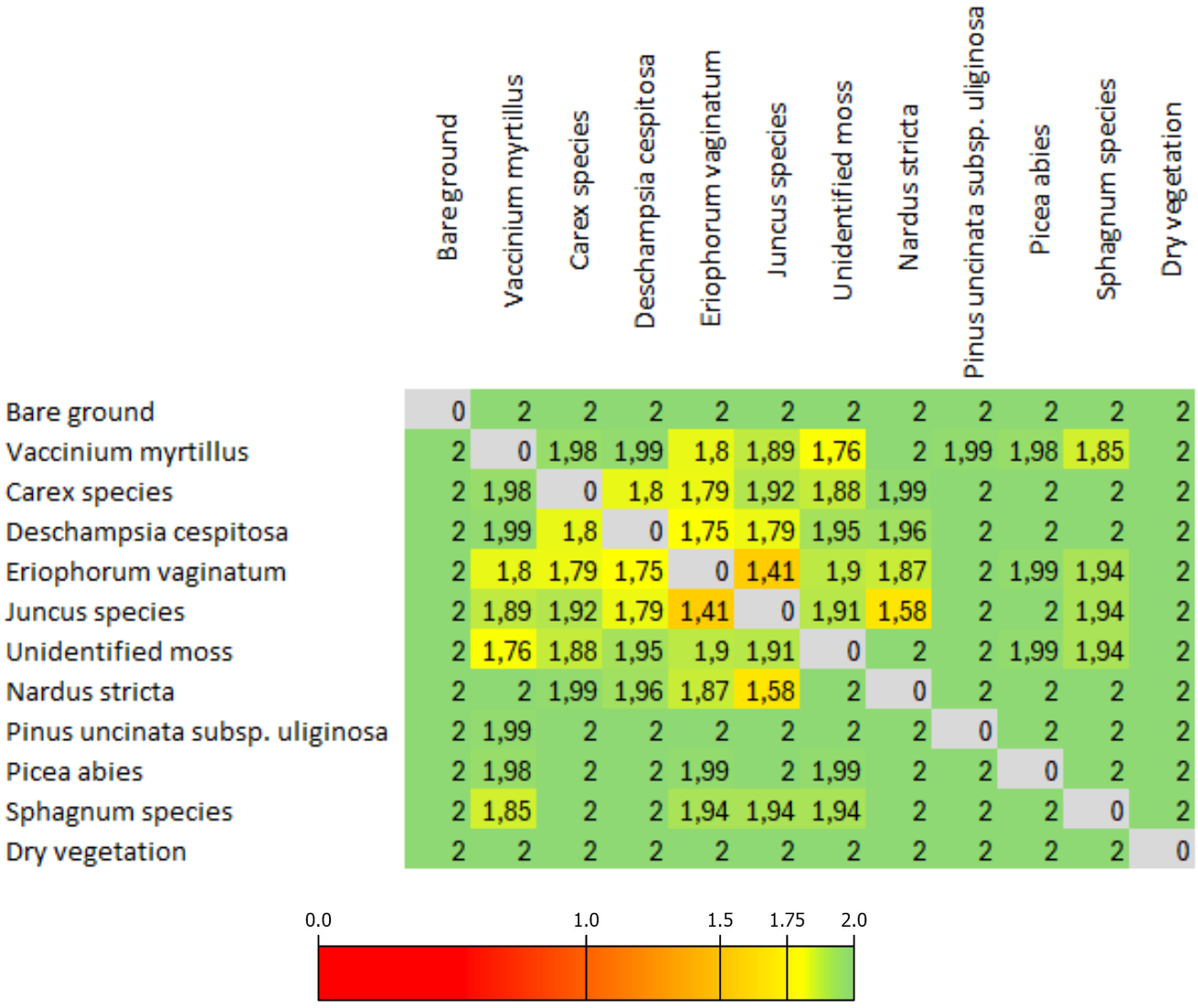
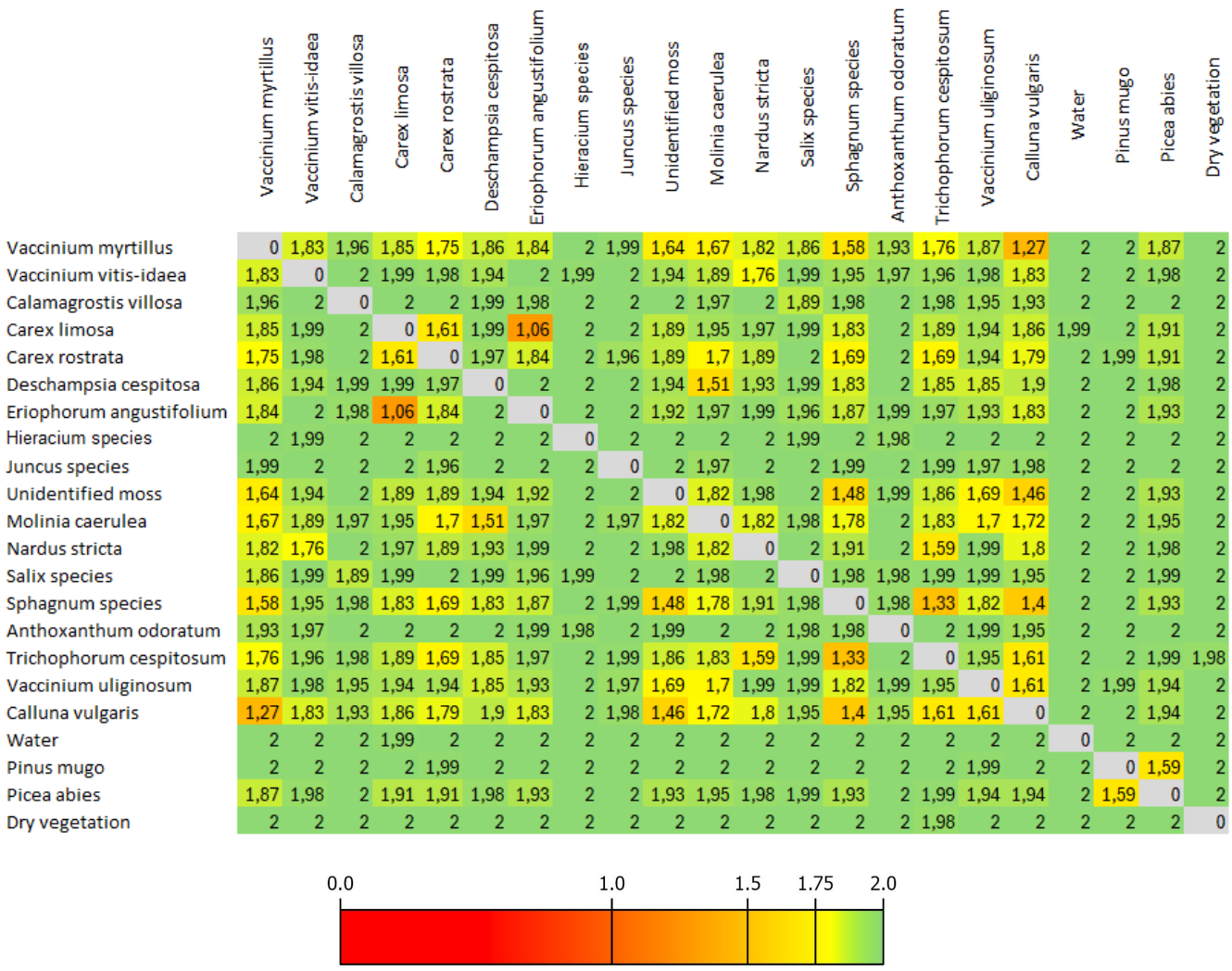


Figure 21: Separability of pairs of classes for Kyselé kouty orthomosaic from July the 25th 2023



	Bare ground	Vaccinium myrtillus	Carex species	Deschampsia cespitosa	Eriophorum vaginatum	Juncus species	Unidentified moss	Nardus stricta	Pinus uncinata subsp. uliginosa	Picea abies	Sphagnum species	Dry vegetation
Bare ground	0	2	2	2	2	2	2	2	2	2	2	2
Vaccinium myrtillus	2	0	1,98	1,99	1,8	1,89	1,76	2	1,99	1,98	1,85	2
Carex species	2	1,98	0	1,8	1,79	1,92	1,88	1,99	2	2	2	2
Deschampsia cespitosa	2	1,99	1,8	0	1,75	1,79	1,95	1,96	2	2	2	2
Eriophorum vaginatum	2	1,8	1,79	1,75	0	1,41	1,9	1,87	2	1,99	1,94	2
Juncus species	2	1,89	1,92	1,79	1,41	0	1,91	1,58	2	2	1,94	2
Unidentified moss	2	1,76	1,88	1,95	1,9	1,91	0	2	2	1,99	1,94	2
Nardus stricta	2	2	1,99	1,96	1,87	1,58	2	0	2	2	2	2
Pinus uncinata subsp. uliginosa	2	1,99	2	2	2	2	2	2	0	2	2	2
Picea abies	2	1,98	2	2	1,99	2	1,99	2	2	0	2	2
Sphagnum species	2	1,85	2	2	1,94	1,94	1,94	2	2	2	0	2
Dry vegetation	2	2	2	2	2	2	2	2	2	2	2	0

Figure 22: Separability of pairs of classes for Pančavská louka orthomosaic from July the 14th 2023



	Vaccinium myrtillus	Vaccinium vitis-idaea	Calamagrostis villosa	Carex limosa	Carex rostrata	Deschampsia cespitosa	Eriophorum angustifolium	Hieracium species	Juncus species	Unidentified moss	Molinia caerulea	Nardus stricta	Salix species	Sphagnum species	Anthoxanthum odoratum	Trichophorum cespitosum	Vaccinium uliginosum	Calluna vulgaris	Water	Pinus mugo	Picea abies	Dry vegetation
Vaccinium myrtillus	0	1,83	1,96	1,85	1,75	1,86	1,84	2	1,99	1,64	1,67	1,82	1,86	1,58	1,93	1,76	1,87	1,27	2	2	1,87	2
Vaccinium vitis-idaea	1,83	0	2	1,99	1,98	1,94	2	1,99	2	1,94	1,89	1,76	1,99	1,95	1,97	1,96	1,98	1,83	2	2	1,98	2
Calamagrostis villosa	1,96	2	0	2	2	1,99	1,98	2	2	2	1,97	2	1,89	1,98	2	1,98	1,95	1,93	2	2	2	2
Carex limosa	1,85	1,99	2	0	1,61	1,99	1,06	2	2	1,89	1,95	1,97	1,99	1,83	2	1,89	1,94	1,86	1,99	2	1,91	2
Carex rostrata	1,75	1,98	2	1,61	0	1,97	1,84	2	1,96	1,89	1,7	1,89	2	1,69	2	1,69	1,94	1,79	2	1,99	1,91	2
Deschampsia cespitosa	1,86	1,94	1,99	1,99	1,97	0	2	2	2	1,94	1,51	1,93	1,99	1,83	2	1,85	1,85	1,9	2	2	1,98	2
Eriophorum angustifolium	1,84	2	1,98	1,06	1,84	2	0	2	2	1,92	1,97	1,99	1,96	1,87	1,99	1,97	1,93	1,83	2	2	1,93	2
Hieracium species	2	1,99	2	2	2	2	2	0	2	2	2	2	1,99	2	1,98	2	2	2	2	2	2	2
Juncus species	1,99	2	2	2	1,96	2	2	2	0	2	1,97	2	2	1,99	2	1,99	1,97	1,98	2	2	2	2
Unidentified moss	1,64	1,94	2	1,89	1,89	1,94	1,92	2	2	0	1,82	1,98	2	1,48	1,99	1,86	1,69	1,46	2	2	1,93	2
Molinia caerulea	1,67	1,89	1,97	1,95	1,7	1,51	1,97	2	1,97	1,82	0	1,82	1,98	1,78	2	1,83	1,7	1,72	2	2	1,95	2
Nardus stricta	1,82	1,76	2	1,97	1,89	1,93	1,99	2	2	1,98	1,82	0	2	1,91	2	1,59	1,99	1,8	2	2	1,98	2
Salix species	1,86	1,99	1,89	1,99	2	1,99	1,96	1,99	2	2	1,98	2	0	1,98	1,98	1,99	1,99	1,95	2	2	1,99	2
Sphagnum species	1,58	1,95	1,98	1,83	1,69	1,83	1,87	2	1,99	1,48	1,78	1,91	1,98	0	1,98	1,33	1,82	1,4	2	2	1,93	2
Anthoxanthum odoratum	1,93	1,97	2	2	2	2	1,99	1,98	2	1,99	2	2	1,98	1,98	0	2	1,99	1,95	2	2	2	2
Trichophorum cespitosum	1,76	1,96	1,98	1,89	1,69	1,85	1,97	2	1,99	1,86	1,83	1,59	1,99	1,33	2	0	1,95	1,61	2	2	1,99	1,98
Vaccinium uliginosum	1,87	1,98	1,95	1,94	1,94	1,85	1,93	2	1,97	1,69	1,7	1,99	1,99	1,82	1,99	1,95	0	1,61	2	1,99	1,94	2
Calluna vulgaris	1,27	1,83	1,93	1,86	1,79	1,9	1,83	2	1,98	1,46	1,72	1,8	1,95	1,4	1,95	1,61	1,61	0	2	2	1,94	2
Water	2	2	2	1,99	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2
Pinus mugo	2	2	2	2	1,99	2	2	2	2	2	2	2	2	2	2	2	1,99	2	2	0	1,59	2
Picea abies	1,87	1,98	2	1,91	1,91	1,98	1,93	2	2	1,93	1,95	1,98	1,99	1,93	2	1,99	1,94	1,94	2	1,59	0	2
Dry vegetation	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1,98	2	2	2	2	2	0

Figure 23: Separability of pairs of classes for Pančavská louka orthomosaic from July the 26th 2023

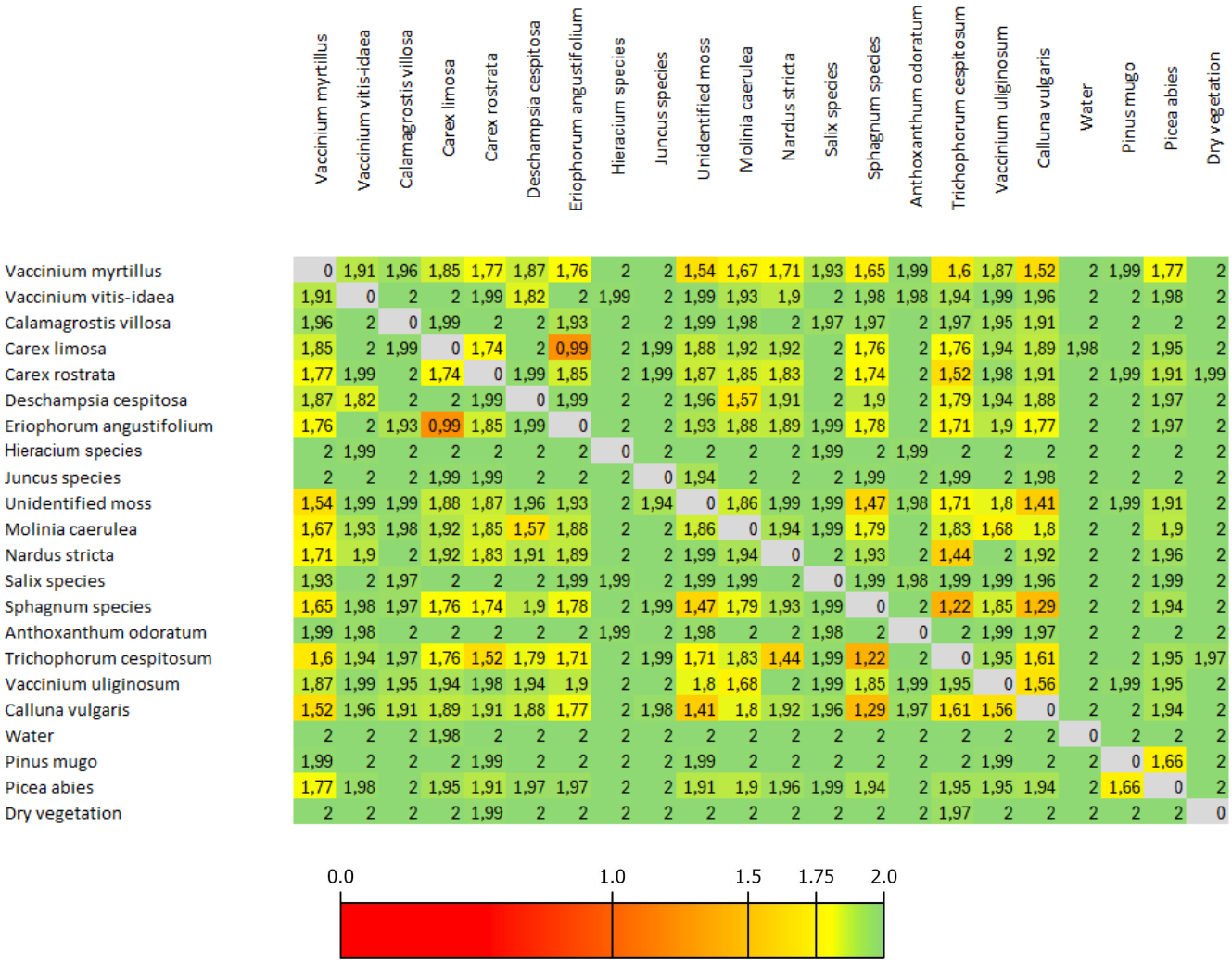
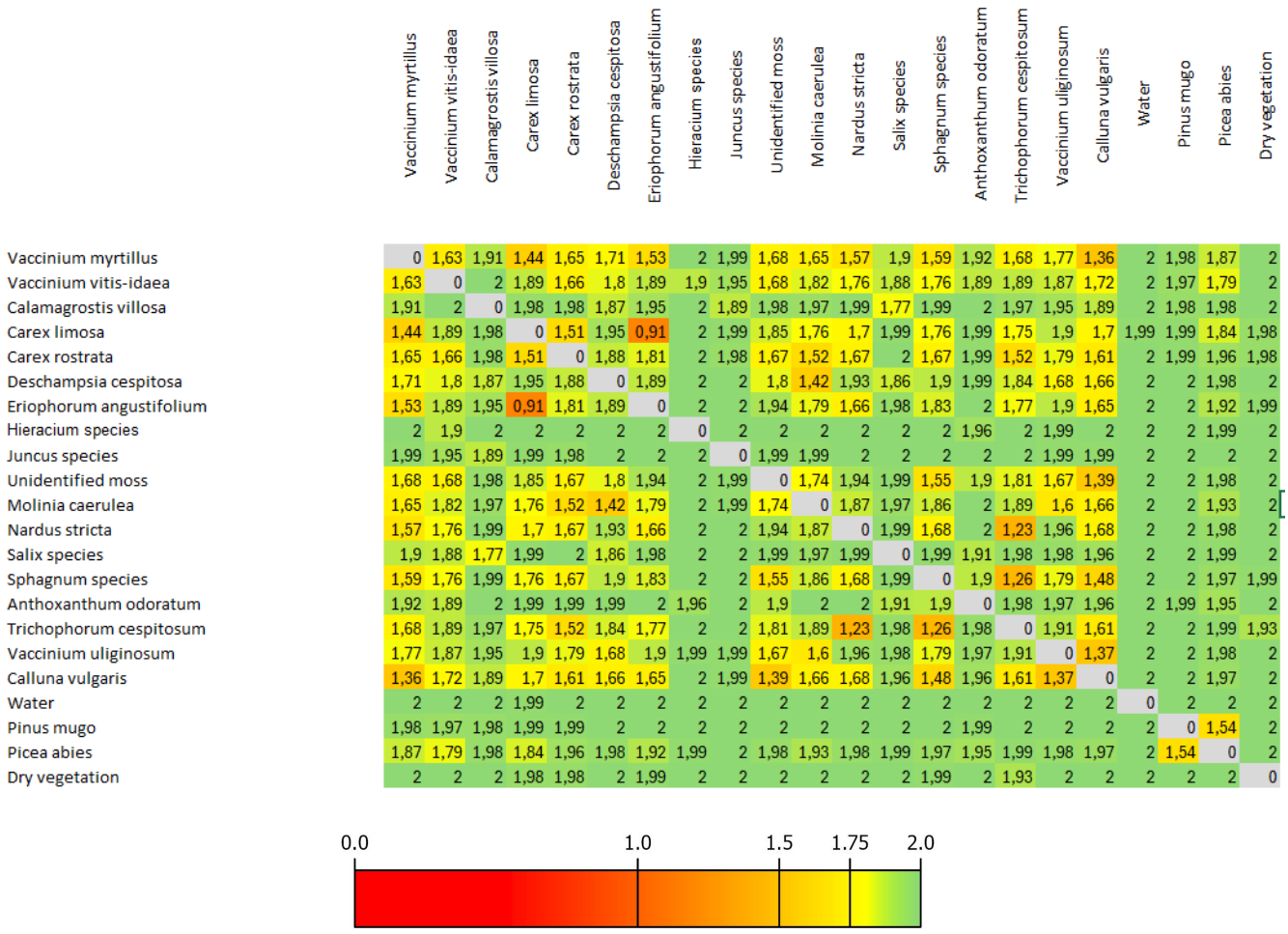


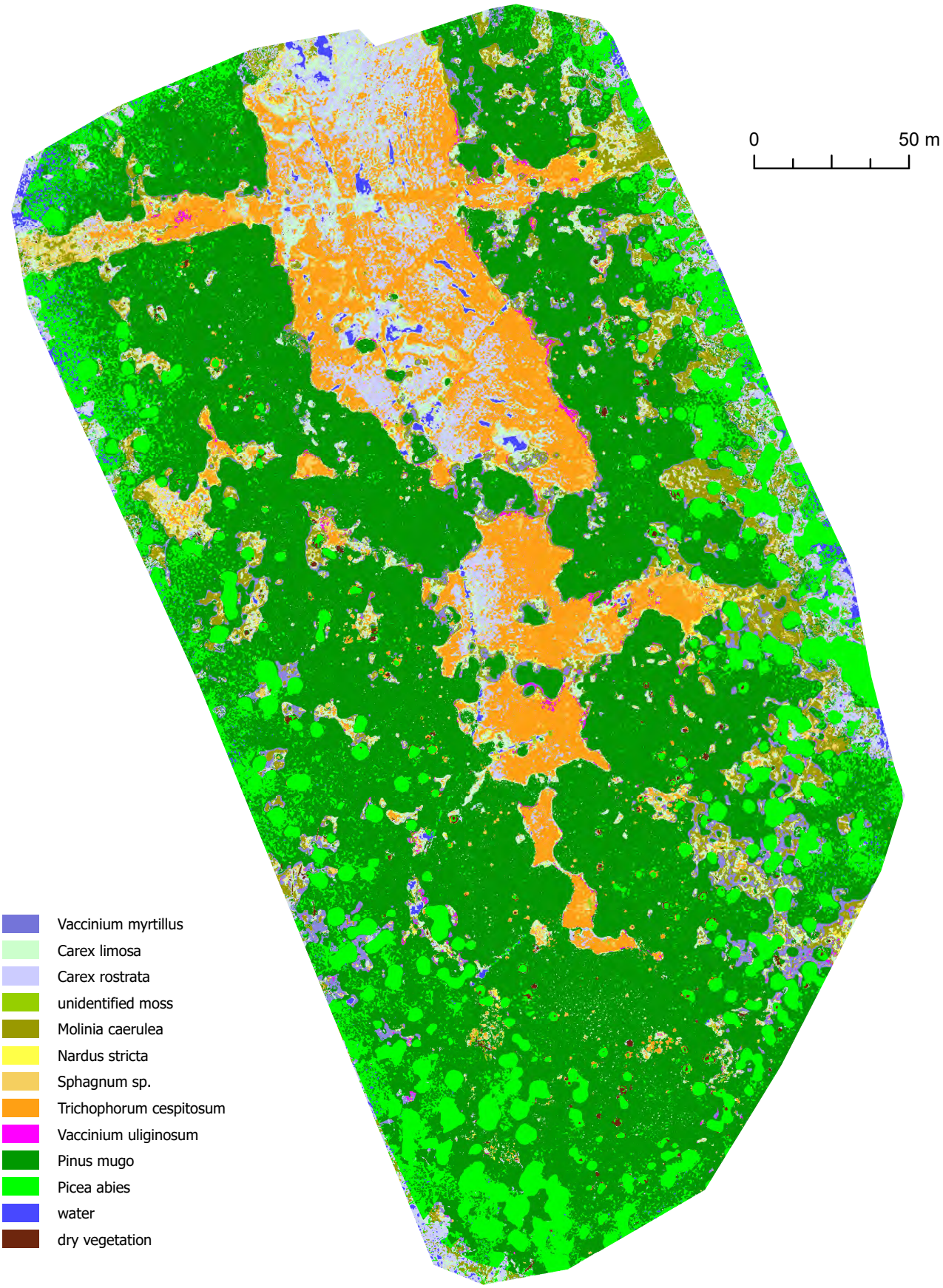
Figure 24: Separability of pairs of classes for Pančavská louka orthomosaic from July the 27th 2023



D Best maps of vegetation cover for each date and area of interest

HRANIČNÍ LOUKA

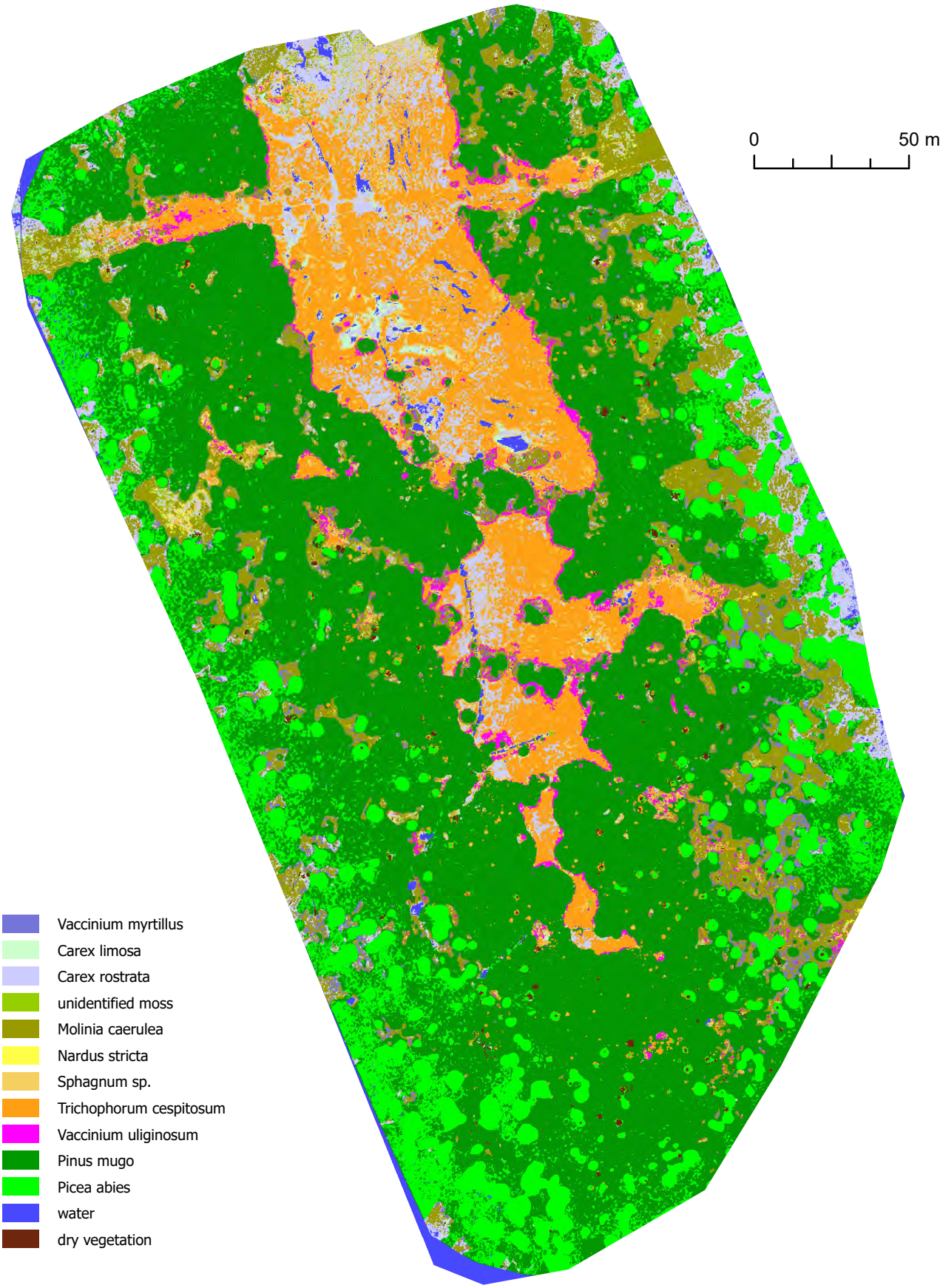
vegetation cover of peat bog in the Krkonoše NP (July 13th, 2023)
classified by Random forest, F-1 score 0.899



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

HRANIČNÍ LOUKA

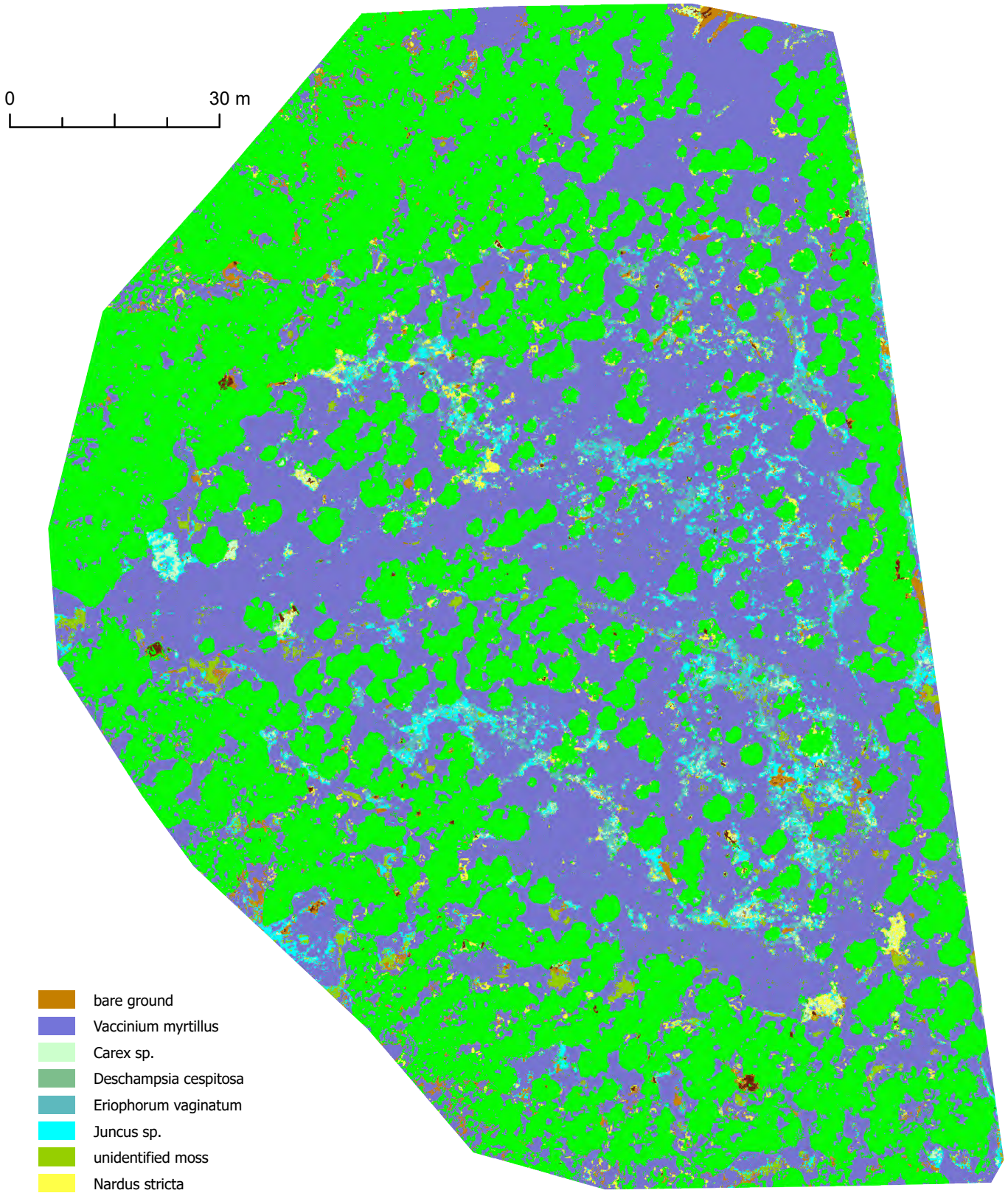
vegetation cover of peat bog in the Krkonoše NP (July 25th, 2023)
classified by Random forest, F-1 score 0.893



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

KYSELÉ KOUTY

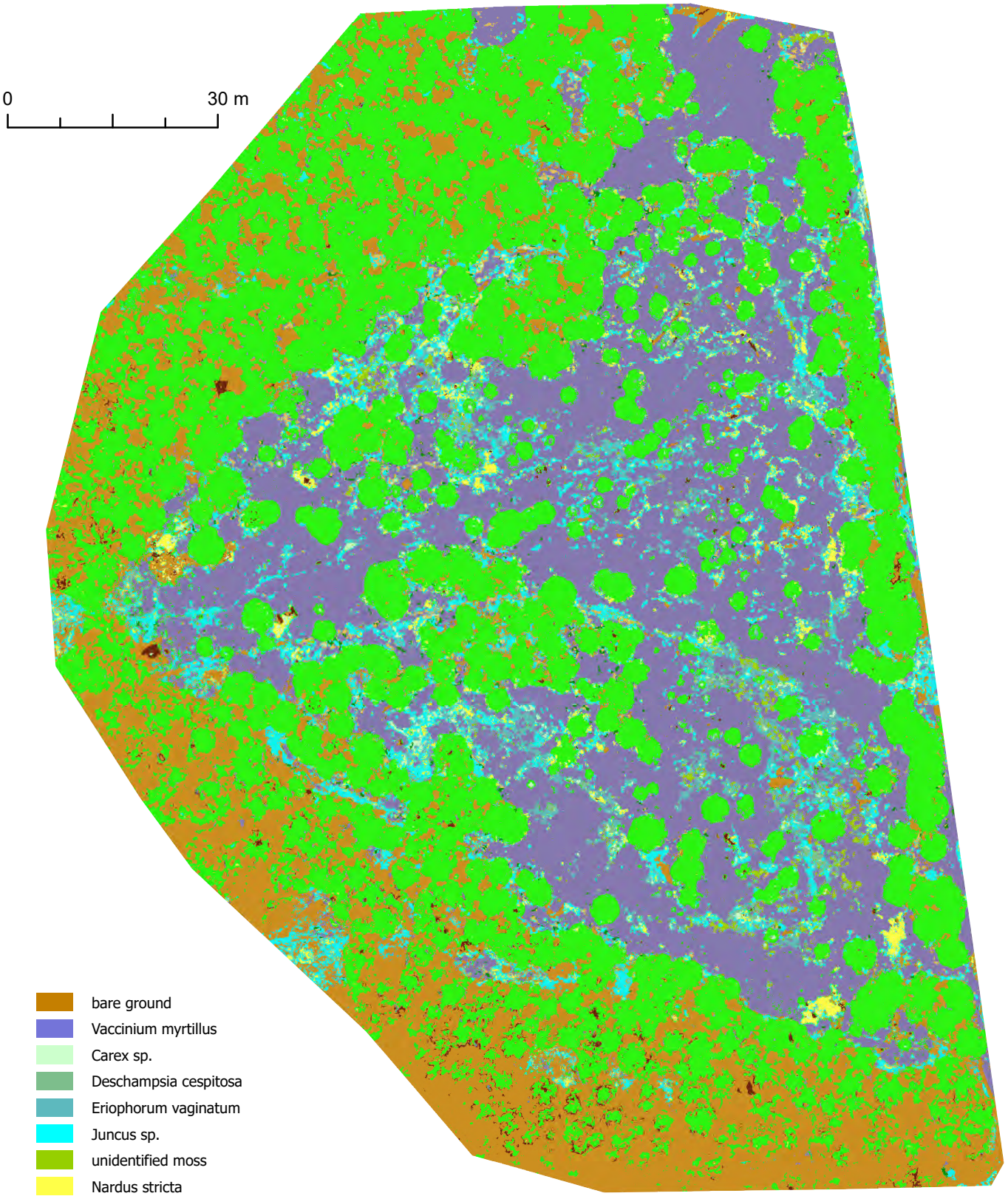
vegetation cover of peat bog in the Krkonoše NP (July 13th, 2023)
classified by Support vector machine, F-1 score 0.931



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

KYSELÉ KOUTY

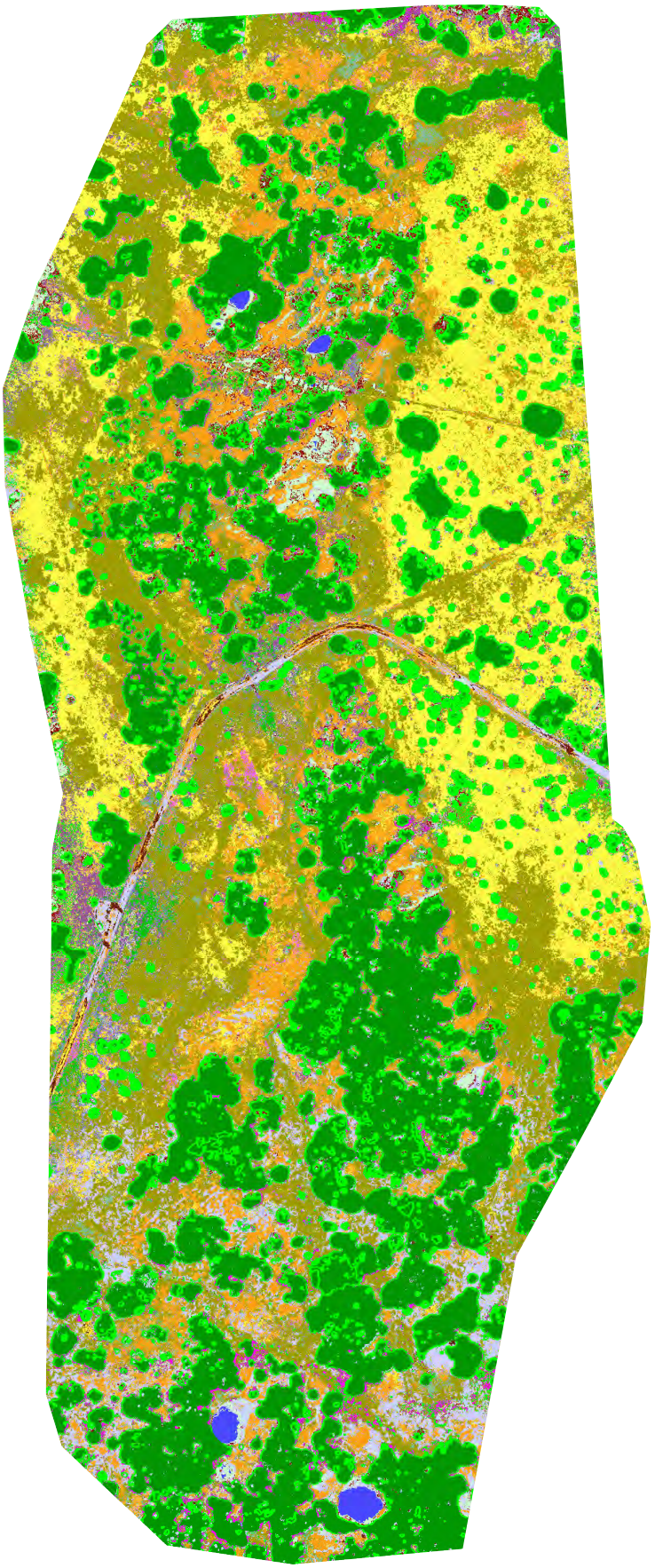
vegetation cover of peat bog in the Krkonoše NP (July 25th, 2023)
classified by Support vector machine, F-1 score 0.957



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

PANČAVSKÁ LOUKA

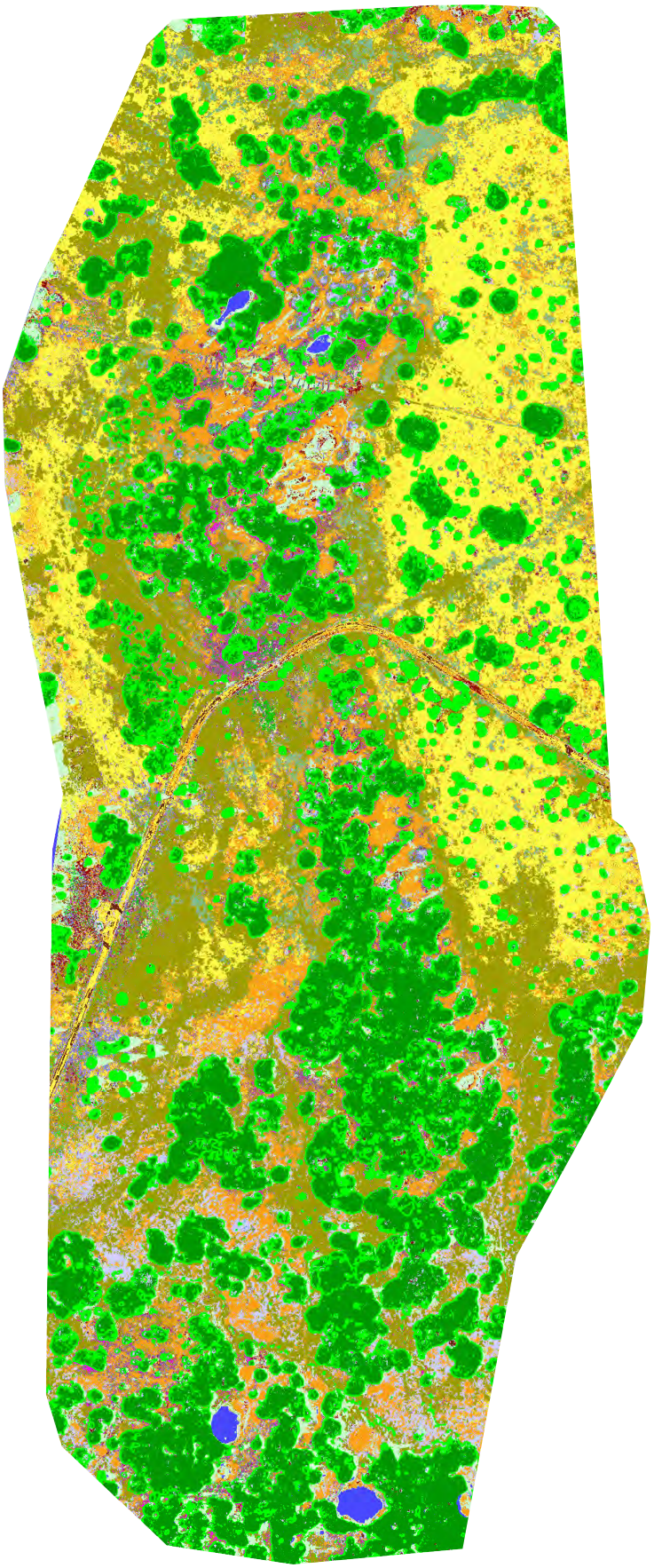
vegetation cover of peat bog in the Krkonoše NP (July 14th, 2023)
ensemble of Random forest and Support vector machine classifications, F-1 score 0.835



Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

PANČAVSKÁ LOUKA

vegetation cover of peat bog in the Krkonoše NP (July 26th, 2023)
ensemble of Random forest and Support vector machine classifications, F-1 score 0.807



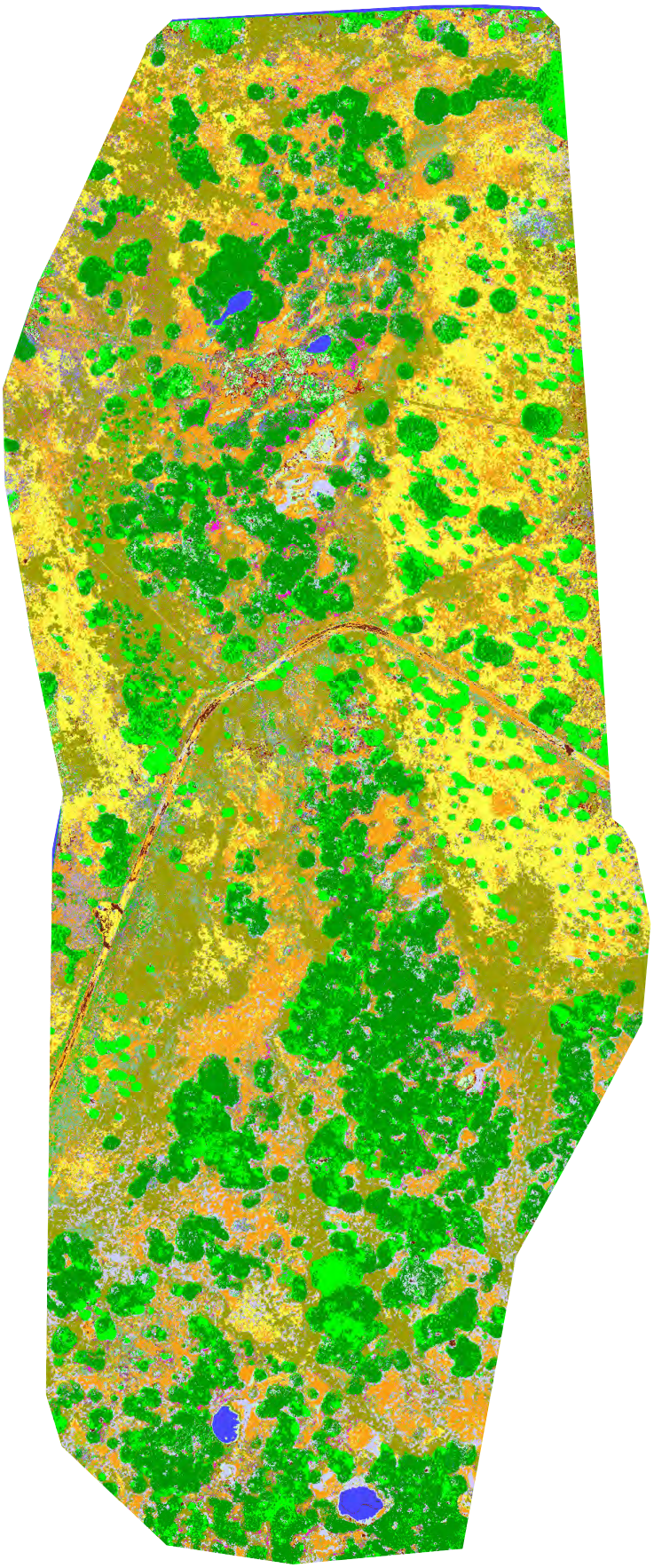
- Vaccinium myrtillus
- Vaccinium vitis-idáea
- Calamagrostis villosa
- Carex limosa
- Carex rostrata
- Deschampsia cespitosa
- Eriophorum angustifolium
- Hieracium sp.
- Juncus sp.
- unidentified moss
- Molinia caerulea
- Nardus stricta
- Salix sp.
- Sphagnum sp.
- Anthoxanthum odoratum
- Trichophorum cespitosum
- Vaccinium uliginosum
- Calluna vulgaris
- water
- Pinus mugo
- Picea abies
- dry vegetation

0 100 m

Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024

PANČAVSKÁ LOUKA

vegetation cover of peat bog in the Krkonoše NP (July 27th, 2023)
classified by Support vector machine, F-1 score 0.769



- Vaccinium myrtillus
- Vaccinium vitis-idáea
- Calamagrostis villosa
- Carex limosa
- Carex rostrata
- Deschampsia cespitosa
- Eriophorum angustifolium
- Hieracium sp.
- Juncus sp.
- unidentified moss
- Molinia caerulea
- Nardus stricta
- Salix sp.
- Sphagnum sp.
- Anthoxanthum odoratum
- Trichophorum cespitosum
- Vaccinium uliginosum
- Calluna vulgaris
- water
- Pinus mugo
- Picea abies
- dry vegetation

0 100 m

Part of master thesis by Adam Kulich
data by TILSPEC (www.tilspec.cz)
Department of Applied Geoinformatics and Carthography, Charles University, Prague
April 2024