

UNIVERZITA KARLOVA

Přírodovědecká fakulta

Katedra aplikované geoinformatiky a kartografie

Studijní program: Geografie

Studijní obor: Kartografie a geoinformatika



Bc. Miroslava Palúchová

**VLIV SPEKTRÁLNÍHO ROZLIŠENÍ NA KLASIFIKACI
KRAJINNÉHO POKRYVU V KRKONOŠSKÉ TUNDŘE**

**THE INFLUENCE OF SPECTRAL RESOLUTION ON LAND COVER
CLASSIFICATION IN KRKONOŠE MTS. TUNDRA**

Diplomová práce

Vedoucí práce: Mgr. Lucie Červená

Praha, duben 2018

Vysoká škola: Univerzita Karlova

Fakulta: Přírodovědecká

Katedra: Aplikované geoinformatiky a kartografie

Školní rok: 2017/2018

Zadání diplomové práce

pro Miroslavu Palúchovou

obor Kartografie a geoinformatika

Název tématu: Vliv spektrálního rozlišení na klasifikaci krajinného pokryvu v krkonošské tundře.

Zásady pro vypracování

Cílem diplomové práce je posoudit míru odlišitelnosti (separabilitu) tříd krajinného pokryvu v oblasti tundry Krkonošského národního parku na jednotlivých vlnových délkách s využitím leteckých hyperspektrálních dat AISA Dual.

Dílčími cíli práce jsou:

- 1) Provést rešerši metod spektrální separability a klasifikace vegetace v tundře.
- 2) Zhodnotit, která pásma AISA Dual dat jsou nevhodnější pro odlišení nedefinovaných tříd krajinného pokryvu v krkonošské tundře. Speciální důraz bude kladen na odlišení tříd jednoděložné vegetace – smilky tuhé, třtiny chloupkaté, bezkolence modrého a metlice trsnaté. Odpovídají vybrané vlnové délky pásmům aktuálních družicových senzorů?
- 3) Převzorkovat AISA Dual data na spektrální rozlišení družicových senzorů Landsat 8, Sentinel-2 a WorldView-2 a zjistit, zda jsou zkoumané třídy odlišitelné i při takto nižším spektrálním rozlišení.
- 4) Provést klasifikace, do kterých budou vstupovat pouze pásma významná pro odlišení tříd a zhodnotit jejich přesnost.

Rozsah grafických prací: dle potřeby

Rozsah průvodní zprávy: 50–100 stran (dle potřeby)

Seznam odborné literatury:

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.

Suchá R., Jakešová L., Kupková L., Červená L., 2016. Classification of vegetation above the tree line in the Krkonoše Mts. National Park using remote sensing multispectral data. *AUC Geographica*, 51(1), 113–129.

Bratsch S. N., Epstein H. E., Buchorn M., Walker D. A., 2016. Differentiating among Four Arctic Tundra Plant Communities at Ivotuk, Alaska Using Field Spectroscopy. *Remote Sensing*, 8(1), 51.

Virtanen T., Ek M., 2014. The fragmented nature of tundra landscape. *International Journal of Applied Earth Observation and Geoinformation*, 27, 4–12.

Van Aardt J., Wynne R., 2007. Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote Sensing*, 28 (2), 431–436.

Vedoucí diplomové práce: Mgr. Lucie Červená

Konzultant diplomové práce:

Datum zadání diplomové práce: 5. 4. 2017

Termín odevzdání diplomové práce: 20. 4. 2018

Platnost tohoto zadání je po dobu jednoho akademického roku.

.....

Vedoucí diplomové práce

.....

Garant oboru

V Praze dne 5. 4. 2017

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Svoluji k zapůjčení této práce pro studijní účely a souhlasím s tím, aby byla řádně vedena v evidenci vypůjčovatelů.

V Praze dne 18. dubna 2018

.....

Bc. Miroslava Palúchová

Poděkování

Ráda bych tímto poděkovala vedoucí mé práce Mgr. Lucii Červené za věnovaný čas, cenné rady a připomínky a poskytnutá data.

Vliv spektrálního rozlišení na klasifikaci krajinného pokryvu v krkonošské tundře

Abstrakt

Diplomová práce je zaměřena na specifikaci požadavků na spektrální rozlišení dat vstupujících do klasifikace a zodpovězení otázky, která pásma jsou stěžejní pro rozlišení tříd předem stanovené legendy. V práci jsou použita letecká hyperspektrální data senzoru AisaDUAL. Aplikovanou metodou výběru významných pásem byla diskriminační analýza provedena v IBM SPSS Statistics. Významná pásma se nacházela v intervalech 1500–1750 nm (začátek oblasti SWIR), 1100–1300 nm (delší vlnové délky NIR), 670–760 (*red-edge*) a 500–600 nm (zelené pásmo). Klasifikace vybraných pásem proběhla v ENVI 5.4 pomocí klasifikátoru Support Vector Machine a dosáhla celkové přesnosti 80,54 %, Kappa koeficient 0,7755. Součástí práce je také hodnocení vhodnosti dostupných družicových dat pro klasifikaci vegetace tundry z hlediska spektrálního rozlišení.

Klíčová slova: tundra, Krkonoše, klasifikace, spektrální rozlišení, separabilita tříd, diskriminační analýza, hyperspektrální data

The influence of spectral resolution on land cover classification in Krkonoše Mts. tundra

Abstract

The aim of this diploma thesis was to specify the spectral resolution requirements for classification and to identify the most important spectral bands to discriminate classes of the predefined legend. Aerial hyperspectral data acquired by AisaDUAL sensor were used. The method applied for the selection of the important bands was discriminant analysis performed in IBM SPSS Statistics. The most discriminative bands were found in intervals 1500–1750 nm (beginning of SWIR), 1100-1300 nm (longer wavelengths of NIR), 670-760 (red-edge) and 500-600 nm (green light). The classification of the selected bands was realized in ENVI 5.4 using the Support Vector Machine classifier, achieving overall accuracy of 80,54 %, Kappa coefficient 0,7755. The suitability of available satellite data for the classification of tundra vegetation in Krkonoše mountains based on spectral resolution was evaluated as well.

Keywords: tundra, Krkonoše, classification, spectral resolution, class separability, discriminant analysis, hyperspectral data

Obsah

Seznam obrázků a tabulek	9
1 Úvod	10
2 Literární rešerše a úvod do problematiky	12
2.1 Výzkum vegetace tundry metodami DPZ	12
2.2 Dosavadní výzkum krkonošské tundry	17
2.3 Analýza spektrální separability	20
2.4 Redukce dimenzionality dat	24
2.4.1 Analýza hlavních komponent.....	25
2.4.2 Diskriminační analýza.....	26
2.5 Spektrální rozlišení vybraných družic DPZ	30
2.5.1 WorldView-2.....	30
2.5.2 Sentinel-2	31
2.5.3 Landsat 8	31
3 Zájmové území	33
3.1 Charakteristika a vymezení	33
3.2 Legenda pro klasifikaci	34
4 Data a metodika	39
4.1 Použitá data a předzpracování	39
4.2 Simulace družicových dat.....	43
4.3 Diskriminační analýza v SPSS Statistics.....	45
4.4 Klasifikace	48
5 Výsledky.....	50
5.1 Pásma senzoru AisaDUAL.....	50
5.2 Pásma družicových senzorů	56
5.3 Odstranění pásem zatížených šumem.....	59
5.4 Výstupy klasifikace a hodnocení přesnosti	61
6 Diskuze	64
7 Závěr.....	69
Použitá literatura a zdroje	71
Seznam příloh	76

Seznam obrázků a tabulek

Obr. 1: Klasifikace krajinného pokryvu tundry z dat senzorů	14
Obr. 2: Prokletí dimenzionality	24
Obr. 3: Základní princip kanonické diskriminační analýzy	29
Obr. 4: Zájmové území.	34
Obr. 5: Křivky odrazivosti tříd.	36
Obr. 6: Fotografie tříd legendy a ukázky trénovacích ploch	37
Obr. 7: Fotografie tříd legendy a ukázky trénovacích ploch	38
Obr. 8.: Vlastní čísla diskriminačních funkcí referenčního souboru se všemi kategoriemi legendy.....	47
Obr. 9: Graf spektrálních pásem významných pro klasifikaci všech tříd legendy.	52
Obr. 10: Graf spektrálních pásem významných pro odlišení tříd jednoděložných travin, výsledek na základě základního referenčního souboru.....	53
Obr. 11: Graf spektrálních pásem významných pro odlišení tříd jednoděložných travin, výsledek na základě rozšířeného referenčního souboru.	54
Obr. 12: Projev šumu dat senzoru AisaDUAL v klasifikaci povrchu.	59
Obr. 13: Výsledek klasifikace SVM	63
Tab. 1: Spektrální pásma družice WorldView-2	30
Tab. 2: Spektrální pásma družic Sentinel-2	31
Tab. 3: Spektrální pásma senzoru OLI družice Landsat 8	32
Tab. 4: Přehled tříd legendy pro klasifikaci krkonošské tundry.	35
Tab. 5: Základní referenční soubor – trénovací a validační plochy.....	41
Tab. 6: Rozšířený referenční soubor kategorií smilka, třtina, bezkoleneček a metlice.....	41
Tab. 7. Počet pásem senzoru AisaDUAL vybraný diskriminační analýzou.....	51
Tab. 8: Rozdělení vybraných pásem do částí spektra.	51
Tab. 9: Spektrální pásma družicových senzorů seřazená sestupně podle důležitosti pro rozlišení tříd klasifikace.....	57
Tab. 10: Úprava výběru pásem na základě zatížení šumem.	60
Tab. 11: Přesnosti klasifikace pro jednotlivé vstupní soubory.	61
Tab. 12: Zpracovatelská a uživatelská přesnost tříd jednoděložných travin	62
Tab. 13: Zpracovatelská a uživatelská přesnost všech tříd.	62

| 1 Úvod

Vegetace v oblastech nad horní hranicí lesa Krkonošského národního parku (KRNAP) tvoří unikátní ekosystém alpínského bezlesí představující ojedinělou kombinaci severské, středoevropské a alpínské flóry. Tato oblast, vytvořená díky specifickým přírodním podmínkám Krkonoš, je v rámci národního parku označována jako oblast arкто-alpínské tundry a tvoří ji společenstva lišejníků, klečových porostů, trávníků a rašelinišť (Správa KRNAP, 2010).

Tundra jako ekosystém patří celosvětově k nejcennějším přírodním prostředím, zároveň je ale velmi citlivá na rušivé faktory a změny. Studium a monitoring vzácných biotopů alpínského bezlesí tvoří důležitou součást výzkumu vlivu různých environmentálních i lidských faktorů na horské ekosystémy. Pro mapování a analýzu vegetace rozsáhlých a těžce dostupných oblastí, jakou je i tundra, je v současnosti běžné využití metod dálkového průzkumu Země (DPZ). Dlouhodobým cílem průzkumu krkonošské tundry metodami DPZ je nalezení optimální kombinace vstupních dat a metod pro co nejpřesnější klasifikaci zdejší vegetace a její monitoring v čase. Tato diplomová práce na probíhající výzkum navazuje. Zabývá se hodnocením spektrálního rozlišení vstupních dat jakožto faktoru, který zásadně ovlivňuje výsledek klasifikace.

Hlavním cílem diplomové práce je specifikace požadavků na spektrální rozlišení dat vstupujících do klasifikace a zodpovězení otázky, která pásma jsou stěžejní pro oddělení definovaných tříd. Z výsledků bude plynout doporučení, jaká z dostupných dat DPZ jsou nejvhodnější pro klasifikaci vegetace nad horní hranicí lesa v KRNAP.

Podle výsledků studií popsaných v rešeršní části byla nejvyšší celková přesnost klasifikace vegetace tundry v KRNAP dosažena s využitím leteckých hyperspektrálních dat s velmi vysokým prostorovým rozlišením. Získání těchto dat je ale nákladné, data nejsou veřejně dostupná a nepořizují se pravidelně. Z těchto důvodů je potřebné zhodnotit možnosti využití běžně dostupných dat komerčních družic dálkového průzkumu a nalézt z nich nejvhodnější pro účely klasifikace do již stanovené, podrobné legendy. Je zřejmé, že velmi vysoké prostorové rozlišení znásobuje výskyt pixelů s homogenní vegetací a umožňuje tak jednoznačnější určení tříd. Na určení třídy vegetace se ale podílí taky spektrální charakteristika jednotlivých druhů.

S ohledem na různé časové rozlišení dostupných dat a vliv fenologické fáze na spektrální vlastnosti rostlin (podrobně řešeno v Roubalová, 2017) budou vstupní obrazová data simulována jednotně na základě hyperspektrálních leteckých snímků pořízených nad územím krkonošské tundry senzorem AisaDUAL. Tento přístup umožní porovnání výsledků pro různé typy dat DPZ bez nutnosti zohlednění vlivu termínu, kdy byla data pořízena. Hyperspektrální data budou převzorkována na spektrální rozlišení odpovídající vybraným současně dostupným sensorům dálkového průzkumu – WorldView-2, Sentinel-2 a Landsat 8.

2 Literární rešerše a úvod do problematiky

2.1 Výzkum vegetace tundry metodami DPZ

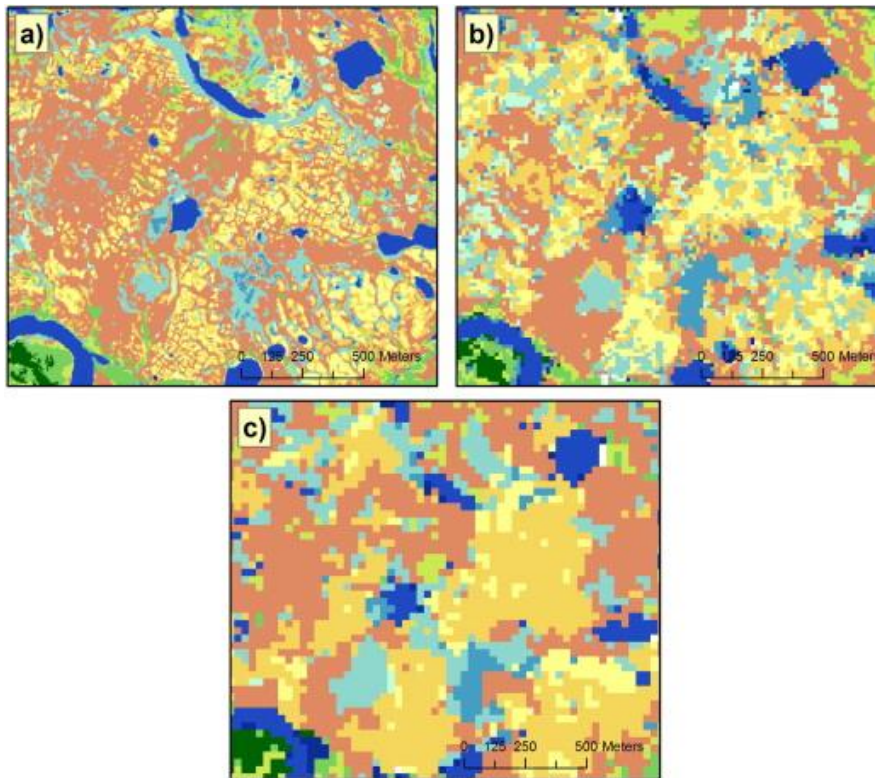
Monitoring vegetace v oblastech nad horní hranicí lesa vyžaduje spolehlivé a opakované mapování. Je nutné sledovat a zaznamenávat druhové složení a diverzitu, rozsah výskytu, úbytek nebo nárůst zastoupení jednotlivých druhů a další možné jevy, které mohou být ovlivněné probíhajícími environmentálními a klimatickými změnami. Zjišťování uvedených skutečností osobním průzkumem v terénu je s ohledem na časovou náročnost související se značnou rozlohou území a vysokými požadavky na správnou interpretaci takřka nemožné. V horském prostředí navíc hraje roli odlehlost a nepřístupnost daných oblastí, nestálost počasí a velmi krátké vegetační období.

Nedostatky vizuálního terénního monitoringu mohou být odstraněny použitím metod DPZ, kdy správně zvolenou metodou lze objektivně a opakovaně klasifikovat určené vegetační třídy, sledovat vývoj vegetace v čase a ze získaných poznatků odvodit důsledky na regionální anebo globální úrovni. Schopnost těchto metod správně rozpoznat a klasifikovat stanovené třídy záleží na řadě faktorů a v současnosti se zlepšuje zejména díky nově vypuštěným sensorům pořizujícím vysoce kvalitní družicová data, která jsou navíc často volně dostupná. Jejich využití dokumentuje mnoho publikovaných studií

zaměřujících se na oblasti tundry ve světě, ať už se jedná o tundru horskou, anebo nížinnou.

V dostupné odborné literatuře se nejčastěji vyskytují studie, které ke klasifikaci tundry využívají multispektrální družicová data se středním až vysokým prostorovým rozlišením. Schneider a kol. (2009) pro klasifikaci vegetace v deltě ruské řeky Lena použili data senzoru Landsat 7 ETM+ a řízenou klasifikaci podle minimální vzdálenosti. Klasifikace vegetace do 9 zvolených tříd dosáhla celkové přesnosti 77,8 %, Kappa indexu 0,74 a plocha území jednotlivých tříd získaná z výsledku klasifikace byla dále použita v analýze emisí metanu v deltě Leny. Data pořízená senzory Landsat, konkrétně Landsat 5 TM a Landsat 7 ETM+, byla použita také ve studii Fraser a kol. (2011), zaměřené na vyhodnocení dlouhodobých změn vegetace tundry na severu Kanady.

Virtanen a Ek (2014) uvádějí, že krajinný pokryv oblastí tundry je v porovnání s jinými biomy podstatně víc rozčleněný. Pro správné určení jednotlivých tříd tundry na severo-východě Ruska proto testovali několik způsobů klasifikace, a kromě běžně používaných dat senzoru Landsat s prostorovým rozlišením 30 m využili také družicových dat senzorů QuickBird (prostorové rozlišení 2,4 m) a Aster (15 m). Pro snímek QuickBird byla použita objektově-orientovaná analýza obrazu metodou řízené segmentace implementované v software eCognition, Aster a Landsat 5 TM data byla klasifikována řízenou pixelovou klasifikací s využitím klasifikátoru maximum likelihood. Přesnosti klasifikací ověřené na základě terénních dat ukázaly, že nejlepších výsledků bylo dosaženo v případě dat QuickBird (celková přesnost klasifikace v rozmezí 72–81 %). Výsledky klasifikace snímků se středním prostorovým rozlišením dosáhly vzájemně podobné celkové přesnosti pouze 50–55 %. Kromě klasifikací vegetace byla v této studii dále zkoumána struktura krajiny, a to analýzou velikosti plošek daných vegetačních tříd v různých rozlišeních. Analýza dat QuickBird s velmi vysokým prostorovým rozlišením podle očekávání odhalila vysoký počet ploch menších, než je prostorové rozlišení ostatních dvou senzorů (obr. 1). Výsledky této studie tak ukázaly, že s využitím dat s pouze vysokým prostorovým rozlišením se ztrácí značná část informace o vegetaci tundry, jelikož mnohé z tříd nelze v tomto rozlišení správně klasifikovat. Tím potvrdily jeden z nejdůležitějších limitujících faktorů pro použití optických družicových dat pro detailní mapování vegetačních tříd tundry, kterým je vysoká heterogenita vegetace a častý výskyt plošek příliš malých pro odlišení na snímku (Reese, 2011).



Obr. 1: Klasifikace krajinného pokryvu tundry z dat senzorů: a) QuickBird, b) Aster, c) Landsat 5 TM (zdroj: Virtanen a Ek, 2014 - upraveno)

Existují i další problémy, které jsou v souvislosti s výzkumem vegetace tundry metodami dálkového průzkumu diskutovány. Prostředí výskytu tundry, tedy nejvýše položené oblasti pohoří a polární oblasti, dále komplikuje použití optických senzorů vzhledem k časté a dlouhotrvající oblačnosti. Kromě toho, jak uvádí Chasmer a kol. (2014), je identifikace typů vegetačního pokryvu v těchto oblastech problematická také kvůli rychle se měnícím spektrálním charakteristikám vegetace a celkově nízkému spektrálnímu kontrastu. Řada autorů se proto snaží přesnost klasifikace zvýšit kombinací různých metod a typů dat.

Reese a kol. (2014) například klasifikují skandinávskou alpínskou a subalpínskou vegetaci s využitím multispektrálního družicového snímku družice SPOT 5, charakteristik reliéfu a dat odvozených z leteckého laserového skenování. Mezi charakteristiky reliéfu v této studii patří nadmořská výška, sklonitost a index vlhkosti. Z bodového mračna laserových dat byla odvozena výška vegetace, hustota porostu a další proměnné. Pro klasifikaci byl vybrán učící algoritmus *random forest* (RF), který byl testovaný pro různé kombinace proměnných vstupujících do klasifikace. Výsledky

klasifikací ukázaly, že oproti klasifikaci samotného družicového snímku dosáhla klasifikace kombinace snímku a dat reliéfu o 8,2 % vyšší celkové přesnosti. Celková přesnost klasifikace kombinace snímku, dat reliéfu a dat LiDAR pak byla o 14,2 % lepší než v případě klasifikace pouze snímku. Díky laserovým datům se povedlo lépe odlišit vegetační třídy, které mají podobné spektrální charakteristiky, což znemožňuje jejich identifikaci pouze z optických dat.

Pro mapování funkčních typů rostlin v arktické tundře na Aljašce využili Langford a kol. (2016) kombinaci multispektrálních dat s vysokým prostorovým rozlišením (~2 m) pořízených senzorem WorldView-2, digitálního modelu terénu vytvořeného z dat LiDAR a opakovaného terénního průzkumu vegetace. Průzkum na místě spočíval v odhadu procentuálního zastoupení jednotlivých funkčních typů rostlin na vybraném počtu ploch o rozloze 1 m². Autoři studie nejdříve popsali statistické vztahy mezi spektrálními a topografickými údaji získanými dálkovým průzkumem a rozložením funkčních typů rostlin v rámci vzorových ploch vegetace. Tyto vztahy potom uplatnili pro modelování rozložení funkčních typů rostlin na celé ploše oblasti, na kterou se studie zaměřuje. Algoritmus pro segmentaci vegetace na celé ploše zahrnoval shlukovou analýzu k-means a interpolaci IDW. Studie představuje dvě verze map rozložení funkčních typů rostlin: verzi vypočtenou z jednoho družicového snímku pořízeného uprostřed vegetačního období a verzi z několika snímků pořízených v průběhu vegetačního období v různých fenologických fázích rostlin. Model, do kterého vstupovalo vícero snímků, dosáhl vyšší přesnosti, což podle autorů dokazuje důležitost fenologických charakteristik, jakou je například měnící se hodnota indexu NDVI, pro zlepšení schopnosti algoritmu odlišit jednotlivé funkční typy rostlin. Ve vztahu k vegetaci tundry a její klasifikaci využili fenologii také Beamish a kol. (2017), kteří zkoumali vzájemnou spektrální variabilitu rostlin a její změny v průběhu vegetačního cyklu.

Problém časté oblačnosti vyskytující se nad tundrovou vegetací řeší využití radarových dat, kdy pořízení snímků je nezávislé na slunečním záření díky aktivnímu ozařování povrchu elektromagnetickými paprsky. Interpretace radarových snímků potom nezáleží na spektrálních charakteristikách, ale na způsobu pořízení snímku a fyzikálních vlastnostech vegetace. Potenciál kombinace radarových dat senzorů TerraSAR-X a Radarsat-2 s optickými daty Landsat 8 pro klasifikaci vegetace tundry rozebírá Ullmann a kol. (2014). V této studii testují řízenou i neřízenou klasifikaci, a to pro jednotlivé typy dat i pro všechny jejich kombinace. Nejlepších výsledků dosáhla řízená klasifikace

provedená nad polarimetrickými radarovými daty a optickými daty (celková přesnost klasifikace 87 %). Zapojení radarových dat přineslo výrazné zlepšení v rozlišení vegetace na základě výšky porostu, a také umožnilo uspokojivou identifikaci mokřadů.

Atkinson a Treitz (2012) ve své studii využili pro klasifikaci tundrové vegetace na území v kanadském teritoriu Nunavut kombinaci spektrálních a ekologických charakteristik. Aplikací několika technik vícerozměrné statistické analýzy na data získaná v terénu definovali shluky odpovídající vegetačním třídám. Následně byla analyzována spektrální separabilita shluků na základě vzdálenosti Jeffries-Matusita. Vzájemnou kombinací shluků s nízkou spektrální separabilitou vznikl finální set shluků, který definoval třídy pro následující klasifikaci. V této studii bylo zvoleno použití klasifikace *maximum likelihood*, a to na multispektrální data družice IKONOS s prostorovým rozlišením 4 m. Celková přesnost klasifikace se v závislosti na území pohybovala v rozmezí 69–79 %. Autoři svým výzkumem dospěli k tvrzení, že kombinací statistických metod používaných v ekologii a metod DPZ lze klasifikovat vegetaci tundry tak, aby výsledky lépe zohledňovaly ekologické vlastnosti jednotlivých tříd a dávaly tak ekologicky větší význam.

2.2 Dosavadní výzkum krkonošské tundry

Probíhající výzkum krkonošské tundry metodami DPZ Přírodovědeckou fakultou UK je popsán v článcích Suchá a kol. (2016) a Kupková a kol. (2017). V prvním článku Suchá a kol. (2016) klasifikuje multispektrální data s různým prostorovým a spektrálním rozlišením. Srovnává 3 typy dat: letecké ortorektifikované snímky s prostorovým rozlišením 12,5 cm pořízené ve čtyřech spektrálních pásmech, družicové snímky družice WorldView-2 s prostorovým rozlišením 2 m a osmi spektrálními pásmy, a družicová data Landsat 8 s prostorovým rozlišením 30 m pořízená v sedmi spektrálních pásmech. Ortofoto snímky byly klasifikovány objektovou klasifikací, pro WorldView-2 data byla použita objektová i pixelová klasifikace, data Landsat 8 pak byla klasifikována pouze pixelově. V rámci základní podrobné legendy klasifikace bylo rozlišeno 12 tříd povrchu. Z důvodu častého výskytu jednotlivých tříd na plochách menších než pixel Landsatu 8 byla pro klasifikaci těchto dat vytvořena zjednodušená legenda obsahující 8 tříd, do které pak byla pro srovnání klasifikována i ostatní data.

Výsledky klasifikací potvrdily předpoklad, že pro dosažení vysoké celkové přesnosti klasifikace krkonošské tundry je zásadním parametrem vysoké prostorové rozlišení dat, a to z důvodu velké rozmanitosti vegetace a malých rozměrů ploch, na kterých se jednotlivé vegetační typy vyskytují. Jako nejúspěšnější se ukázala objektová klasifikace ortofot, která v případě obou typů legendy dosáhla nejvyšší celkové přesnosti (72–84 %). Vhodnost objektového přístupu pro klasifikaci snímků s velmi vysokým prostorovým rozlišením je podmíněna zejména skutečností, že plochy jednotlivých vegetačních typů, i když nejsou plošně rozsáhlé, jsou většinou jasně oddělené a nedochází k jejich smíšenému výskytu. Nejlepší výsledek objektové klasifikace dat WorldView-2 byl 68 %, v případě pixelové klasifikace těchto dat pak byla celková přesnost pouze 61 %. Vytvoření zjednodušené legendy optimalizované pro data Landsat 8 přineslo vysokou přesnost klasifikace těchto dat, až 78 %, a potvrdilo se tak, že tato data jsou vhodná k výzkumu základních vegetačních tříd v krkonošské tundře. Detailní legenda ale samozřejmě druhovou a prostorovou rozmanitost místní vegetace vystihuje lépe.

Hodnocením a srovnáním vhodnosti leteckých hyperspektrálních dat a dat Sentinel-2A pro klasifikaci vegetace v tundře v Krkonoších se zabývá článek Kupkové a kol. (2017). Klasifikovaná hyperspektrální data s velmi vysokým prostorovým

rozlišením (1-5 m) byla pořízena senzory AISA Dual a APEX. Pro jejich klasifikaci byla použita legenda s 11 kategoriemi. Data Sentinel-2A, u kterých je velikost pixelu podstatně větší (10 a 20 m), byla klasifikována také do zjednodušené legendy stejné jako v Suchá a kol. (2016). Samotným klasifikacím předcházelo rozsáhlé předzpracování dat a pořízení a úprava referenčních dat. Hyperspektrální data byla klasifikována objektivě i pixelově, Sentinel-2A pak pouze pixelově. Pro dosažení co nejlepších výsledků byly jednotlivé klasifikační algoritmy testovány opakovaně s různým nastavením parametrů a různým počtem pásem vstupujících do klasifikace. Klasifikována byla také transformovaná pásma získaná analýzou hlavních komponent (PCA).

Celkově nejlepších výsledků dosáhla pixelová klasifikace 40 PCA pásem dat AISA Dual algoritmem Support Vector Machines (SVM), kdy celková přesnost klasifikace dat AISA Dual dosáhla 84,3 % a u dat APEX byla jenom o 1,7 procenta nižší. Přesnost nejlepší objektové klasifikace hyperspektrálních dat byla jenom o 4 procenta nižší než přesnost nejlepší pixelové klasifikace. Jako opodstatněné se ukázalo zavedení zjednodušené legendy pro klasifikaci dat Sentinel-2A. Přesnost klasifikace těchto dat do detailní legendy byla až o 26 procent nižší než nejlepší výsledek hyperspektrálních dat, při použití zjednodušené legendy ale přesnost dosáhla 77,7%. Tato studie celkově potvrdila vhodnost zkoumaných dat pro klasifikaci vegetace tundry, a zároveň dokázala nutnost použití správné legendy a klasifikační metody.

Z hlediska přesnosti klasifikace jednotlivých tříd, v případě obou studií dosáhla nejvyšší uživatelskou a zpracovatelskou přesnost klasifikace nevegetačních tříd, jakými jsou vodní a antropogenní plochy anebo kamenná moře, a to pro všechny typy dat. S vysokou přesností byla také z dat s vysokým prostorovým rozlišením klasifikována třída kosodřeviny. V případě legendy zjednodušené pro prostorové rozlišení Landsatu 8 byla tato třída rozdělena podle hustoty porostu na kosodřevinu hustou a kosodřevinu řídkou. Správná klasifikace třídy řídké kosodřeviny dosáhla horších výsledků z důvodu smíšeného výskytu kosodřeviny a travnatých porostů v rámci jednoho pixelu. Třída subalpínské brusnicové vegetace byla klasifikována přesněji z multispektrálních dat, na druhou stranu hyperspektrální data se ukázala jako vhodnější pro třídy vřesoviště, mokřady a rašeliniště. Největší komplikací bylo podle očekávání rozlišení jednotlivých druhů trav, jejichž spektrální signál je podobný. V tomto případě byla přesnější klasifikace hyperspektrálních dat, ze kterých bylo částečně možné některé druhy trav od sebe odlišit.

K výzkumu krkonošské tundry metodami DPZ dále přispívají také absolventské práce studentů Katedry aplikované geoinformatiky a kartografie Přírodovědecké fakulty. Objektově-orientovanou klasifikací vegetace nad horní hranicí lesa v KRNAP se zabývá práce Jakešové (2014). Porovnává v ní dva přístupy k objektově-orientované klasifikaci ortofota s vysokým prostorovým rozlišením (12,5 cm) do 13 tříd. Všechny třídy se podařilo odlišit pouze pomocí přístupu *Example-based* a celková přesnost klasifikace dosáhla v tomto případě 75,95 %. Nejlepších výsledků klasifikace bylo dosaženo u tříd kosodřevina, rašeliniště a subalpínská brusnicová vegetace, naopak nejhorší výsledky byly zaznamenány pro třídy subalpínských vysokostéblých trávníků. Tyto výsledky se shodují s těmi z výše popsaných článků. Jiná studentská práce, Roubalová (2017), klasifikuje vybrané vegetační kategorie krkonošské tundry z dat senzoru Sentinel-2A. Práce je zaměřena na zlepšení přesnosti klasifikace pomocí multitemporálního přístupu, který vychází z předpokladu, že zohlednění změn spektrálního projevu vegetace v průběhu vegetační doby může přispět k odlišení vegetačních kategorií. Tento předpoklad se potvrdil pouze částečně, jelikož celková přesnost klasifikace byla podobná jako v případě klasifikace jedné scény (53,4 % a 51,2%).

Kromě uvedených článků a prací se klasifikací horské vegetace v Krkonoších věnuje také Marcinkowska a kol. (2014). Vegetace v krkonošském pohraničí byla klasifikována do 15 kategorií s využitím hyperspektrálních dat senzoru APEX a klasifikačního algoritmu SVM. Senzor APEX pořizuje data v 288 spektrálních pásmech v rozsahu 0,3937 – 2,4673 μm s prostorovým rozlišením 1,75 m. Pro klasifikaci bylo použito všech 288 pásem a postupně byly testovány různé typy jádrových funkcí (kernel type). Hodnocení přesnosti klasifikace proběhlo s využitím terénních GPS dat a vektorových mapových podkladů získaných v předchozích studiích. Nejlepších výsledků klasifikace bylo dosaženo s použitím lineární funkce a celková přesnost klasifikace dosáhla 79,13 %, Kappa koeficient 0,77. Použitá metoda v kombinaci s danými daty byla vyhodnocena jako vhodný nástroj pro identifikaci a monitorování vegetačních typů v horském prostředí nad hranicí lesa. Další studií využívající hyperspektrální data senzoru APEX je Jarocińska a kol. (2016), cílená na analýzu zdravotního stavu vegetace v oblastech krkonošského bezlesí.

2.3 Analýza spektrální separability

Moderní programy pro zpracování obrazových dat nabízejí řadu algoritmů pro řízenou klasifikaci. Abychom dosáhli vysoké přesnosti klasifikace, je nutné předem znát charakteristiky spektrální odrazivosti klasifikačních tříd. Podle naměřených spekter lze posoudit míru jejich vzájemné odlišitelnosti (separabilitu) na jednotlivých vlnových délkách a tím splnit cíl diplomové práce – určit ideální spektrální rozlišení vstupních dat, resp. vybrat spektrální pásma vhodná pro zvýšení přesnosti klasifikace.

Velmi podobný cíl popisuje Mureriwa a kol. (2016), kteří úspěšně určili konkrétní vlnové délky spektrálních dat stěžejní pro odlišení spekter invazní rostliny *Prosopis* a čtyř druhů okolní vegetace. Výzkum byl založen na využití spektroskopických dat z terénu v kombinaci s nově vyvinutým algoritmem *guided regularized random forest* (GRRF). Tato metoda je zdokonalením běžně používané metody *random forest* (RF) – náhodný les (Breiman, 2001), která řeší problém zpracování hyperspektrálních dat označovaný jako prokletí dimenzionality (*curse of dimensionality*). Jedná se o skutečnost, že počet trénovacích vzorků potřebných pro učení konkrétního klasifikátoru exponenciálně roste s počtem spektrálních pásem (Hsu, 2007). U hyperspektrálních dat je tak většinou počet trénovacích vzorků nedostatečný ve srovnání s velkým množstvím pásem, což způsobuje nepřesný a nespolehlivý výsledek klasifikace. Metoda RF při učení vytváří více rozhodovacích stromů, které vedou k různým výsledkům, a následně použije pro určení dané třídy jejich modus. Tato metoda ale nedokáže zvolit optimální počet proměnných (pásem), které vedou k nejvyšší přesnosti klasifikace. Metoda GRRF proto implementuje výsledky běžné RF do procesu výběru proměnných (*feature selection*) a pro konstrukci klasifikačního modelu používá pouze vybranou podmnožinu relevantních proměnných (Deng, Runger 2013).

Analýza dat popsaná v Mureriwa a kol. (2016) začíná odstraněním celkem 325 vlnových délek, na kterých se projevil šum způsobený absorpcí vody atmosférou, ze spekter vegetačních druhů. Na zbývajících 1825 pásem byly postupně použity metody RF a GRRF. RF kromě samotného procesu rozhodovacích stromů poskytuje také informaci o důležitosti jednotlivých proměnných, nazývanou *Gini* významnost. Tato hodnota významnosti je založená na výpočtu *Gini* indexu: vždy, když se daná proměnná použije při dělení množiny výsledků (klasifikačních tříd) ve stromě, zaznamená se hodnota,

o kolik se snížila nečistota množiny, tj. o kolik je množina homogennější. Suma těchto hodnot je pak významnost proměnné pro daný strom a průměrováním sum přes množinu všech stromů je každé proměnné přidělena konečná hodnota významnosti (Gábik, 2010). Kvůli vysoké korelaci sousedících pásem při spektrálním intervalu 1 nm je ale zřejmé, že mnoho vlnových délek bude mít totožnou hodnotu *Gini* významnosti. Následně použitá metoda GRRF využívá významnost proměnných pro eliminaci nerelevantních a přebytečných pásem a pro výběr těch, které jsou optimální pro rozlišení daných vegetačních druhů. Výsledných 11 pásem pak bylo vstupem do klasifikátoru RF. Celková přesnost klasifikace se zvýšila z 79,19 % (Kappa index 0,7201) při použití všech vlnových délek dat na 88,59 % (Kappa index 0,8524) při klasifikaci pouze vybraných vlnových délek.

Jiná studie zabývající se otázkou separability spekter, Van Aardt a Wynne (2007), zkoumá odlišitelnost tří druhů borovic na území amerického státu Virginia z leteckých hyperspektrálních dat senzoru AVIRIS. Spektra byla získána z obrazu, z oblastí s homogenním výskytem daného druhu ověřeným a geolokalizovaným v terénu. Byly provedeny dvě samostatné analýzy, a to pro spektra získaná z jednotlivých pixelů, a pro spektra z plošek 3x3 pixely. Pro redukci dimenzionality dat na maximálně 10 spektrálních pásem stěžejních pro rozlišení druhů byla použita lineární diskriminační analýza (LDA). Vybraná pásma byla vstupem pro klasifikaci. Celková přesnost klasifikace dosáhla hodnot 65 % pro spektra ze samostatných pixelů a 85 % pro spektra z okolí 3x3. Nižší přesnost klasifikace v prvním případě byla odůvodněna možnými odchylkami v poloze pixelů způsobenými chybami během předzpracování dat, a také vysokým prostorovým rozlišením vstupních dat (3,4 m). Autoři uvádí, že spektra získaná z plochy 3x3 pixely lépe vystihují spektrální charakteristiky seskupení stromů, protože zohledňují odraz od různých částí vegetace. Vlnové délky vyhodnocené jako klíčové pro rozpoznání druhů se převážně nacházely v oblastech viditelného a blízkého infračerveného záření.

Lineární diskriminační analýza je vícerozměrná statistická metoda. Její dvě základní funkce jsou 1) určení vlivu sledovaných proměnných na odlišení jednotlivých skupin a sestavení klasifikačního pravidla pro zařazení jednotek do těchto skupin, 2) klasifikace neznámých objektů do skupin. Identifikací proměnných významných pro klasifikaci se zabývá kanonická diskriminační analýza, jejíž cílem je nalezení takové lineární kombinace proměnných, která by co nejlépe separovala uvažované skupiny tak,

aby vnitroskupinová variabilita byla co nejmenší a meziskupinová variabilita co největší (Jarkovský a Littnerová, 2011). Tento způsob byl použit v mnoha studiích jako prostředek snížení dimenzionality dat a zpřesnění výsledků klasifikace vegetace, a to i v případě většího počtu klasifikačních tříd (Prospere a kol., 2014, Pomahačová, 2012, Cingolani a kol., 2004 a další).

Spektrální odrazivost vegetace ovlivňují různé parametry, jako je vlhkost, vertikální struktura rostlin, výskyt podrostu, buněčná struktura, a další. Tyto parametry jsou proměnlivé v čase, a i jejich vliv na odrazivost se v průběhu vegetačního období mění. Tato skutečnost byla jedním ze závěrů studie Bratsch a kol. (2016), která dokázala, že oblasti spektra stěžejné pro odlišení vegetace se mění s fenologickou fází. Hlavním cílem této studie bylo zhodnocení možností pozemní spektroskopie pro odlišení čtyř vegetačních společenstev aljašské tundry. Pro snížení počtu přebytečných úzkých hyperspektrálních pásem vstupujících do klasifikace a nalezení těch, které jsou pro klasifikaci vegetace stěžejní, byl použit model pro výběr proměnných a redukcii dimenzionality nazývaný *sparse partial least squares* (SPLS). 10 % všech pásem s nejlepšími výsledky ze SPLS (8 pásem pro data ze začátku vegetační doby, 12 pásem pro vrcholní fázi vegetační doby) pak bylo vstupem do klasifikace LDA. Vybraná pásma se převážně nacházela v infračervené části spektra. Celková přesnost klasifikace byla vyšší v případě dat z vrcholní fáze vegetační doby (94 %, resp. 84 %).

Beamish a kol. (2017) uvádí, že spektrální charakteristiky jednotlivých druhů vegetace tundry jsou často velmi podobné a spektrální separaci zde označuje jako náročnou. Dále uvádí, že zlepšení by mohl přinést rozdílný přístup k různým fenologickým fázím rostlin a cílený výběr pouze těch vlnových délek, na kterých je v dané fenologické fázi separabilita druhů nejvyšší. Ve studii byla zkoumána spektrální variabilita vegetačních druhů tundry ve třech fenologických fázích – na začátku, uprostřed a na konci vegetační doby. Spektrální charakteristiky byly získány pozemním spektroskopickým měřením na úrovni porostu. Cílem studie bylo určit, která fenologická fáze je nejvhodnější pro co nejpřesnější klasifikaci vegetace a která spektrální pásma jsou pro tento účel stěžejní. Dalším cílem byla také simulace spektrálních rozlišení senzorů EnMAP a Sentinel-2 a hodnocení jejich využitelnosti ve srovnání s daty z terénu. Pro určení rozlišujících vlnových délek byl pro všechna tři spektra odrazivosti (terénní, EnMAP, Sentinel-2) vypočítán tzv. InStability Index (ISI), také nazývaný Separability Index (Ghulam a kol., 2016). Index ISI pochází z klasifikační metody označované stable

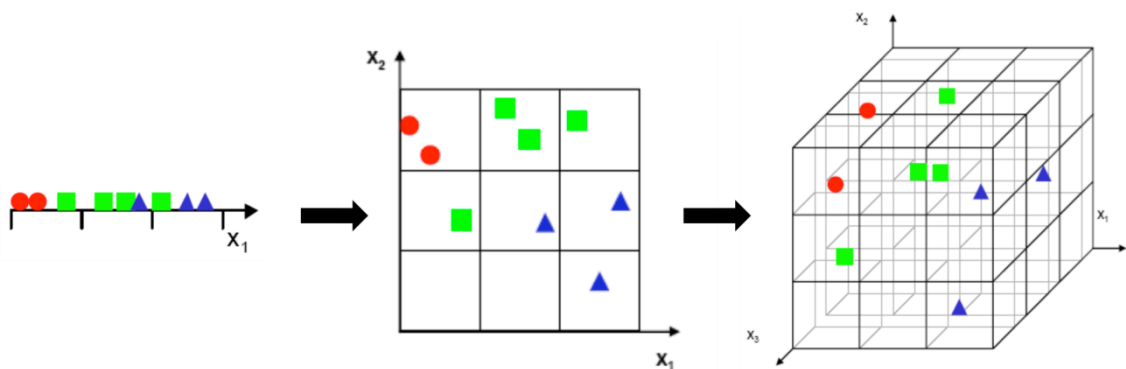
zone unmixing (SZU), která je modifikací analýzy smíšených pixelů (spectral mixture analysis – SMA) používané pro určení dílčího zastoupení koncového členu v pixelu. Je podílem vnitroskupinové a meziskupinové variability předem určených vegetačních skupin (tříd) a počítá se pro každou vlnovou délku spektrálního rozsahu dat. Nízké ISI hodnoty indikovali nízkou vnitroskupinovou spektrální variabilitu, a naopak vysokou meziskupinovou spektrální variabilitu, což je žádoucí stav pro výběr rozlišujících vlnových délek. Ty byly vybrány nalezením lokálních minim hodnot ISI. Výsledky studie ukázaly, že pro tři z pěti kategorizovaných druhů rostlinných společenstev (suchá tundra, vlhká acidická tundra, vlhká neacidická tundra, mechová tundra, keřová tundra) je nejvhodnější fenologickou fází pro spektrální odlišení druhů fáze senescence, tedy konec vegetačního období. Jelikož tyto tři druhy mají v studované oblasti většinové zastoupení, autoři označují fázi senescence jako celkově nejvhodnější. Vlnové délky stěžejní pro identifikaci jednotlivých druhů se nacházely v modrém, červeném a red-edge pásmu spektra. Simulací spektrálních charakteristik družicových senzorů bylo zjištěno, že fenologický přístup k analýze vegetace může přinést zlepšení i u tohoto typu dat. Analýza spektra senzoru EnMap určila jako klíčovou fázi senescence, zatímco data Sentinel-2 identifikovala fázi uprostřed vegetačního období.

Ve studii autorů Ghulam a kol. (2016) byl index ISI použit v návaznosti na statistickou analýzu nazývanou jednofaktorová analýza rozptylu (*one-way ANOVA*), kterou byly charakterizovány rozdíly ve spektrální odezvě vegetačních druhů na koncentraci ozonu v prostředí. Cílem studie bylo ukázat, zda je porovnáním spektrálních charakteristik druhů s různou úrovní citlivosti na ozon možné určit spektrální projev poškození rostliny ozonem. Výsledky ANOVA určily oblasti spektra, na kterých všech pět sledovaných druhů vegetace mělo specifickou, a tedy odlišitelnou, spektrální odezvu. Výpočtem indexu ISI byly stanoveny konkrétní vlnové délky s největší vahou pro rozlišení spektrálního projevu rostlin (hodnoty $ISI < 1$). Analýza ANOVA se efektivně používá také pro výběr spektrálních proměnných vhodných pro vstup do klasifikačních algoritmů na základě spektrální separability vegetace (Pu, 2008; Potůčková a kol. (2016).

2.4 Redukce dimenzionality dat

Vysoká dimenzionalita neboli rozměrnost dat vstupujících do statistické analýzy může mít na analýzu negativní vliv. Tento vliv bývá označován jako prokletí dimenzionality a má za následek zhoršení efektivity datových struktur, které v případě klasifikace může snižovat úspěšnost klasifikátoru. „*Redukce dimenze představuje důležitý krok statistické analýzy či extrakce informace z mnohorozměrných dat, který může být v případě vysoce dimenzionálních dat zcela nezbytný*“ (Kalina a Tebbens, 2013).

Platí, že se stoupajícím množstvím příznaků (prvků vstupujících do klasifikace) stoupá také úspěšnost klasifikátoru, ale pouze do chvíle, kdy je dosaženo optimálního počtu prvků. Další zvyšování dimenzionality dat bez zvýšení počtu prvků trénovací množiny úspěšnost klasifikátoru snižuje, jelikož trénovací množina se stává s rostoucí rozměrností příznakového prostoru stále řidčeji rozložená a její prvky jsou proto stále méně reprezentativní (obr. 2). Vysoká komplexita klasifikátoru a malý rozsah trénovací množiny vedou k tzv. přeučení (*overfitting*), což je stav, kdy klasifikátor je příliš přizpůsobený trénovacím datům a je tak velmi nepřesný pro nové prvky – selhává na validační množině dat. S dimenzionalitou dat navíc také roste výpočetní náročnost systému.



Obr. 2: Prokletí dimenzionality (zdroj: Nikhil Buduma, 2015)

Cílem redukce dimenzionality dat je transformovat vstupní data z prostoru vyšší dimenze do prostoru nižší dimenze s co nejmenší ztrátou informace. V případě klasifikačních algoritmů se jedná o proces výběru nejinformativnějších příznaků. Jak

uvádí Blažek a kol. (2008), „*metodika redukce dimenzionality problému klasifikačního typu je velmi rozsáhlá a matematicky náročná*“, proto nebude v této diplomové práci objasněna podrobně, s výjimkou vybraných metod PCA a DA.

2.4.1 Analýza hlavních komponent

Analýza hlavních komponent (PCA – *principal component analysis*) je metoda aplikovaná na data DPZ ve fázi předzpracování za účelem zvýraznění obrazu a snížení dimenzionality dat. Jedná se o statistický postup, založený na rotaci os vícerozměrného prostoru, ve kterém jsou data zobrazena, ve směru jejich maximálního rozptylu. Původní datový soubor pásem obrazu, která jsou korelovaná, je tak transformován do méně rozměrného souboru nekorelovaných proměnných, který ale zároveň nese téměř celou informaci obsaženou v původním souboru. Tyto proměnné se nazývají hlavní komponenty a jejich využití může přispět ke zlepšení výsledků klasifikace (Rencher, 2002).

První hlavní komponenta datového souboru je orientována ve směru maximálního rozptylu původních dat. Druhá komponenta je na první komponentu kolmá a protíná jí tak, aby byla orientována ve směru druhého největšího rozptylu dat. Takto jsou vytvořeny všechny další komponenty a platí, že každá komponenta nese méně informace než předcházející komponenta, zároveň každá komponenta nese část informace, která nebyla obsažena v předcházejících. Ke každé komponentě existuje její vlastní číslo (*eigenvalue*), které vyjadřuje, kolik procent celkového rozptylu dat je v ní zahrnuto. Se zvětšujícím se číslem komponenty klesá procento rozptylu a narůstá výskyt šumu, a tedy snižuje se hodnota obrazové informace. Většina informace datového souboru je obsažena v prvních několika komponentách, další komponenty lze vyloučit z dalšího zpracování. „*Metoda hlavních komponent umožňuje namísto vyšetřování velkého počtu původních znaků s komplexními vnitřními vazbami analyzovat pouze malý počet nekorelovaných hlavních komponent*“ (Tonhauserová, 2013).

Výsledkem analýzy hlavních komponent jsou proměnné, které popisují maximální rozptyl dat, nicméně, jedná se o neřízený algoritmus, který nezohledňuje zařazení prvků datového souboru do skupin. Provedením analýzy dojde k sumarizaci

informace obsažené v původním datovém souboru a snížení jeho rozměrnosti, výsledek analýzy ale neříká nic o separabilitě proměnných.

2.4.2 Diskriminační analýza

Metodou, ve které redukce dimenzionality dat úzce souvisí s nalezením proměnných, které nejlépe separují skupiny datového souboru, je diskriminační analýza (DA). Je to metoda vícerozměrné statistické analýzy, která se zabývá zkoumáním závislosti jedné kvalitativní závisle proměnné na skupině několika kvantitativních nezávisle proměnných (Rencher, 2002). Nezávisle proměnné se označují také jako diskriminátory a tvoří sloupce zdrojové matice. Na druhé straně zdrojové matice jsou skupiny kategoriální proměnné, do kterých jsou zařazené objekty vstupních dat. Pro všechny objekty jsou známy všechny hodnoty diskriminátorů. Cílem DA je odpovědět na otázku, nakolik je příslušnost objektu ke skupině ovlivněna hodnotami těchto proměnných.

Jednou ze základních funkcí DA je sestavení klasifikačního pravidla na základě vstupních dat pro nalezení skupinové příslušnosti nezařazených objektů (klasifikační neboli predikční diskriminační analýza). Pro účely této diplomové práce je důležitější druhá funkce, kterou je určení vlivu diskriminátorů na odlišení jednotlivých skupin a identifikace diskriminátorů, které k tomuto odlišení přispívají největší mírou (kanonická diskriminační analýza). Tato funkce je běžně využívána ve fázi předzpracování dat jako prostředek redukce jejich dimenzionality pro účely klasifikace anebo strojového učení (Huberty a Olejnik, 2006). Kanonická DA se také označuje jako Fisherova, podle jejího autora Ronalda A. Fishera, který jí popsal již v roce 1936 pod názvem *canonical variate analysis* (Fisher, 1936). Původně byla použita k řešení otázky zařazení objektu do jedné ze dvou tříd. Později, v roce 1948, jí C.R. Rao generalizoval pro více než dvě skupiny (Rao, 1948). Vychází z předem známých příslušností objektů do skupin, a souboru proměnných, kterými jsou objekty charakterizovány.

Analýza hledá takovou lineární kombinaci původních nezávisle proměnných, která by nejlépe ze všech separovala dané skupiny tak, aby jejich vnitroskupinová variabilita byla co nejmenší a meziskupinová variabilita co největší, tzn. aby bylo maximalizováno Fisherovo diskriminační kritérium. Zároveň platí, že nezávisle

proměnné ovlivňují závisle proměnnou různou mírou, tzn., přispívají různou mírou k diskriminaci skupin závisle proměnné.

Následující odvození bylo převzato z Rabasová (2012). Uvažujme soubor prvků s rozsahem n , který je členěn do H skupin s rozsahy n_h , $h = 1, \dots, H$. Mějme p různých proměnných a každý prvek ze souboru ať je charakterizován hodnotami těchto proměnných X_1, \dots, X_p , které jsou obsaženy ve vektoru pozorování $x = (x_1, \dots, x_p)^T$. Označme x_i vektor p hodnot proměnných X_1, \dots, X_p zjištěných u i -té jednotky souboru, a x_{ih} vektor p hodnot proměnných X_1, \dots, X_p zjištěných u i -té jednotky v h -té skupině. Dále označme:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{vektor průměrů hodnot proměnných}$$

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{ih} \quad \text{vektor průměrů hodnot proměnných v } h\text{-té skupině.}$$

Potom:

$$\mathbf{T} = \sum_{h=1}^H \sum_{i=1}^{n_h} (x_{ih} - \bar{x})(x_{ih} - \bar{x})^T \quad \text{je celková variabilita proměnných,}$$

$$\mathbf{E} = \sum_{h=1}^H \sum_{i=1}^{n_h} (x_{ih} - \bar{x}_h)(x_{ih} - \bar{x}_h)^T \quad \text{je vnitroskupinová variabilita proměnných,}$$

$$\mathbf{B} = \sum_{h=1}^H \sum_{i=1}^{n_h} (\bar{x}_h - \bar{x})(\bar{x}_h - \bar{x})^T \quad \text{je meziskupinová variabilita proměnných.}$$

Pro libovolnou lineární kombinaci $Y = b^T x$ proměnných X_1, \dots, X_p , kde $b^T = (b_1, \dots, b_p)$ je vektor parametrů, určíme míru její meziskupinové, resp. vnitroskupinové variability jako $Q_B(Y) = b^T B b$ resp. $Q_E(Y) = b^T E b$. Potom Fisherovo diskriminační kritérium je:

$$F = \frac{Q_B(Y)}{Q_E(Y)} = \frac{b^T B b}{b^T E b}$$

Výpočet DA směřuje k nalezení kanonických diskriminačních funkcí. Každá funkce představuje takovou projekci dat do vícerozměrného prostoru, ve které jsou prvky skupin závisle proměnné co nejlépe odděleny. Prvky jsou odděleny podél přímky, jedná se tedy o lineární diskriminační analýzu (LDA). Počet funkcí je buďto rovný počtu diskriminátorů (když počet skupin závisle proměnné je větší než počet diskriminátorů) anebo rovný počtu skupin mínus jedna. Kanonická diskriminační funkce má obecný tvar:

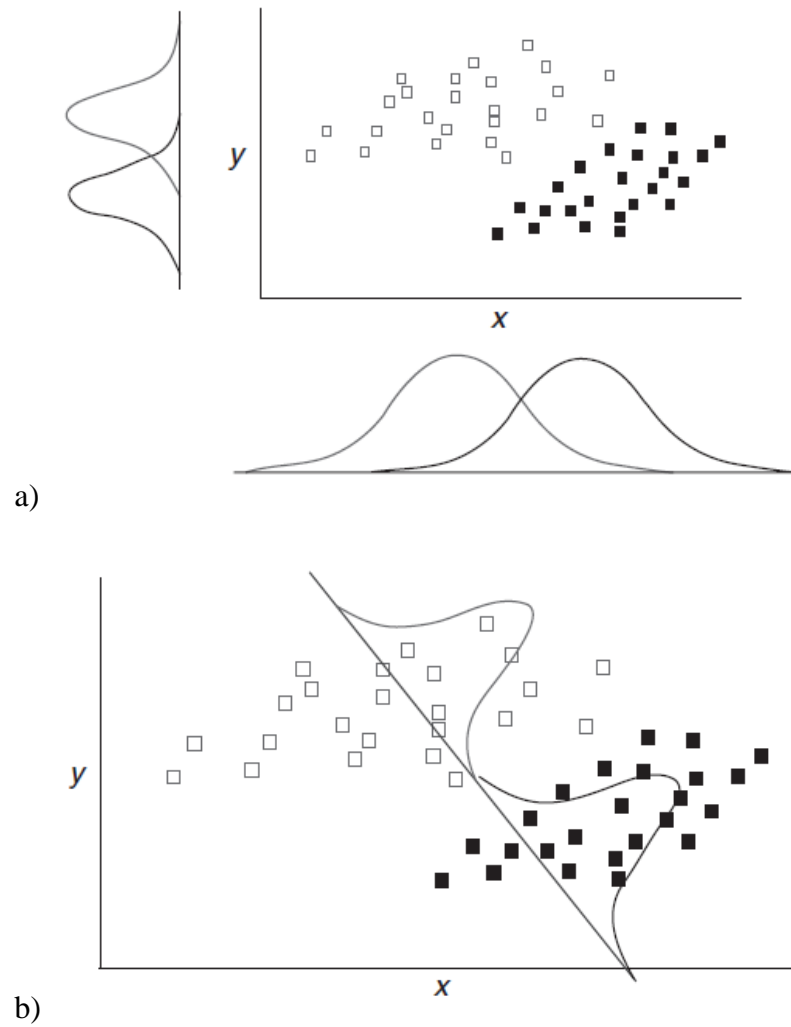
$$z = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_px_p$$

Standardizované koeficienty každé diskriminační funkce představují váhu jednotlivých proměnných v dané funkci a platí, že proměnné s největšími koeficienty přispívají nejvíce k diskriminaci skupin. Koeficienty pro první funkci se odvodí tak, aby skupinové těžiště neboli centroidy, byly maximálně vzdálené ve smyslu Mahalanobisovy vzdálenosti. Koeficienty každé další funkce musí dále maximalizovat rozdíly mezi skupinovými centroidy a současně hodnoty obou funkcí nesmí být korelovány. Centroidy skupin jsou tedy reprezentace každé skupiny ve formě bodu, ležícího v prostoru, který je vymezen diskriminačními funkcemi jako jeho osami. Každý prvek analýzy je také charakterizován jako bod v tomto prostoru a jeho zařazení do skupiny odpovídá jeho poloze vůči centroidům.

Výslednou hodnotou kanonické diskriminační funkce je kombinace původních proměnných, která se nazývá diskriminační skóre. Předpokládá se, že pro každou skupinu bude platit normální rozdělení diskriminačních skóre a míra překryvu jejich rozdělení pak určuje úspěšnost metody. V části a) obrázku 3 je zobrazený příklad dvou skupin popsaných dvěma proměnnými. Je vidět, že proměnná y je lepším diskriminátorem než proměnná x, jelikož překryv zařazení do skupiny je zde menší, nicméně žádná z proměnných x a y nedokáže skupiny odlišit úplně. Řešením této situace je nová kanonická osa popsaná diskriminační funkcí, která je kombinací původních proměnných (část b) obrázku 3). „*Princip diskriminační analýzy spočívá v nalezení kombinace proměnných odlišujících existující skupiny objektů, které nelze odlišit žádnou z proměnných samostatně*“ (Pomahačová, 2011).

S diskriminačními funkcemi jsou také spojena vlastní čísla (*eigenvalues*), která stejně jako u PCA popisují míru rozptylu zachycenou těmito funkcemi. V případě, že existuje víc než jedna diskriminační funkce, procentuální podíl vlastního čísla funkce vzhledem k součtu všech vlastních čísel všech funkcí určuje důležitost diskriminační

funkce – jedná se podíl celkové schopnosti diskriminátorů odlišit skupiny prvků, obsažený v dané funkci.



Obr. 3: Základní princip kanonické diskriminační analýzy (zdroj: Neil, 2002 - upraveno)

2.5 Spektrální rozlišení vybraných družic DPZ

Družicový dálkový průzkum podle Radoux a kol. (2016) vstoupil v posledních letech do nové éry senzorů pořizujících globální a volně dostupná data typická opakovaným snímáním v krátkém časovém horizontu. Rychlý technologický vývoj umožňuje neustálé zlepšování prostorového rozlišení dat a přidávání nových spektrálních pásem cílených na konkrétní výzkumné otázky. Dostupnost družicových snímků vede k jejich vzrůstajícímu využití nejen v oblastech výzkumné, ale i komerční sféry.

Tato práce se proto zaměřuje na tři současně využívané senzory: WorldView-2, Sentinel-2 a Landsat 8. Jejich základní charakteristiky jsou uvedeny v tabulkách 1–3.

2.5.1 WorldView-2

WorldView-2 je družicový senzor americké společnosti DigitalGlobe snímající v osmi spektrálních a jednom panchromatickém pásmu. Prostorové rozlišení spektrálních pásem je 1.84 m, v případě panchromatického pásma je to 0.46 m. Kromě čtyř tradičních pásem (2,3,4,7 v tab. 1) poskytuje také čtyři nová pásma, která mají zvyšovat informační obsah a přinést nové možnosti využití těchto dat v oblastech spektrální analýzy vegetace, monitoringu prostředí, plánování využití krajiny a objektově orientované klasifikace (LANDinfo, 2018).

Prostorové rozlišení (m)	Číslo pásma	Označení	Střední vlnová délka (nm)	Šířka pásma (nm)
~2	1	Coastal	429,3	47,3
	2	Blue	478,8	54,3
	3	Green	547,5	63,0
	4	Yellow	607,8	37,4
	5	Red	658,5	57,4
	6	Red Edge	723,5	39,3
	7	NIR 1	825,0	98,9
	8	NIR 2	919,4	99,6

Tab. 1: Spektrální pásma družice WorldView-2 (zdroj: LANDinfo, 2018 - upraveno)

2.5.2 Sentinel-2

Mise Sentinel-2 je tvořena dvěma identickými družicemi Sentinel-2A a Sentinel-2B, které se ve své oběžné dráze následují. Perioda snímkování jednoho místa na Zemi je tak přibližně 5 dní. Každá družice nese multispektrální senzor snímkující ve 13 pásmech. Zcela specifická jsou tři pásma v oblasti *red-edge* na červeném okraji viditelné části spektra, která jsou svými spektrálními parametry ideální pro monitorování vegetace se zaměřením na klasifikaci a sledování zdravotního stavu. Mise Sentinel-2 je zaměřena především na monitoring krajinného pokryvu a mapování změn v území a využití ploch (Copernicus, 2016).

Prostorové rozlišení (m)	Číslo pásma	Označení	Sentinel 2-A		Sentinel 2-B	
			Střední vlnová délka (nm)	Šířka pásma (nm)	Střední vlnová délka (nm)	Šířka pásma (nm)
10	2	Blue	496,6	98	492,1	98
	3	Green	560	45	559	46
	4	Red	664,5	38	665	39
	8	NIR	835,1	145	833	133
20	5	Red Edge 1	703,9	19	703,8	20
	6	Red Edge 2	740,2	18	739,1	18
	7	Red Edge 3	782,5	28	779,7	28
	8a	Red Edge 4	864,8	33	864	32
	11	SWIR 1	1613,7	143	1610,4	141
	12	SWIR 2	2202,4	242	2185,7	238
60	1	Coastal	443,9	27	442,3	45
	9	Water Vapor	945	26	943,2	27
	10	Cirrus	1373,5	75	1376,9	76

Tab. 2: Spektrální pásma družic Sentinel-2 (zdroj: ESA - 2018, upraveno)

2.5.3 Landsat 8

Historie družic Landsat sahá až do roku 1972 a Landsat 8 je nejnovější z osmidílné sady družic. Na palubě nese dva senzory: OLI (*Operational Land Imager*) a TIRS (*Thermal Infrared Sensor*). Oproti předchozím multispektrálním senzorům řady Landsat senzor OLI disponuje dvěma novými pásmy (1 a 9 v tab.3), která jsou zaměřena na výzkum vodních zdrojů, pobřežních oblastí a detekci oblaků typu cirrus. Nejlepší prostorové

rozlišení je v případě Landsat-8 15 m pro panchromatické pásmo, spektrální pásma jsou dostupná v prostorovém rozlišení 30 m. Perioda snímkování Landsat 8 je 16 dní, resp. 8 dní v konstelaci s Landsat 7. Díky návaznosti používaných senzorů a obsáhlým archivům se data družic Landsat hodí pro detekce změn v čase, a to zejména v souvislosti s vegetačním pokryvem a využitím krajiny (ARCDATA Praha, 2018).

Prostorové rozlišení (m)	Číslo pásma	Označení	Střední vlnová délka (nm)	Šířka pásma (nm)
30	1	Coastal	443,0	16,0
	2	Blue	482,0	60,0
	3	Green	561,4	57,3
	4	Red	654,6	37,5
	5	NIR	864,7	28,3
	6	SWIR 1	1608,9	84,7
	7	SWIR 2	2200,7	186,7
	9	Cirrus	1373,4	172,4

Tab. 3: Spektrální pásma senzoru OLI družice Landsat 8 (zdroj: Landsat Science, 2018 - upraveno)

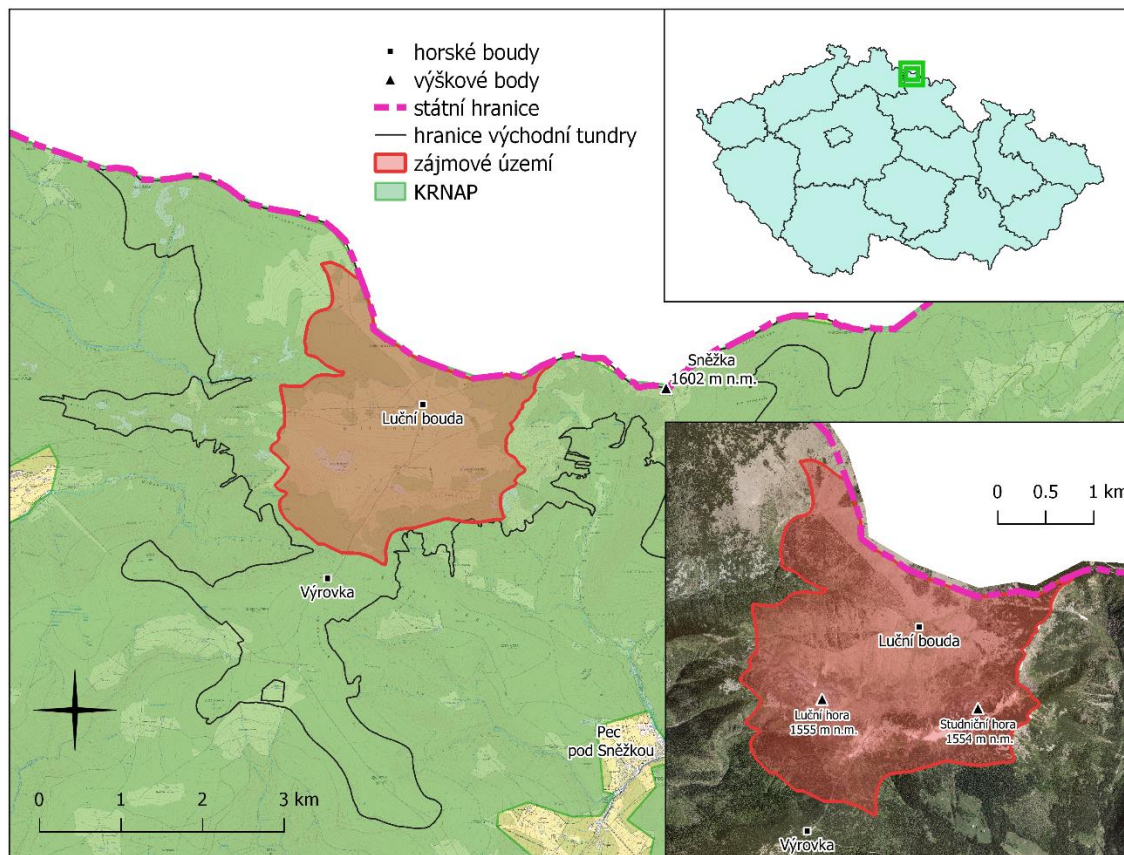
3 Zájmové území

3.1 Charakteristika a vymezení

Oblast arкто-alpínské tundry v Krkonoších tvoří nejvyšší polohy pohoří nad horní hranicí lesa. Přírodní charakteristika této oblasti je podobná jak subarktickým, tak vysokohorským regionům a je výrazně ovlivněná extrémními klimatologickými podmínkami, jako je například vystavení silným severozápadním větrům, dlouhotrvající sněhová pokrývka a nízká průměrná roční teplota vzduchu (Müllerová, 2005). Kromě mechů, lišejníků a vřesu zde převládají alpínské travní porosty s převažujícím zastoupením smilky tuhé a kosodřevina. Jedná se o unikátní reliktní výskyt tohoto typu prostředí zahrnutý v mezinárodním programu pro monitoring tundry INTERACT. Plocha území tundry činí celkem 47 km², 32 km² na území Česka a 15 km² na území Polska.

Území arкто-alpínské tundry v Krkonoších se dělí na dvě prostorově oddělené části, západní a východní. Tato práce se zaměřuje na východní část. Zájmové území, pro které jsou dostupná hyperspektrální data, představuje okolí Luční boudy ohraničené vrstevnicemi 1300 až 1475 m n. m. a terénními zlomy (obr. 4). Tvoří plochu přibližně 6,6

km², což je 28,9 % plochy východní části a 14 % celkové plochy krkonošské tundry. Oblast se nachází na území I. zóny KRNAP.



Obr. 4: Zájmové území (zdroj: vlastní zpracování, použitá data: ARCDATA PRAHA, ArcČR 500 verze 3.2; Geoportál ČÚZK, Prohlížečí služba WMS – Ortofoto).

3.2 Legenda pro klasifikaci

Klasifikace vegetace tundry, která je cílem této práce, bude probíhat do již stanovené legendy vytvořené pro účely předchozích publikací ve spolupráci s botaniky KRNAP. Počet kategorií této legendy se pohybuje v rozmezí 9-12 tříd, v závislosti na specifických požadavcích a územním rozsahu analýz, které byly předmětem předešlých publikací a prací zaměřených na klasifikaci krkonošské tundry (viz část 2.2 práce).

Pro klasifikace v této práci je použita legenda o 12 třídách obsahující všechny základní vegetační kategorie tundry. Přehled všech tříd s číselným a zjednodušeným

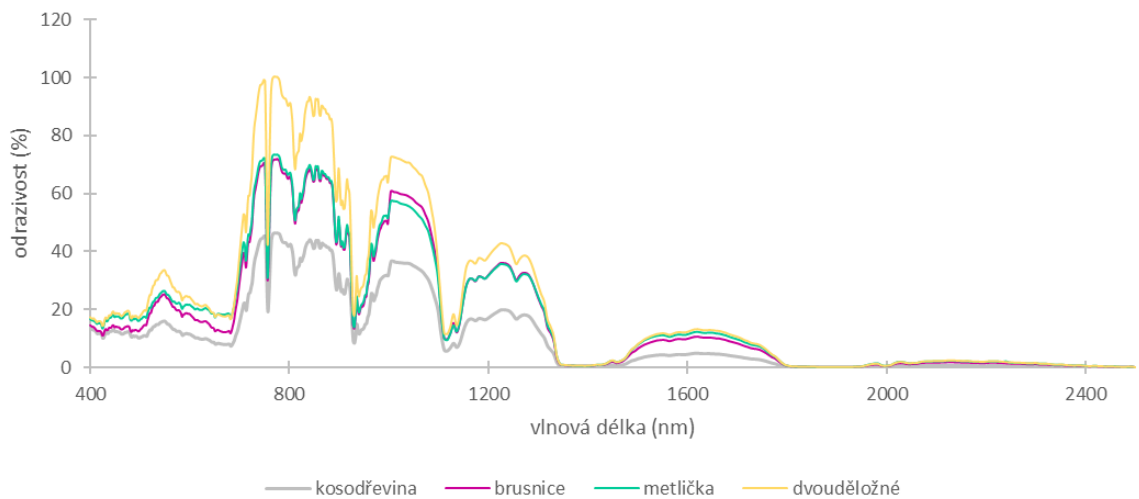
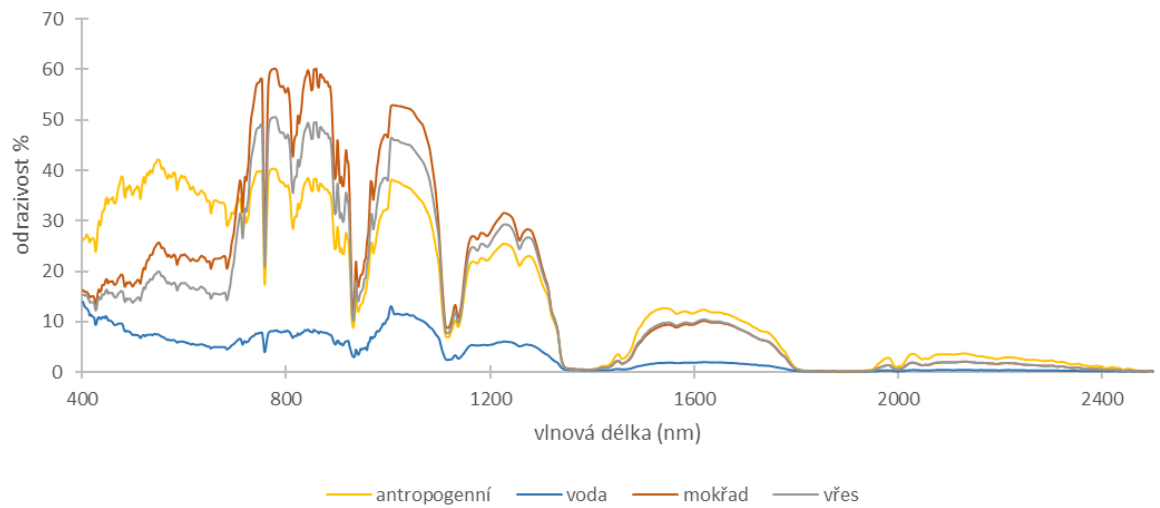
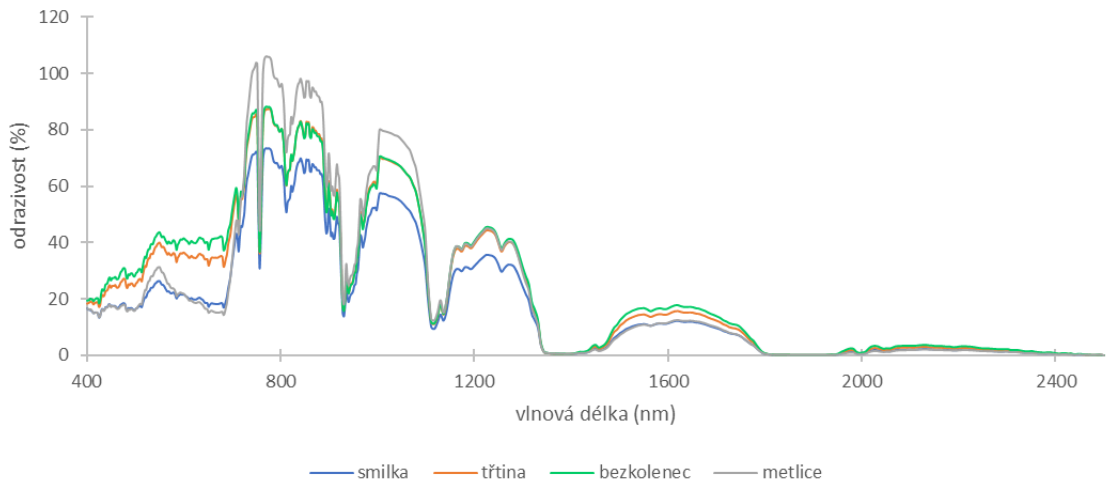
slovným označením uvádí tabulka 4. Uveden je také latinský název, v případě tříd zahrnujících několik různých rostlinných druhů je uveden latinský název převažujícího druhu (označeny *). Číselné označení tříd v této práci je ponecháno v souladu se zažitým označením. Označením 6a-6c je naznačena příslušnost těchto tříd do společné skupiny druhů – alpínské trávníky zapojené. Stejně tak třídy 7a-7c jsou součástí skupiny subalpínské vysokostébelné trávníky. Práce se pak více zaměřuje na vybrané převažující a pro botaniky KRNAP zásadní druhy jednoděložné vegetace – smilku jako pro tundru významnou kategorii (6a) a konkurenčně silných subalpínských vysokostébelných trávníků (7a-7c). Obrázky 6 a 7 zobrazují fotografie tříd ve fenologické fázi shodné s termínem pořízení dat (červen) a příklady trénovacích ploch v barevné syntéze hyperspektrálního snímku (kombinace vybraných pásem SWIR-NIR-červená). Širší charakteristika jednotlivých tříd zde není uvedena. Informace o výskytu, druhovém složení v rámci třídy a dalších specifikách lze vyhledat např. v práci Jakešová (2014).

Třída	Název	Latinský název	Označení
1	Kosodřevina	<i>Pinus mugo</i> *	Kosodřevina
2	Mokřady a rašeliniště	<i>Sphagnum magellanicum</i> *	Mokřad
3	Vodní plochy	-	Voda
4	Alpínská vřesoviště	<i>Calluna vulgaris</i> *	Vřes
5	Subalpínská brusnicová vegetace	<i>Vaccinium myrtillus</i> *	Brusnice
6a	Smilka tuhá	<i>Nardus stricta</i>	Smilka
6b	Metlička křivolaká	<i>Deschampsia flexuosa</i>	Metlička
6c	Druhově bohaté porosty s vysokým zastoupením dvouděložných	<i>Festuca supina</i> *	Dvouděložné
7a	Třtina chloupkatá	<i>Calamagrostis villosa</i>	Třtina
7b	Bezkolenec modrý	<i>Molinia caerulea</i>	Bezkolenec
7c	Metlice trsnatá	<i>Deschampsia cespitosa</i>	Metlice
8	Kamenná moře a antropogenní plochy	-	Antropogenní

* převažující druh

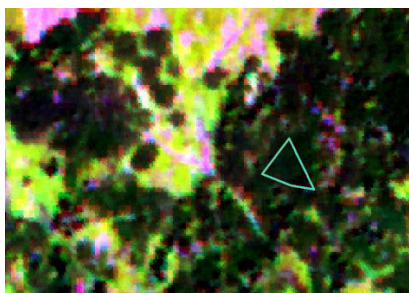
Tab. 4: Přehled tříd legendy pro klasifikaci krkonošské tundry.

Zprůměrováním hodnot odrazivosti trénovacích polygonů tříd ze základního referenčního souboru (definován v části 4.1) ve všech spektrálních pásmech dat AisaDUAL byly vytvořeny grafy spektrálních křivek tříd (obr. 5).

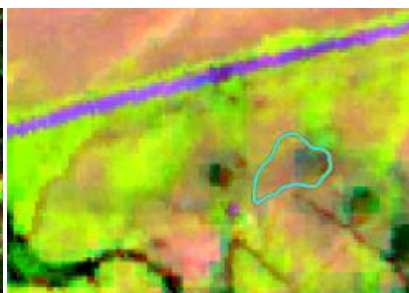


Obr. 5: Křivky odrazivosti tříd.

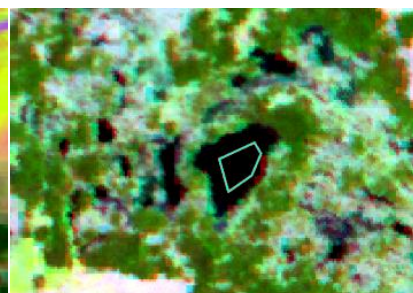
1 – Kosodřevina



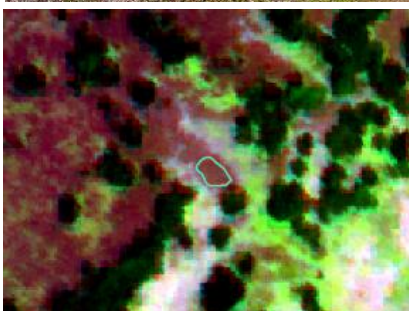
2 – Mokřady a rašeliniště



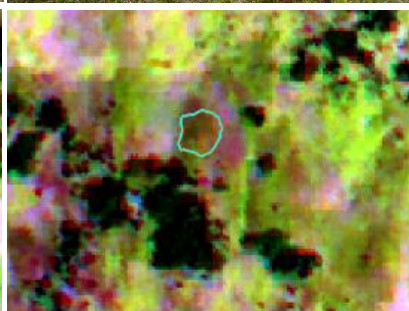
3 – Vodní plochy



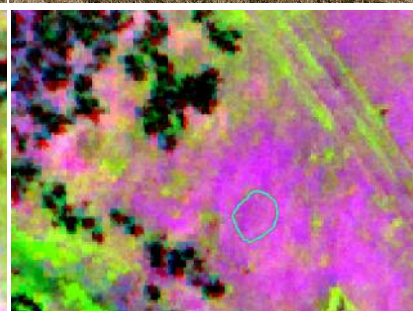
4 – Alpínská vřesoviště



5 – Subalpínská brusnicová
vegetace

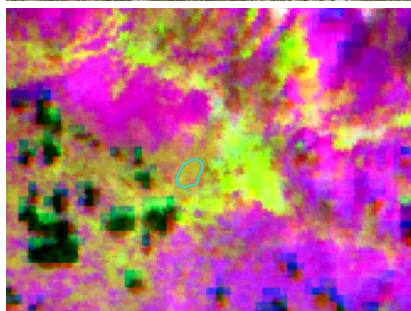


6a – Smilka tuhá

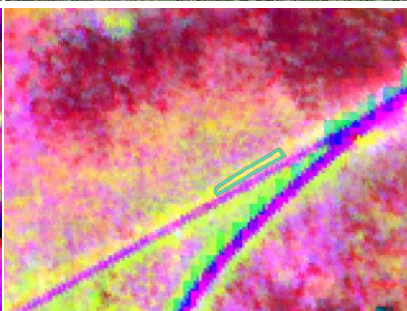


Obr. 6: Fotografie tříd legendy a ukázky trénovacích ploch (autor fotografií: Lucie Červená)

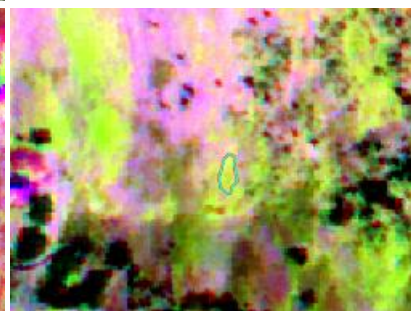
6b – Metlička křivolaká



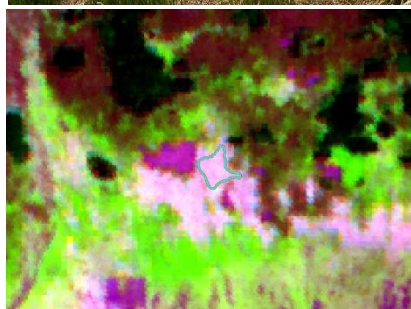
6c – Druhově bohaté porosty,
dvouděložné



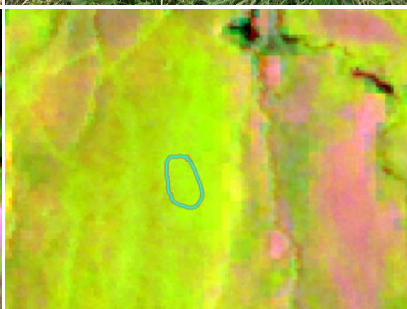
7a – Třtina chloupkatá



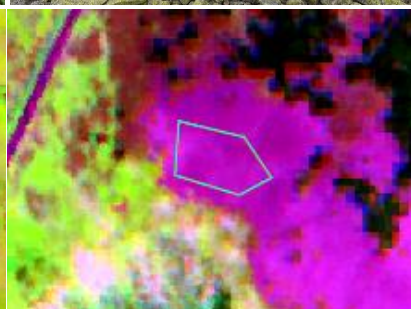
7b – Bezkolenec modrý



7c – Metlice trsnatá



8 – Kamenná moře a
antropogenní plochy



Obr. 7: Fotografie tříd legendy a ukázky trénovacích ploch (autor fotografií: Lucie Červená)

4 Data a metodika

4.1 Použitá data a předzpracování

Tato práce je založena na použití hyperspektrálních leteckých dat s velmi vysokým prostorovým rozlišením. Jedná se o data systému AisaDUAL pořízená 19. června 2013. Jejich geometrické, radiometrické a atmosférické korekce byly provedeny společností Geodis Brno, která data pořizovala. Kompletní dokumentace k pořízení a zpracování dat je uvedena v příslušné technické zprávě (Geodis Brno, 2013). Systém AisaDUAL kombinuje senzory AisaEAGLE, pořizující data ve viditelné a blízké infračervené části spektra (VNIR), a AisaHAWK snímající v pásmu krátkých vlnových délek infračerveného záření (SWIR). Data pořízená senzory byla sloučena do jednoho datového souboru (mozaiky) s výsledným spektrálním rozsahem 400–2500 nm, který obsahuje celkem 494 pásem. Datový soubor byl dále převzorkován na velikost pixelu 1 m, což odpovídá prostorovému rozlišení senzoru AisaEAGLE. Původní prostorové rozlišení u senzoru AisaHAWK bylo 3 m.

Ve spektrech odrazivosti tříd uvedených v části 3.2 lze pozorovat chybu dat projevující se hodnotami odrazivosti vyššími než 100 %. Tato chyba je s největší

pravděpodobností důsledkem nepřesné atmosférické a radiometrické korekce provedené společností Geodis Brno na základě referenčních spektroradiometricky zaměřených ploch změřených v terénu (Geodis Brno, 2013). Nepřesné korekce se projevují také odlišností tvaru spektrálních křivek tříd od konvenčního tvaru křivek vegetace. Jelikož pro celé zájmové území byla použita stejná korekce, lze předpokládat, že datový soubor je zatížen stejnou chybou a tím pádem je použitelný pro analýzu z hlediska hodnocení odlišitelnosti tříd. Negativní vliv, související s nepřesností atmosférické korekce, může mít pouze průběh spektrální odrazivosti tříd v oblasti NIR, kde se nachází několik výrazných propadů hodnot způsobených atmosférickou absorpcí. Intervaly vlnových délek propadů jsou nicméně pouze úzké.

Referenční data byla naměřena v terénu v několika fázích v průběhu let 2014 až 2017. Plochy odpovídající vegetačním kategoriím legendy byly identifikované botanikem a geolokalizované pomocí GPS přístroje Trimble Geoexplorer 3000 Geo XT s přesností 10 cm. Plochy nevegetačních kategorií byly získány interpretací a vektorizací ortofoto snímků. Základní referenční soubor, použitý v této práci, je za účelem porovnatelnosti výstupů totožný s referenčním souborem již použitým v Kupková a kol. (2017). Tento soubor obsahuje celkem 123 polygonů zařazených do jedné z dvanácti skupin legendy. Je rozdělen na trénovací a validační polygony tak, že pro každou kategorii bylo 40 % dat použitých jako trénovací a 60 % jako validační. Trénovací datový soubor obsahuje 51 polygonů a validační datový soubor, míněný pro hodnocení přesnosti klasifikace, obsahuje 72 polygonů (tab. 5).

Kompletní referenční soubor výzkumu krkonošské tundry obsahuje aktuálně 258 polygonů sesbíraných v terénu, představujících jak homogenní plochy, tak plochy smíšených rostlinných společenstev. Pro účely diplomové práce byly z tohoto souboru použity pouze vybrané plochy samostatných kategorií. Na základě zaměření diplomové práce na odlišení druhů jednoděložných rostlin byly vytvořeny dva referenční soubory obsahující pouze kategorie smilka, třtina, bezkolenek a metlice. První ze souborů byl pouze výběrem polygonů čtyř daných kategorií z trénovacích polygonů základního referenčního souboru. Vybráno bylo 21 trénovacích příslušných polygonů. V druhém souboru byly tyto polygony doplněny 40 vybranými polygony daných kategorií z kompletního referenčního souboru. Doplněné polygony byly vybrány tak, aby se nacházeli ve výřezu dostupných hyperspektrálních dat a aby daná vegetační kategorie

byla v datech dobře identifikovatelná. Vznikl rozšířený referenční soubor obsahující 58 trénovacích polygonů rozdělených do čtyř kategorií travin (tab. 6).

Kategorie		Celý referenční soubor			Trénovací plochy		Validační plochy	
		Počet polygonů	Počet pixelů	Pixelů v %	Počet polygonů	Počet pixelů	Počet polygonů	Počet pixelů
1	Kosodřevina	7	8559	29,6	3	3620	4	4939
2	Mokřad	9	2156	7,5	4	943	5	1213
3	Voda	4	361	1,2	2	74	2	287
4	Vřes	16	2979	10,3	7	1138	9	1841
5	Brusnice	10	893	3,1	3	410	7	483
6a	Smilka	14	4338	15,0	6	1694	8	2644
6b	Metlička	13	1561	5,4	5	371	8	1190
6c	Dvouděložné	6	396	1,4	3	180	3	216
7a	Třtina	10	1411	4,9	4	438	6	973
7b	Bezkoleneč	16	1736	6,0	7	803	9	933
7c	Metlice	11	2131	7,4	4	776	7	1355
8	Antropogenní	7	2380	8,2	3	936	4	1444
Celkem		123	28901	100,0	51	11383	72	17518

Tab. 5: Základní referenční soubor – trénovací a validační plochy.

Kategorie		Rozšířený referenční soubor		
		Počet polygonů	Počet pixelů	Pixelů v %
6a	Smilka	18	5642	47,2
7a	Třtina	12	1772	14,8
7b	Bezkoleneč	15	2021	16,9
7c	Metlice	13	2511	21,0
Celkem		58	11946	100,0

Tab. 6: Rozšířený referenční soubor kategorií smilka, třtina, bezkoleneč a metlice.

Dalším krokem po stanovení referenčních souborů bylo sestavení vstupních souborů pro diskriminační analýzu. Soubory byly vytvořeny pomocí nástrojů programu ArcMap 10.6. Diskriminační analýza vyhodnocuje hodnoty odrazivosti v jednotlivých pásmech obrazu pro každý z pixelů, vstupem je tedy tabulka hodnot pixelů ve všech pásmech a zařazení pixelu do kategorie legendy. Pro vytvoření takovéto tabulky byla

nejprve z datového souboru AisaDUAL pomocí funkce *Raster to Point* vytvořena vrstva bodů, ve které každý bod reprezentuje centroid jednoho pixelu. Z těchto bodů pak byly pomocí *Select by Location* vybrány body, které se překrývaly s polygonovou vrstvou trénovacích ploch základního referenčního souboru. Propojením bodové vrstvy s polygonovou vrstvou pomocí funkce *Spatial Join* byl těmto bodům přiřazen atribut popisující jejich příslušnost do jedné z kategorií. Posledním krokem bylo použití funkce *Extract Multi Values to Point* na výslednou bodovou vrstvu. Vstupem do této funkce byl datový soubor AisaDUAL, složený z celkem 494 pásem. Bodům bylo přiřazeno 494 atributů, ze kterých každý obsahoval hodnoty odrazivosti jednotlivých pásem v daném bodu (pixelu). Z atributové tabulky byl následně odstraněn poslední sloupec, jelikož v pásmu 494 byla hodnota všech pixelů 0. Stejný postup byl aplikován na dva referenční soubory obsahující pouze travní kategorie. Atributové tabulky tří bodových souborů byly vyexportovány ve formátu *.txt.

4.2 Simulace družicových dat

Data optických družicových senzorů byla v této práci simulována jednotně z hyperspektrálních dat senzoru AisaDUAL. Tímto způsobem se zamezilo rozdílům v spektrálním chování vegetace způsobeným posunem fenologické fáze v případě rozdílných termínů pořízení dat a také rozdílům způsobeným atmosférickými vlivy a atmosférickými korekcemi. Porovnáním výsledků klasifikace se tak přímo hodnotí použité klasifikační algoritmy a vhodnost vstupních dat.

Simulace spektrálních charakteristik družicových senzorů byla provedena v programu ENVI 5.4 pomocí nástroje *Spectral Resampling*. Tento nástroj převede vstupní data na data se spektrálním rozlišením odpovídajícím vstupnímu nastavení. Jako vzorové spektrální rozlišení lze nastavit 1) předdefinovaná spektrální funkce známého senzoru (výběr dostupný přímo v nástroji), 2) uživatelem definovaná spektrální funkce ve formátu ENVI spektrální knihovny 3) soubor vlnových délek ve formátu ASCII, s možností specifikovat také rozsah pásma pomocí parametru FWHM (*full-width-half-maximum*), 4) obrazový soubor se spektrálním rozlišením, které požadujeme. Charakter vzorových dat ovlivňuje výběr metody převzorkování, tzn. například při zadání pouze vlnových délek jako hodnot středů požadovaných pásem nástroj přepočte vlnové délky na základě Gaussovy křivky a automaticky nastaví FWHM jako polovinu rozdílu mezi středy pásem (Harris Geospatial Solutions, 2018a).

Všechny senzory, jejichž spektrální rozlišení bylo potřeba pro účely této diplomové práce simulovat, se nacházely v nabídce předdefinovaných funkcí programu ENVI 5.4. Tímto způsobem byly vytvořeny datové soubory odpovídající senzorům WorldView-2, Sentinel-2 a Landsat 8 OLI. Počty pásem a jejich spektrální rozsahy jsou pro všechny senzory uvedeny v části 2.5. V případě senzoru Landsat 8 OLI bylo v simulaci programu ENVI vynecháno pásmo 9 - Cirrus zaměřené na monitorování vysoké oblačnosti. Vizuálním zhodnocením nasimulovaných pásem senzoru Sentinel-2 bylo zjištěno, že pásmo 10 – Cirrus tohoto senzoru rovněž nevyhovuje (též je zaměřené na monitorování vysoké oblačnosti), takže bylo před dalším zpracováním odstraněno. Předdefinované funkce rovněž nezahrnovaly panchromatická pásma senzorů WorldView-2 a Landsat 8 a termální pásma senzoru Landsat 8 TIRS.

Prostorové rozlišení výstupních souborů zůstalo shodné s daty senzoru AisaDUAL tzn. simulované soubory nereflktují prostorové rozlišení družicových senzorů. To souvisí s cílem práce, kterým je stanovení vlnových délek stěžejních pro odlišení druhů a hodnocení vhodnosti družicových dat DPZ z hlediska spektrálního rozlišení. Důležitost prostorového rozlišení pro klasifikaci je nesporná a byla již dokázána v mnoha publikacích (Virtanen a Ek, 2014; Suchá a kol., 2016; Kupková a kol., 2017), pro účely této práce bude ale v dalších částech prostorové rozlišení zanedbáno.

Převzorkované datové soubory byly načteny do programu ArcMap 10.6 a postupem popsaným v části 4.1 byly vytvořeno dalších devět vstupních souborů pro diskriminační analýzu. Jednalo se o tabulky hodnot pásem každého družicového senzoru pro pixely odpovídající všem trénovacím polygonům referenčního souboru, trénovacím polygonům referenčního souboru pouze čtyř vybraných tříd, a polygonům rozšířeného referenčního souboru čtyř vybraných tříd.

4.3 Diskriminační analýza v SPSS Statistics

Existuje několik způsobů, jak diskriminační analýzu provést. Jedním z běžných postupů je výpočet v programovacím jazyku a prostředí R, které jsou určeny pro statistickou analýzu dat a jejich grafické zobrazení. Lze nalézt také případy výpočtu v jazyku Python. Diskriminační analýza bývá také součástí statistických programů, jakými jsou například SAS System anebo SPSS Statistics. Manly (2005, citováno Savić a kol., 2008) zdůrazňuje potřebu důkladného nastudování analytických postupů, terminologie a výstupů diskriminační analýzy, která probíhá přes uživatelské rozhraní, jelikož tyto se mohou pro různé programy lišit.

Pro diskriminační analýzu nad hyperspektrálními daty řešenou v této diplomové práci byl vybrán statistický program SPSS Statistics od společnosti IBM, který disponuje řadou statistických analýz a příklady jeho využití ve vědeckých pracích lze najít například v Kalaczka a Sanchez (2008), Weisa a Gerhards (2007) a Pomahačová (2012). Program nabízí dva přístupy k provedení diskriminační analýzy. V prvním přístupu vstupují do analýzy všechny diskriminátory najednou a jsou vyhodnocovány současně, druhou možností je kroková metoda, kdy se podle zvolených kritérií vybírají nejvhodnější diskriminátory postupně až do chvíle, kdy je dosaženo optimálního stavu. Vstupem do programu je v obou případech připravený datový soubor ve formě tabulky, obsahující prvky s přiřazenými hodnotami diskriminátorů a zařazením do skupin. Cesta pro spuštění nástroje DA v programu SPSS Statistics je *Analyze – Classify – Discriminant Analysis*.

Dokumentace StatSoft (2013) popisuje krokovou (*stepwise*) diskriminační analýzu jako nejběžnější postup pro určení proměnných stěžejních pro oddělení tříd v případě vysokého počtu proměnných vstupujících do analýzy. Oproti základní diskriminační analýze, která vyhodnocuje všechny proměnné a jejich vliv na separaci skupin najednou (rovnou pro vstupní soubor počítá diskriminační funkce), v případě krokové analýzy výpočtu diskriminačních funkcí předchází hodnocení významnosti pro každou proměnnou jednotlivě a výběr, resp. odstranění proměnných, které byly vyhodnoceny jako redundantní. K tomu jsou použity hodnoty F jednotlivých proměnných založené na výpočtu *Wilks' lambda*. F hodnota proměnné vyjadřuje její statistickou významnost pro diskriminaci mezi skupinami, tedy jakou mírou proměnná přispívá

k zařazení prvků souboru do skupiny. Platí, že vysoká hodnota F minimalizuje *Wilks' lambda* pro danou proměnnou.

Kroková diskriminační analýza může být buď dopředná (*forward*) nebo zpětná (*backward*). *Forward* analýza začíná výběrem pouze jedné proměnné, která samotná nejlépe separuje skupiny vstupního souboru. V každém dalším kroku analýzy je přidána další proměnná tak, aby její přidání co nejvíce zlepšilo separaci skupin (maximalizuje hodnotu F). Opakem tohoto přístupu je *backward* analýza, tzv. zpětná eliminace, kdy na začátku výpočtu se uvažují všechny dostupné proměnné a v každém kroku se odstraní ta, která přispívá k separaci skupin nejméně (nejnižší hodnota F). Kroková diskriminační analýza v programu SPSS Statistics tyto dva přístupy kombinuje. Výběr proměnných probíhá postupně a v každém kroku analýzy se kontroluje, zda se některá z dříve vybraných proměnných nestala v přítomnosti nově přidaných proměnných redundantní. Uživatel vstupuje do algoritmu zadáním hraničních hodnot F pro přidání, resp. odebrání proměnných (*F to enter*, *F to remove*). Výběr proměnných se zastaví v okamžiku, kdy nejvyšší hodnota F dostupných proměnných již nepřesahuje hraniční hodnotu F pro přidání proměnné do výběru. Kromě metody založené na testu *Wilks' lambda*, která je defaultním nastavením, lze nastavit také výběr na základě míry nevysvětleného rozptylu, Mahalanobisovy vzdálenosti a dalších.

V programu SPSS byly postupně načtené tabulky hodnot pro všechny referenční soubory, kterých bylo celkem 12 (3 na základě dat AisaDUAL a 3 pro každý družicový senzor), a pro každou byla provedena diskriminační analýza. V nastaveních nástroje diskriminační analýzy byla definována závisle proměnná, tedy atribut tabulky, který nesl informaci o zařazení prvků do kategorií legendy, a její rozsah, tedy v našem případě 1-13 pro referenční soubor se všemi třídami, a 1-4 pro referenční soubory obsahující pouze třídy travin. Jako nezávisle proměnné byla zvolena v každém souboru všechna pásma. Pro tabulky dat AisaDUAL byl tedy počet nezávisle proměnných 493, v případě družicových senzorů počet nezávisle proměnných odpovídal počtu použitých pásem senzoru. Byla vybrána metoda *stepwise* na základě testu *Wilks' lambda* a všechny parametry byly ponechány v původním nastavení.

Výsledné výběry pásem stěžejních pro oddělení definovaných skupin byly dále hodnoceny podle výše standardizovaných koeficientů ve vytvořených diskriminačních funkcích. V případě referenčního souboru všech tříd bylo hodnoceno pouze prvních

7 diskriminačních funkcí z celkem 11, jelikož na základě vlastních čísel tyto vysvětlovaly celkem 96.6 % variance vstupních dat (obr. 10). V případě referenčních souborů obsahujících pouze třídy travin byly hodnoceny všechny tři diskriminační funkce. Z výběrů pásem, které byly výsledkem diskriminační analýzy, byla odstraněna ta pásma, kterým byla ve všech diskriminačních funkcích přiřazena nízká absolutní hodnota koeficientu. Za hraniční hodnotu byla zvolena hodnota 1. Výběr a odstranění pásem s nízkou důležitostí z výběru bylo provedeno v programu Microsoft Excel 2016.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	11.496 ^a	44.1	44.1	.959
2	6.525 ^a	25.0	69.1	.931
3	2.682 ^a	10.3	79.4	.853
4	1.851 ^a	7.1	86.5	.806
5	1.175 ^a	4.5	91.0	.735
6	.851 ^a	3.3	94.3	.678
7	.595 ^a	2.3	96.6	.611
8	.455 ^a	1.7	98.3	.559
9	.212 ^a	.8	99.1	.418
10	.133 ^a	.5	99.6	.342
11	.097 ^a	.4	100.0	.297

a. First 11 canonical discriminant functions were used in the analysis.

Obr. 8.: Vlastní čísla diskriminačních funkcí referenčního souboru se všemi kategoriemi legendy (zdroj: SPSS Statistics).

4.4 Klasifikace

Klasifikace do zvolené legendy probíhaly v programu ENVI 5.4. Jako klasifikační algoritmus byl zvolen *Support Vector Machine* (SVM), statistická učící metoda založená na principu tzv. strukturální minimalizace rizika (Vapnik, 1998). Jedná se o řízenou pixelovou klasifikaci, velmi častou využívanou pro potřeby DPZ. Tato klasifikace patří do skupiny neparаметrických klasifikačních algoritmů, které jsou typické tím, že u nich nepředpokládáme znalost tvaru pravděpodobnostních charakteristik tříd. Princip SVM popisuje Pejčoch, 2011: „*Algoritmus SVM je založen na převedení úlohy klasifikace do nelineárně separovatelných tříd na úlohu klasifikace do tříd lineárně separovatelných. V prostoru transformovaných atributů se poté snažíme nalézt dělicí nadrovinu, která má co největší vzdálenost od příkladů různých tříd.*“ Dělicí nadrovina je zvolena ve fázi učení na základě tzv. pomocných vektorů, které se nacházejí na okrajích klasifikovaných skupin a algoritmus je vybírá automaticky.

Funkce SVM v programu ENVI 5.4 nabízí 3 typy kernelové funkce jakožto vstupního parametru pro klasifikaci – lineární, polynomicou a radiální bázovou (RBF). Podle vzoru článku Kupková a kol. (2017) byla pro tuto práci zvolena funkce RBF, která dosahovala pro klasifikaci krkonošské tundry nejlepších výsledků. U této funkce je dále potřebné nastavení parametrů g – šířka kernelové funkce a C – regularizační parametr. Rovněž podle vzoru předchozí publikace byly tyto parametry nastaveny následovně: g jako převrácená hodnota počtu pásem vstupujících do klasifikace a C na hodnotu 100, což jsou i defaultní nastavení této funkce (Petropoulos, 2012). Výsledky klasifikace byly upraveny pomocí mediánového filtru o velikosti 3x3 pixely, čímž bylo docíleno odstranění izolovaných pixelů.

Přesnost klasifikací byla hodnocena na základě referenčního souboru validačních polygonů (viz část 4.1) pomocí chybové matice. Opět bylo využito programu ENVI 5.4, konkrétně funkce *Confusion Matrix Using Ground Truth ROIs*. Počet řádků a sloupců chybové matice odpovídá počtu klasifikovaných tříd. Sloupce matice představují třídy klasifikace a všechny pixely ve sloupci dané třídy do této třídy skutečně patří. Řádky matice pak nesou informaci o tom, do které třídy byly pixely zařazeny klasifikační metodou. Správně klasifikované pixely se nachází na diagonále matice, pixely mimo diagonály jsou zařazeny chybně. Výstupem funkce je matice s absolutními hodnotami

a také matice s procentuálním vyjádřením. Kromě samotné matice tento výstup udává také celkovou přesnost klasifikace, kappa koeficient, poměr chyb z opomenutí, chyb z nesprávného zařazení a uživatelskou a zpracovatelskou přesnost pro každou třídu. Uživatelská přesnost je definována jako poměr počtu správně klasifikovaných pixelů dané třídy a celkového počtu pixelů, které do dané třídy byly zařazeny. Vyjadřuje pravděpodobnost, s jakou pixel zařazený do dané třídy tuto skutečně reprezentuje. Zpracovatelská přesnost je poměr počtu správně klasifikovaných pixelů a počtu pixelů použitých pro testování dané třídy (Harris Geospatial Solutions, 2018b).

5 Výsledky

5.1 Pásma senzoru AisaDUAL

Přehled počtů pásem vybraných pomocí DA z datového souboru AisaDUAL pro tři různé referenční soubory uvádí tabulka 7. Na základě referenčního souboru obsahujícího trénovací prvky ze všech kategorií legendy bylo vybráno 33 pásem. Z tohoto výběru bylo následně odebráno 10 pásem, která se ukázala jako málo významná pro sedm prvních diskriminačních funkcí. Výsledný počet pásem důležitých pro odlišení všech kategorií legendy je 23, což je pouze 4,7 % z celkového počtu pásem vstupujících do DA. V případě zaměření analýzy na čtyři kategorie jednoděložných travin byl výstupem z DA větší počet pásem: 93 v případě menšího trénovacího souboru s pouze 3500 pixely kategorií travin, a 145 pásem pro rozšířený trénovací soubor s počtem pixelů travin přibližně odpovídajícím základnímu souboru se všemi kategoriemi vegetace (tj. okolo 12 000). Ukázalo se, že pro referenční soubory travin bylo sice vybráno více pásem, ale výběr zahrnoval vysoký počet pásem s nízkou významností pro všechny diskriminační funkce. Počet pásem výběru byl tak po úpravě 42 pásem pro základní a 47 pásem pro rozšířený trénovací soubor travin, což je 8,5 % a 9,5 % původních 493 pásem senzoru AisaDUAL.

	polygony	pixely	vybraná pásma (výstup z DA)	vybraná pásma (výběr podle koef.)	%
základní trénovací soubor všechny třídy	53	11933	33	23	4,7
základní trénovací soubor pouze traviny	20	3500	93	42	8,5
rozšířený trénovací soubor pouze traviny	58	11946	145	47	9,5

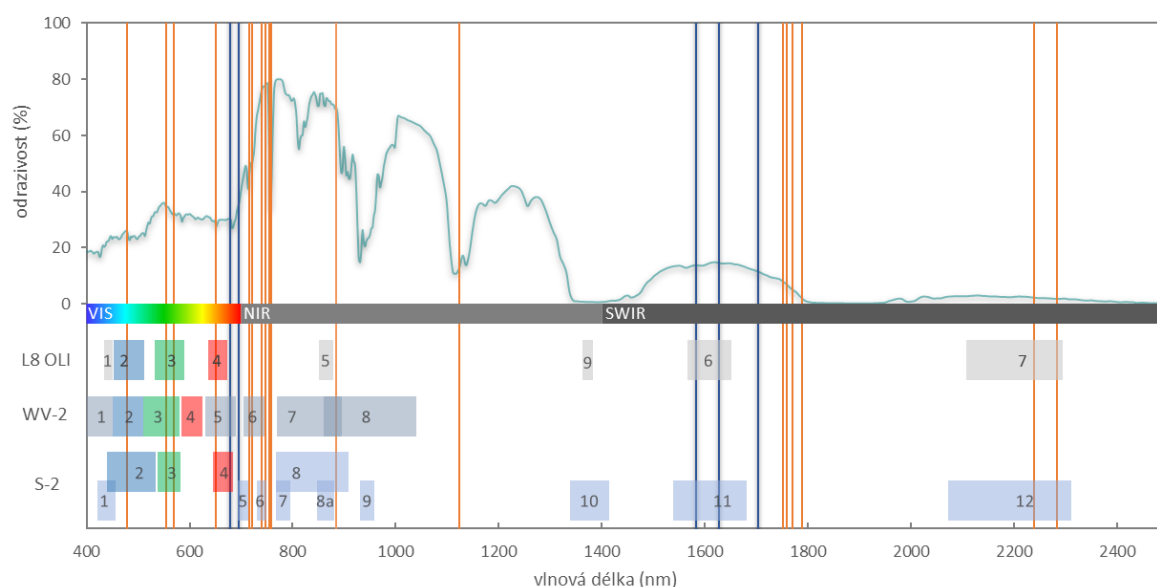
Tab. 7. Počet pásem senzoru AisaDUAL vybraný diskriminační analýzou.

	pásma ve VIS	pásma ve VIS (%)	pásma v NIR	pásma v NIR (%)	pásma v SWIR	pásma v SWIR (%)	celkem
základní trénovací soubor všechny třídy	6	26,1	8	34,8	9	39,1	23
základní trénovací soubor pouze traviny	6	14,3	15	35,7	21	50,0	42
rozšířený trénovací soubor pouze traviny	9	19,1	16	34,0	22	46,8	47

Tab. 8: Rozdělení vybraných pásem do částí spektra.

Tabulka 8 uvádí celkové a procentuální zastoupení pásem vyhodnocených jako důležitých pro oddělení skupin v částech spektra vlnových délek. V případě všech referenčních souborů byl pozorován stejný trend. Nejmenší zastoupení měla pásma viditelné části spektra, následovaná pásmy z oblasti NIR. Největší podíl významných pásem se ve všech třech případech nacházel v oblasti SWIR, přičemž pro základní trénovací soubor jednoděložných travin se v této části spektra nacházela až polovina všech vybraných pásem. Z tabulky 8 je také zřejmé, že podíly významných pásem v částech spektra VIS, NIR a SWIR jsou si podobnější v případě oddělení všech kategorií legendy. Při zaměření analýzy na čtyři travní kategorie kleslo procentuální zastoupení důležitých pásem ve viditelné části spektra a významnost pásem se přesunula do oblasti SWIR. Procentuální zastoupení významných pásem v oblasti NIR zůstalo pro všechny případy referenčních souborů přibližně stejné.

Obrázky 9–11 představují grafické znázornění pásem, která byla vybrána jako stěžejní na základě jednotlivých trénovacích souborů. Tmavě modré čáry v grafech znázorňují 5 nejlépe vyhodnocených vlnových délek (pásma s nejvyššími hodnotami koeficientů v první diskriminační funkci), oranžové čáry jsou ostatní vlnové délky vyhodnocené jako významné. Pro názornost je zobrazena také spektrální křivka jedné z tříd legendy, konkrétně se jedná o spektrum smilky. Ve spodních částech grafů jsou vyznačena spektrální pásma družic, které jsou předmětem zájmu této práce. Průniky linií vybraných důležitých pásem s polygony, představující oblasti snímání družic, tak demonstrují vhodnost použití družic pro klasifikace krkonošské tundry.

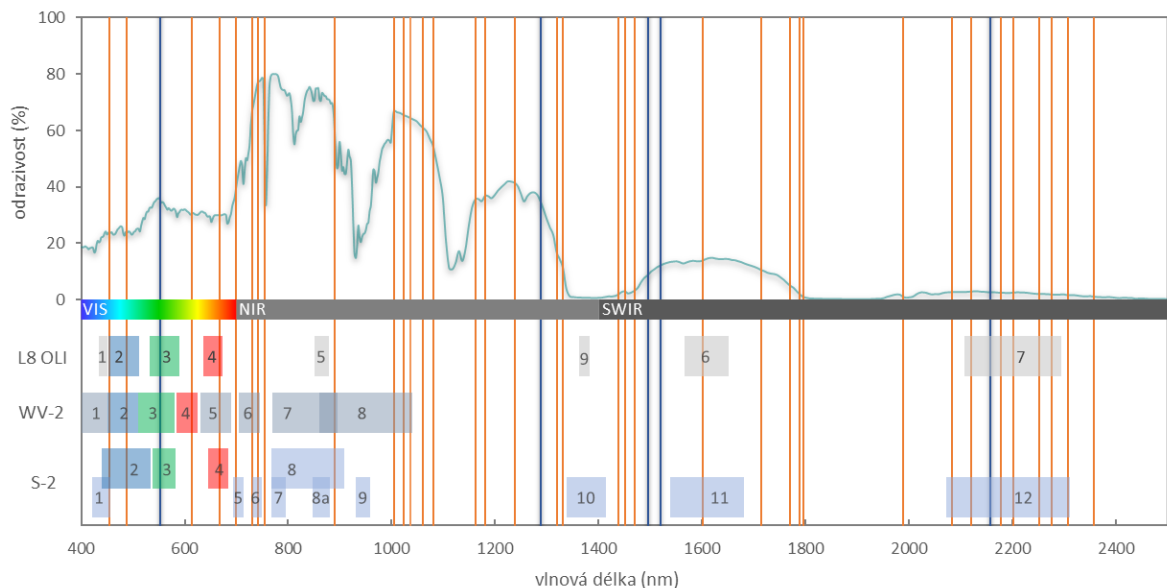


Obr. 9: Graf spektrálních pásem významných pro klasifikaci všech tříd legendy.

Na základě obrázku 9 lze definovat dvě základní oblasti vlnových délek, které vyšly jako stěžejní pro klasifikaci všech tříd legendy. První je oblast červeného záření, její konec, a přechod do oblasti NIR. Tato oblast je zdola ohraničená vlnovou délkou 650 nm a pokračuje až do vlnové délky 760 nm pásma NIR. Celkem se zde nachází 9 důležitých pásem. Pásma se středními vlnovými délkami 678 nm a 695 nm patří mezi 5 pásem, které významnost byla pro daný referenční soubor vyhodnocena jako nejvyšší. Druhou oblastí, do které spadá značná část pásem z výběru, je oblast SWIR. Významná pásma se zde vyskytují v rozmezí vlnových délek 1580 nm – 1790 nm. Z grafu na obrázku 9 je patrný koncentrovaný výskyt několika významných pásem v části spektra těsně před hranicí absorpčního pásma vodní páry, které začíná vlnovou délkou 1800 nm. Těmto pásmům byla při interpretaci výsledků věnována zvýšená pozornost, jelikož mohla být

již zatížena šumem. Pásma byla ale vyhodnocena jako vyhovující. Nejvýznamnější pásma v oblasti SWIR se nachází blíže k dolní hranici tohoto pásma a jejich středové vlnové délky jsou 1583 nm, 1627 nm a 1703 nm. Další pásma významná pro rozlišení tříd se podle očekávání nacházejí i ve viditelné části spektra, konkrétně na vlnových délkách zelené a modré.

Oblasti snímání družic naznačené ve spodní části grafu na obrázku 9 ukazují na zásadní nedostatek senzoru WorldView-2, kterým je absence spektrálních pásem v oblasti SWIR. Ostatní dva senzory zde disponují dvěma pásmy, která mohou být pro zlepšení přesnosti klasifikace klíčová. Ani jeden ze senzorů nicméně nesnímá v oblasti těsně před hranicí 1800 nm absorpčního pásma vodních par. Nespornou výhodou mise Sentinel-2 je rozšířená působnost v oblasti *red-edge*, která též vyšla jako důležitá a tuto oblast snímá velmi dobře i senzor WorldView-2. Naproti tomu Landsat 8 snímá v NIR až v delších vlnových délkách a pásmo *red-edge* v jeho spektrálním rozlišení chybí.

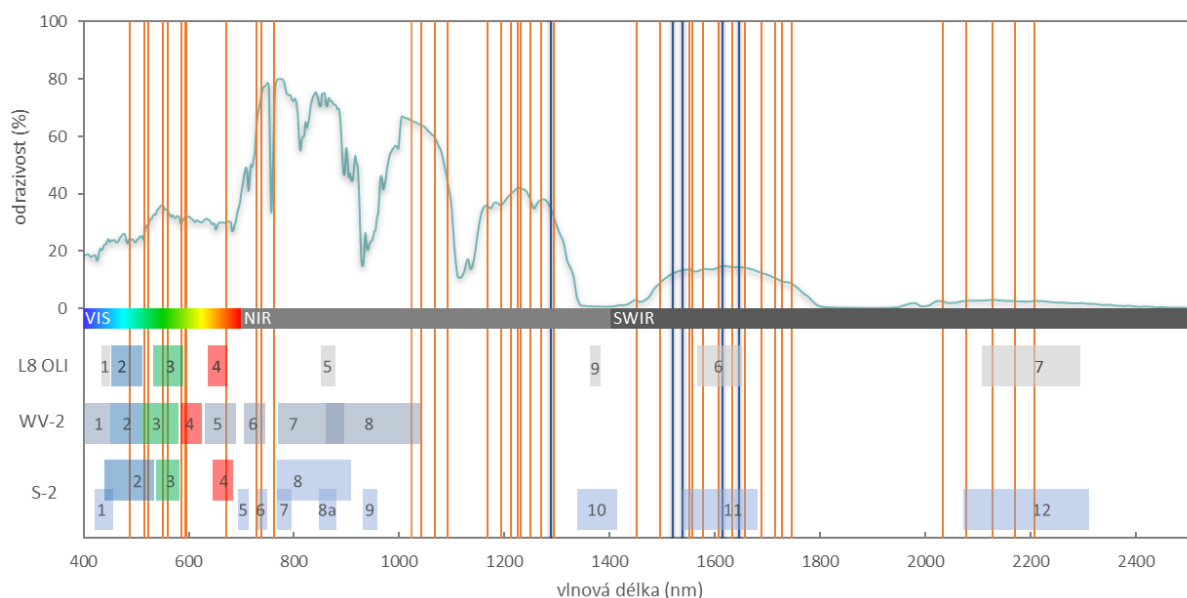


Obr. 10: Graf spektrálních pásem významných pro odlišení tříd jednoduchých travin, výsledek na základě základního referenčního souboru.

Když byl vstupem do DA základní trénovací soubor obsahující pouze třídy travin, počet vybraných pásem se zvýšil přibližně trojnásobně (obr. 10). Lze předpokládat, že to souvisí s podobností spektrální odezvy těchto tříd, kvůli které je náročné tyto třídy odlišit. Mezi tímto větším počtem vybraných pásem se ale také nacházelo více takových, která měla v porovnání s jinými nízkou významnost v diskriminačních funkcích, proto byla na

základě hodnocení koeficientů později vynechána. Z grafu obrázku 10 je patrné rovnoměrnější rozložení významných pásem po celém rozsahu hyperspektrálních dat. Porovnání s grafem na obrázku 9 ukazuje zvýšení výskytu těchto pásem v oblasti SWIR, a to i ve vzdálenější části, kde se nachází i jedno z pěti hlavních pásem vybraných pomocí DA. Další takové pásmo se v souladu s očekáváním nachází v oblasti zeleného záření, zbylé tři ohraničují absorpční pásmo vodních par na rozhraní oblasti NIR a SWIR v rozmezí vlnových délek 1340 nm – 1420 nm. Významná pásma se opět nachází také v oblasti *red-edge*, tentokrát na užším intervalu 670 nm – 755 nm a jejich počet je zde nižší. Rozdílem oproti předchozímu výsledku je výskyt důležitých pásem ve druhé polovině oblasti NIR.

Z hlediska vhodnosti družicových senzorů se opět ukazuje důležitost pásem snímaných v oblasti SWIR senzorů Landsat 8 a Sentinel-2. Dále je zřejmé, že pásma těchto senzorů ležící v oblasti absorpčního pásma vodních par (pásma „Cirrus“), zaměřená na výzkum oblačnosti, nejsou významná pro klasifikaci vegetace. Na druhou stranu, přínosem může být pásmo „Coastal“, které leží na začátku viditelné části spektra.



Obr. 11: Graf spektrálních pásem významných pro odlišení tříd jednoděložných travin, výsledek na základě rozšířeného referenčního souboru.

Rozšíření referenčního souboru polygonů travin o další polygony těchto kategorií přineslo lepší rozřazení významných pásem do shluků (obr. 11). To umožňuje přesněji definovat významné vlnové délky pomocí spojitých intervalů. Nejvyšší výskyt důležitých pásem je v tomto případě na začátku oblasti SWIR, kde se v rozmezí vlnových délek 1450 nm – 1750 nm nachází až čtyři z pěti pásem s nejlepším výsledkem a řada dalších významných pásem. Jako velmi významná se na základě analýzy tohoto referenčního souboru ukázala také vzdálenější oblast NIR ohraničená vlnovými délkami 1170 nm a 1300 nm. Lze pozorovat další pokles významnosti oblasti *red-edge*, který otvírá prostor pro diskuzi. Naopak, v oblasti viditelného záření významná pásma přibyla a došlo k jejich téměř rovnoměrnému rozložení v oblastech modrého a zeleného světla.

Na základě těchto výsledků se pro klasifikaci vegetace krkonošské tundry opět jako nejdůležitější ukazuje první spektrální pásmo SWIR senzorů Landsat 8 a Sentinel-2. Dále by podle důležitosti následovala druhé pásmo SWIR, pásma viditelného záření a pásma NIR těchto senzorů. Z obrázku 11 je patrné, že v rozmezí vlnových délek 1170 nm – 1300 nm, které bylo v této části analýzy identifikované jako velmi významné pro odlišení druhů, nesnímá žádný z dostupných senzorů. Do druhé poloviny oblasti NIR zasahuje svým spektrálním rozlišením pouze senzor WorldView-2, jeho rozsah ale zahrnuje pouze malou část pásem, která zde byla vyhodnocena jako významná.

5.2 Pásma družicových senzorů

Pro porovnání výsledků byla provedena DA i na základě simulovaných spektrálních rozlišení popisovaných senzorů, a to pro všechny tři vytvořené referenční soubory. Vzhledem k nižšímu počtu pásem vstupujících do DA jako nezávisle proměnných byl charakter výsledku analýzy odlišný od případu analýzy hyperspektrálních dat. Algoritmus ve většině případů označil jako významná všechna vstupní pásma, čímž byla znemožněna interpretace ve smyslu výběru pouze těch důležitých a eliminace těch, které nemají žádný nebo velmi malý vliv na separaci tříd.

Významnost pásem byla proto hodnocena na základě pořadí, ve kterém byla pásma v jednotlivých krocích analýzy zařazena do výběru. Lze předpokládat, že první vybrané pásmo bude mít největší vliv na odlišení kategorií, po něm druhé atd. Seřazení pásem podle významnosti vyhodnocené na základě jednotlivých referenčních souborů je obsahem tabulky 9. Pásma jsou v nich seřazena sestupně podle významnosti pro oddělení tříd definovaných v referenčním souboru.

V případě spektrálních pásem senzorů Landsat 8 a WorldView nebyla z krokové DA vyřazena žádná pásma. Když pomineme skutečnost, že Landsat 8 disponuje navíc pásmem SWIR, výsledky pro tyto senzory jsou si značně podobné. Pro rozlišení všech kategorií legendy se jako nejvýznamnější ukázalo pásmo Coastal (a to i v případě Sentinel-2), což je zajímavý výsledek, který bude dále rozebrán v diskuzní části práce. Ve výsledku pro tento trénovací soubor následují pásma NIR/SWIR a Red/Red Edge, což odpovídá rozložení důležitých pásem podle dat AisaDUAL. Když uvažujeme rozlišení pouze druhů travin, významnost těchto pásem vzrůstá. V případě senzoru Landsat 8 jsou pro oba referenční soubory travin méně významná pásma Blue, Green a Coastal, které si pouze mění pořadí ve spodních částech tabulky. Z výsledků pro pásma Red, NIR a SWIR je zřejmé, že na základě rozšířeného trénovacího souboru travin stoupá důležitost pásem SWIR, a klesá důležitost pásmo Red, čemuž odpovídá i míra koncentrace významných pásem vybraných v předcházející části práce v těchto oblastech spektra. Z pásem senzoru WorldView-2 jsou podle výsledků DA pro odlišení jednoděložní vegetace nejdůležitější pásma Red, Red Edge a NIR.

	základní trénovací soubor, všechny třídy		základní trénovací soubor, pouze traviny		rozšířený trénovací soubor, pouze traviny		
	číslo pásma	označení	číslo pásma	označení	číslo pásma	označení	
Landsat 8	1	Coastal	4	Red	5	NIR	
	5	NIR	5	NIR	6	SWIR 1	
	6	SWIR 1	6	SWIR 1	7	SWIR 2	
	4	Red	7	SWIR 2	4	Red	
	3	Green	2	Blue	1	Coastal	
	7	SWIR 2	3	Green	3	Green	
	2	Blue	1	Coastal	2	Blue	
WorldView-2	1	Coastal	5	Red	7	NIR 1	
	8	NIR 2	6	Red Edge	8	NIR2	
	7	NIR 1	3	Green	5	Red	
	5	Red	2	Blue	2	Blue	
	6	Red Edge	8	NIR 2	3	Green	
	3	Green	7	NIR 1	6	Red Edge	
	2	Blue	1	Coastal	4	Yellow	
	4	Yellow	4	Yellow	1	Coastal	
Sentinel 2	1	Coastal	4	Red	7	Red Edge 3	
	8	NIR	7	Red Edge 3	8a	Red Edge 4	
	7	Red Edge 3	11	SWIR 1	11	SWIR 1	
	11	SWIR 1	12	SWIR 2	12	SWIR 2	
	4	Red	1	Coastal	5	Red Edge 1	
	6	Red Edge 2	3	Green	3	Green	
	3	Green	2	Blue	2	Blue	
	12	SWIR 2	5	Red Edge 1	4	Red	
	2	Blue	6	Red Edge 2	6	Red Edge 2	
	pásma nezahrnuta ve výběru						
	9	Water Vapor	8	NIR	8	NIR	
	5	Red Edge 1	8a	Red Edge 4	9	Water Vapor	
	8a	Red Edge 4	9	Water Vapor			

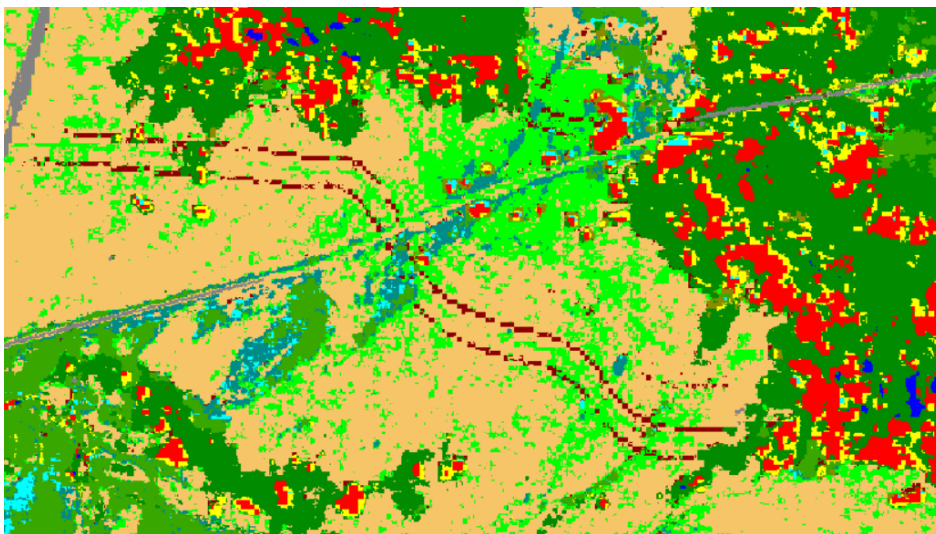
Tab. 9: Spektrální pásma družicových senzorů seřazená sestupně podle důležitosti pro rozlišení tříd klasifikace.

Výsledek DA pro pásma senzoru Sentinel-2 se ukázal jako nejnáročnější na interpretaci. Jedině v případě tohoto senzoru byl výsledný výběr pásem dosažený eliminací pásem s významností nízkou pro klasifikaci do skupin referenčních souborů. Společným znakem pro všechny referenční soubory byla eliminace pásma „Water Vapor“ z výběru, což je, vzhledem k poloze v absorpčním pásmu vodních par atmosféry a s tím souvisejícímu účelu pásma, správný výsledek. V souladu se stěžejními pásmy vybranými podle hyperspektrálních dat se se změnou klasifikovaných kategorií pouze na travinné

stávají důležitější vlnové délky SWIR. Platí to zejména pro pásmo SWIR 2, které bylo na základě referenčního souboru všech tříd vyhodnoceno jako méně důležité, což se ale změnilo v případě referenčních souborů travin. Rozdílným výsledkem oproti analýze hyperspektrálních dat je zejména významnost pásma Red Edge 3, které bylo ve všech případech vyhodnoceno jako poměrně důležité, kdežto pro hyperspektrální data nevyšla žádná význačná pásma s vlnovými délkami, které mu odpovídají. Z výsledků DA pro pásmo Sentinel-2 je rovněž zajímavý výrazný pokles důležitosti pásma NIR mezi referenčním souborem všech tříd a referenčními soubory travin, pro které bylo pásmo NIR dokonce vyřazeno z výběru.

5.3 Odstranění pásem zatížených šumem

V další části zpracování se přistoupilo ke klasifikaci. Výsledky klasifikací nicméně odhalily výskyt šumu ve formě pruhů (obr. 12). Původ pruhů není znám, může se jednat o mechanickou závadu skeneru. Nejvýraznější negativní dopad byl evidentní na výsledku klasifikátoru maximální pravděpodobnosti, který byl také testován spolu s klasifikátorem SVM. Výsledky klasifikace SVM byly šumem zasaženy méně, pruhy se zde ale vyskytly také. Na základě těchto výsledků byla zpětně provedena analýza vybraných pásem. Ukázalo se, že některá pásma, která byla výstupem diskriminační analýzy, byla šumem výrazně postižena, i když byla vyhodnocena jako důležitá. Pro odstranění vlivu šumu dat na klasifikaci bylo zkoumáno, zda nevyhovující pásma (s výskytem pruhů) nelze nahradit jiným pásem z jeho bezprostřední blízkosti, na kterém by se pruhy nevyskytovaly. Hodnota FWHM popisující šířku pásma se pro data AISA pohybuje v rozmezí 2,2 – 6,3 nm, proto rozdíl ve vlnové délce sousedních pásem je minimální. Pro každé nevyhovující pásmo byla vizuálně hodnocena sousední pásma maximálně do vzdálenosti tří pásem. Nevyhovující pásmo bylo pak ve výběru pásem nahrazeno blízkým pásmem nezasáženým šumem. V případě, že se v okolí nevyhovujícího pásma nenacházelo žádné vyhovující pásmo, toto pásmo bylo z výběru odstraněno bez nahrazení. Tabulka 10 je obměnou tabulky 7 po zohlednění výskytu šumu ve vybraných pásmech.



Obr. 12: Projev šumu dat senzoru AisaDUAL v klasifikaci povrchu.

	vybraná pásma (výstup z DA)	vybraná pásma (podle koef.)	pásma s pruhy	pásma s pruhy v %	pásma po úpravě
základní trénovací soubor všechny třídy	33	23	7	30,4	20
základní trénovací soubor pouze traviny	93	42	15	35,7	36
rozšířený trénovací soubor pouze traviny	145	47	10	21,3	44

Tab. 10: Úprava výběru pásem na základě zatížení šumem.

5.4 Výstupy klasifikace a hodnocení přesnosti

Na základě výběrů pásem nezatižených šumem byly z původního datového souboru senzoru AisaDUAL vytvořeny tři dílčí soubory složené pouze z vybraných pásem. Tuto možnost nabízí funkce *Spectral Subset* programu ENVI 5.4. Dalším krokem bylo provedení klasifikací SVM a hodnocení jejich přesnosti. Chybové matice a související hodnocení přesnosti výsledných klasifikací jsou v práci uvedeny jako přílohy 1, 2 a 3.

	počet pásem klasifikovaného obrazu	celková přesnost klasifikace (%)	Kappa koeficient
základní trénovací soubor, všechny třídy	20	80,49	0,7747
základní trénovací soubor, pouze traviny	36	80,54	0,7755
rozšířený trénovací soubor, pouze traviny	44	80,5	0,7742

Tab. 11: Přesnosti klasifikace pro jednotlivé vstupní soubory.

Všechny tři klasifikaci dosáhly velmi podobné celkové přesnosti i hodnot Kappa koeficientu (tab. 11). K rozdílným výsledkům došlo až na úrovni uživatelské (UA) a zpracovatelské (PA) přesnosti, nicméně ani zde není pozorovatelný trend (tab. 12), což je podrobněji komentováno v části 5. Tabulka 13 nese informaci o UA a PA všech tříd klasifikace 20 pásem získaných analýzou základního trénovacího souboru. Kompletní přehled UA a PA všech tříd pro všechny klasifikace je součástí příslušných příloh.

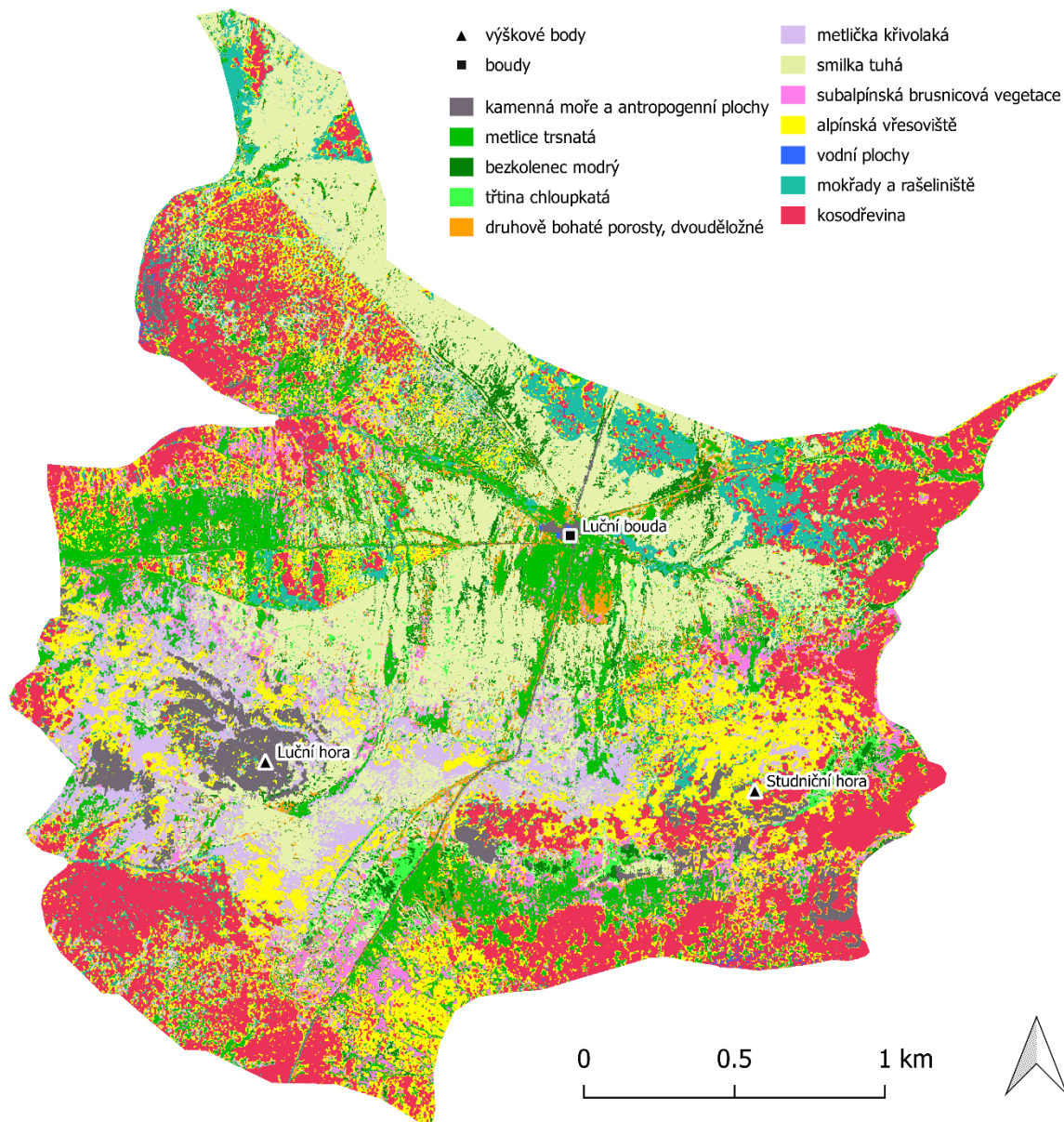
Pro názornost výsledku klasifikace byly v programu QGIS Desktop 2.18 vytvořeny mapové výstupy zobrazující třídy povrchu arкто-alpínské krkonošské tundry. Na obrázku 13 je výsledek klasifikace 44 pásem získaných analýzou rozšířeného referenčního souboru čtyř tříd jednoděložní vegetace. Další 2 výsledky klasifikace jsou v práci uvedeny jako přílohy 4 a 5.

Třída		základní trénovací soubor, všechny třídy		základní trénovací soubor, pouze traviny		rozšířený trénovací soubor, pouze traviny	
		PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
6a	smilka	82,5	77	79	76,9	83,1	72,6
7a	třtina	72,6	61,4	66,5	67,9	45,7	61,1
7b	bezkoleneec	48,5	71,7	62	64,3	61,3	68,5
7c	metlice	94,3	67	83,8	65,4	84,9	63,1

Tab. 12: Zpracovatelská a uživatelská přesnost hodnocených tříd jednoděložných travin.

Třída		PA (%)	UA (%)
1	Kosodřevina	99,9	99,29
2	Mokřad	58,64	59,91
3	Voda	98,49	100
4	Vřes	56,43	86,3
5	Brusnice	52,08	75,12
6a	Smilka	82,46	76,98
6b	Metlička	67,16	62,17
6c	Dvouděložné	72,34	53,97
7a	Třtina	72,57	61,42
7b	Bezkoleneec	48,47	71,73
7c	Metlice	94,32	66,97
8	Antropogenní	99,94	98,65

Tab. 13: Zpracovatelská a uživatelská přesnost všech tříd.



Obr. 13: Výsledek klasifikace SVM.

6 Diskuze

Klasifikace vegetace krkonošské arko-alpínské tundry je náročným úkolem, jehož úspěšnost závisí na mnoha faktorech. Složitost rozlišení jednotlivých klasifikačních tříd souvisí zejména s heterogenitou vegetace, podobností jejích spektrálních charakteristik a početným druhovým složením. Jak již bylo dokázáno v publikacích popsaných v části 2.2., bez závislosti na druhu použitých dat lze s poměrně vysokou přesností klasifikovat třídy vizuálně rozpoznatelné z ortofota, kterými jsou zejména kosodřevina, kamenná moře a antropogenní povrchy. U ostatních kategorií, které mají v mnoha případech velmi podobné spektrální vlastnosti (často dochází k jejich záměně), přesnost klasifikace záleží na rozlišení dostupných dat.

Jedním ze stanovených cílů práce bylo zjistit, která spektrální pásma jsou stěžejní pro odlišení vybraných vegetačních tříd a tímto způsobem zhodnotit vliv spektrálního rozlišení na přesnost klasifikace. Jako prostředek pro hodnocení separability tříd v konkrétních pásmech hyperspektrálního leteckého snímku, pořízeného pro výzkum krkonošské tundry, byla zvolena metoda krokové diskriminační analýzy (DA). Tuto metodu použili pro určení spektrálních pásem významných pro klasifikaci vegetace také například Prince a kol. (2002), Van Aardt a Wynne (2007), Prospere a kol. (2014) a Bajwa a kol. (2017).

Do DA vstupovaly tři datové soubory odpovídající trénovacím polygonům všech tříd základního referenčního souboru, trénovacím polygonům vybraných tříd základního referenčního souboru a trénovacím polygonům vybraných tříd rozšířeného referenčního souboru (viz část 4.1). Vybrané třídy představovaly čtyři druhy jednoděložných trav: smilky tuhé, třtiny chloupkaté, bezkolence modrého a metlice trsnaté. Tyto vegetační druhy jsou typické podobnými spektrálními charakteristikami a s tím související náročností odlišení v klasifikaci povrchu, zároveň je však nutný monitoring jejich rozšíření, neboť existuje předpoklad, že konkurenčně silné traviny (třtina, bezkolence) se rozpínají na úkor původní a pro tundru významné smilky tuhé (Hejcman a kol., 2009; Hejcman a kol., 2010). Účelem vytvoření tří vstupních souborů pro DA bylo ověření vlivu počtu a typu tříd obsažených ve vstupním souboru (všechny vs. pouze vybrané) a dále také zjištění, zda větší počet trénovacích polygonů pro vybrané třídy nepřinese zlepšení výsledku (základní vs. rozšířený referenční soubor).

Výstupem DA byly výběry pásem významných pro odlišení vstupních tříd (viz část 5.1). Poměr počtu vybraných pásem k celkovému počtu pásem hyperspektrálních dat byl 4,7 % pro všechny třídy v základním trénovacím souboru, 8,5 % pro traviny též v základním trénovacím souboru a 9,5 % pro traviny v rozšířeném souboru (23, 42 a 47 pásem). V případě většího počtu tříd stoupá meziskupinová variabilita vstupního souboru (centroidy skupin jsou více vzdálené), s čímž může souviset nižší počet pásem potřebných pro diskriminaci skupin.

Důležitým ukazatelem z hlediska hodnocení výběrů pásem je jejich rozložení v rámci spektrálního rozlišení vstupních dat (obr. 9, 10 a 11). Významné oblasti se nacházely zejména v zelené části spektra, na rozhraní částí červené a NIR a na začátku části SWIR. Ukázalo se, že zaměřením vstupního souboru na vybrané třídy travin dochází k vzrůstu důležitosti vlnových délek SWIR. Tucker (1980) uvádí, že důležitost snímání v oblasti SWIR souvisí se silnou korelací hodnot odrazivosti s obsahem vody ve vegetaci. Voda v zelených částech vegetace velkou část záření SWIR absorbuje a odrazivost v tomto pásmu je proto vůči obsahu vody nepřímou závislou. V pásmu NIR jsou naproti tomu hodnoty odrazivosti vegetace nezávislé na obsahu vody ve vegetaci. Zvýšená odrazivost v pásmu SWIR je indikátorem abiotického stresu rostliny z nedostatku vody a nepříznivého stavu vegetace obecně (Panigrahy a kol., 2009). Zjednodušeně řečeno, projevuje se zde „suchost“ vegetace. Vyhodnocení pásem SWIR jako důležitých pro rozlišení druhů trav může na základě těchto poznatků souviset s různou rychlostí růstu

vegetace po zimním období. Hyperspektrální letecká data byla pořízena v polovině června, kdy některé z druhů, zejména bezkoleneček modrý, ještě neprorostly suchou vegetací z minulého vegetačního období, což se projeví v pásmech SWIR.

Další oblast s vysokou koncentrací pásem významných pro odlišení vegetace vyšla v přechodu červené a infračervené části spektra. Tato oblast je v literatuře označována jako *red-edge* a je dobře známá jako oblast spektra významná pro klasifikaci vegetace. Odrazivost vegetace na tomto rozhraní prudce stoupá, protože zatímco chlorofyl v rostlinách pohlcuje značnou část červeného viditelného světla, v infračerveném pásmu nemá vliv a na odrazivosti se podílí spíše buněčná struktura rostliny (Tucker, 1980). Důležitost pásem v oblasti *red-edge* a jejím okolí byla očekávaným výsledkem, který se potvrdil. Zajímavým výsledkem nicméně zůstává skutečnost, že ani v jednom případě se v této oblasti nenacházela nejdůležitější pásma.

Klasifikace obrazů, obsahujících pásma dat AisaDUAL vybraná diskriminační analýzou, byly provedeny pomocí klasifikátoru SVM. Celkové přesnosti klasifikací dosáhly velmi podobných hodnot kolem 80 %, což je uspokojivý výsledek porovnatelný s výsledky dosaženými v Kupková a kol. (2017). Nejlepší celková přesnost klasifikace dat AisaDUAL dosažená v této studii byla 84,31 %, v případě použití klasifikátoru SVM na 40 transformovaných PCA pásem. Klasifikace všech ostatních kombinací vstupních dat (všechny původní pásma, 7 pásem PCA) a klasifikátorů (SVM, *neural network*) popsané ve studii dosáhly celkové přesnosti nižší než 80 %. Nejhorší výsledek klasifikátoru SVM byl 74,24 % při použití všech původních hyperspektrálních pásem. Diskriminační analýza tak byla prokázána jako vhodný nástroj pro redukci dimenzionality dat, rovnocenný s transformací PCA. Na rozdíl od PCA, která zohledňuje pouze variabilitu dat, DA ale zohledňuje také rozdělení dat podle skupin.

Na základě chybových matic byly zhodnoceny uživatelské a zpracovatelské přesnosti čtyř zkoumaných tříd travin. Ukázalo se, že žádná z klasifikací provedených v této práci odlišení tříd výrazně nezlepšuje (v porovnání s výsledkem nejlepší klasifikace z Kupková a kol. (2017)). V případě klasifikace 20 pásem vybraných na základě trénovacích ploch všech tříd byla s vysokými hodnotami PA a UA klasifikována smilka (82,5 % a 77%). Nejhorší výsledek byl zaznamenán pro bezkoleneček – pouze přibližně polovina validačních pixelů této třídy byla klasifikována správně (PA 48,4 %). Ostatní pixely byly přiřazeny převážně zbylým třem hodnoceným třídám travin, nejvíce třídě

třtiny. S tím souvisí nízká UA v případě třtiny (61,4 %) – došlo tedy k záměně těchto dvou tříd. Hodnoty PA a UA tříd v klasifikaci 36 pásem získaných diskriminační analýzou trénovacích ploch travin ze základního referenčního souboru se ukázaly jako více vyrovnané, tzn. došlo ke zvýšení PA bezkolence, avšak na úkor snížení PA ostatních tříd. Hodnoty PA a UA v tomto případě neklesly pod 62 % pro všechny třídy. Nejpřesněji byla opět klasifikována smilka. Poslední klasifikace 40 pásem vybraných podle rozšířeného referenčního souboru travin přinesla mírné zlepšení přesnosti pro tři ze čtyř tříd. Nepřesně klasifikovanou třídou byla v tomto případě třtina, jejíž PA byla pouze 45,7 %. Většina ze špatně klasifikovaných pixelů třtiny připadla třídě smilka. Porovnání uživatelských a zpracovatelských přesností klasifikací mezi sebou neukázalo žádný výrazný trend vedoucí ke zlepšení jejich odlišení, tedy nepotvrdily se předpoklady vlivu počtu a typu tříd/počtu prvků vstupního souborů na odlišitelnost tříd.

Dalším z cílů této práce bylo zhodnocení využitelnosti volně dostupných dat současných družic DPZ z hlediska jejich spektrálního rozlišení. Jednalo se o družice Sentinel-2, WorldView2 a Landsat 8. Porovnáním výsledků DA na základě dat AisaDUAL s rozsahy spektrálních pásem, na kterých tyto družice snímají, byla nejdůležitější pásma určena následovně: 6, 7, 4 a 3 pro Landsat 8; 3, 4 a 6 pro WorldView-2; 12, 11, 6, 5, 4, 3 pro Sentinel-2 (označení pásem podle části 2.5).

Data AisaDUAL byla následně převzorkována na spektrální rozlišení družic. Porovnáním výsledků DA hyperspektrálních dat a DA simulovaných spektrálních rozlišení byla odhalena pouze částečná shoda ve výběru vlnových délek stěžejních pro odlišení kategorií legendy krkonošské tundry. Hlavním rozdílem bylo, že jako důležitá byla vyhodnocena v případě všech družic téměř všechna pásma. Pořadí, ve kterém kroková metoda DA vybírala pásma postupně podle důležitosti odpovídalo předpokládané významnosti pásem. V tabulce 9 je vidět, že na prvních pozicích v seznámkách pásem jsou převážně pásma SWIR, NIR, *red-edge* a červené. V případě družice Sentinel-2 došlo správně k vyřazení pásma Water Vapor, které neneslo informace vztahované k vegetaci. Vyřazení pásma NIR a některého z pásem *red-edge* nejspíš souvisí s vysokou vzájemnou korelací pásem 3, 4, 5, 6 a 8. Po přidání jednoho nebo vícera z těchto pásem do výběru se informace obsažená v dalším stává redundantní. Ne úplně očekávaným výsledkem, který se nepodařilo vysvětlit, bylo umístění pásma Coastal na pozici nejdůležitějšího pro diskriminaci tříd v případě analýzy základního trénovacího souboru se všemi třídami, a to pro všechny tři senzory. Pásma Coastal je zaměřené na

výzkum mělkých pobřežních vod a aerosolů v atmosféře a je výrazně korelované s modrým pásmem.

V celkovém hodnocení vybraných družic pro účel klasifikace krkonošské tundry vychází nejhůře senzor WorldView-2, který nedisponuje spektrálním rozlišením ve vzdálenějších částech oblasti NIR a v oblasti SWIR. Diskriminační analýza však v těchto oblastech určila významný podíl stěžejních pásem pro odlišení daných vegetačních tříd. Ostatní dva senzory svým spektrálním rozlišením tyto oblasti pokrývají. Tuto skutečnost reflektují výsledky studie Suchá a kol. (2016), ve které přesnost klasifikace dat Landsat 8 předčila klasifikaci dat WorldView-2 řádově o 10–20 % v případě všech použitých klasifikačních algoritmů (nejlepší dosažená celková přesnost klasifikace 78,31 %). V souvislosti s touto prací je nutno zmínit také prostorové rozlišení, které, jak se ukázalo, hraje v případě heterogenního vegetačního pokryvu tundry velkou roli. V pracích Suchá a kol. (2016) a Kupková a kol. (2017) bylo v případě dat s nižším prostorovým rozlišením přistoupeno k použití zjednodušené legendy shlukující některé z původních kategorií. Pro legendu popsanou v této práci je použití prostorového rozlišení uvedených družic rovněž nevhodné.

7 Závěr

Cílem práce bylo popsat a zhodnotit vliv spektrálního rozlišení na klasifikaci krajinného pokryvu v krkonošské tundře. Hlavním datovým souborem pro práci byla letecká hyperspektrální data senzoru AisaDUAL se spektrálním rozsahem 400–2500 nm a celkem 493 spektrálními pásmy. Byly vytvořeny referenční soubory trénovacích dat, zaměřené jak na odlišení všech tříd krajinného pokryvu, tak vybraných vegetačních tříd. V případě vybraných tříd se jednalo o tyto třídy jednoděložné vegetace: smilka tuhá, třtina chloupkatá, bezkoleneček modrý a metlice trsnatá. Za účelem ověření vlivu většího počtu trénovacích polygonů pro vybrané třídy byl vytvořen také rozšířený referenční soubor.

Jako prostředek pro hodnocení separability tříd v konkrétních pásmech senzoru AisaDUAL byla zvolena metoda krokové diskriminační analýzy implementovaná v programu IBM SPSS Statistics. Touto metodou byl pro každý referenční soubor vytvořen výběr hyperspektrálních pásem stěžejních pro rozlišení definovaných tříd. Pásma vybraná diskriminační analýzou byla graficky naznačena v rozsahu elektromagnetického spektra snímaného senzorem AisaDUAL (obr. 9, 10, 11). Podle četnosti jejich výskytu byly definovány intervaly vlnových délek stěžejní pro diskriminaci druhů krajinného pokryvu v krkonošské tundře. Nejvýznamnějšími intervaly byly 1500–1750 nm, což odpovídá začátku oblasti krátkých vlnových délek infračerveného záření (SWIR); 1100–1300 nm, tedy delší vlnové délky oblasti blízkého infračerveného záření (NIR); interval 670–760 nm odpovídající přechodu červeného pásma a pásma NIR (*red-edge*) a 500–600 nm, tudíž zelené pásmo ve viditelné části spektra.

Klasifikace datových souborů tvořených pásmo vybranými pomocí diskriminační analýzy byly provedeny v programu ENVI 5.4. Vybraným klasifikátorem byl *Support Vector Machine*. Všechny klasifikace dosáhly velmi podobných celkových přesností kolem 80 % (nejlepší 80,54 %, Kappa koeficient 0,7755). Tento výsledek potvrdil významnost vybraných spektrálních pásem pro klasifikaci krajinného pokryvu v krkonošské tundře. Na druhou stranu, shodnost výsledku pro všechny referenční soubory nepotvrdila předpoklad, že zaměřením trénovacího souboru pro diskriminační analýzu na konkrétní třídy, anebo jeho rozšířením, lze dosáhnout zpřesnění klasifikace.

V další části práce byl datový soubor senzoru AisaDUAL převzorkován pomocí funkce *Spectral Resampling*, čímž byla simulována spektrální rozlišení družic Sentinel-2, WorldView-2 a Landsat 8. Pásma družic byla rovněž analyzována pomocí diskriminační analýzy. Ta potvrdila významnost pásem SWIR, NIR a *red-edge*. Vhodnost dat vybraných družic pro odlišení druhů krkonošské tundry z hlediska spektrálního rozlišení tedy vyžaduje snímání v těchto pásmech. Z toho důvodu se jeví jako nejméně vhodný zdroj satelitních dat družice WorldView-2, která nesnímá v oblasti SWIR. Naopak nejlepším zdrojem je z hlediska spektrálního rozlišení Sentinel-2, který disponuje rozšířeným snímáním pásma *red-edge* a dvěma pásmo v oblasti SWIR. V realitě pak však nastává problém s prostorovým rozlišením, které je pro účely krkonošské tundry v případě Sentinelu-2 (i Landsat 8) malé.

Práce stanovila vliv spektrálního rozlišení na klasifikaci krajinného pokryvu v krkonošské tundře definováním spektrálních intervalů, které jsou nejvýznamnější pro odlišení jednotlivých tříd. Tato informace může být využita v dalším studiu vegetace této oblasti. Příkladem využití je správné nastavení multispektrálních kamer UAV, které pak mohou pořizovat data „na míru“ pro účely klasifikace vegetace. Postup navržený v této diplomové práci může být dále aplikován, například ve smyslu zohlednění vlivu fenologické fáze vegetace, tj. lze zkoumat změnu významnosti vlnových délek s postupem vegetačního období.

Použitá literatura a zdroje

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

Bajwa S.G., Rupe J.C., Mason J., 2017. Soybean Disease Monitoring with Leaf Reflectance. *Remote Sensing*, 9 (127).

Beamish A. L., Coops N., Chabrillat S., Heim B., 2017. A Phenological Approach to Spectral Differentiation of Low-Arctic Tundra Vegetation Communities, North Slope, Alaska. *Remote Sensing*, 9(11), 1200.

Bratsch S. N., Epstein H. E., Buchorn M., Walker D. A, 2016. Differentiating among Four Arctic Tundra Plant Communities at Ivotuk, Alaska Using Field Spectroscopy. *Remote Sensing*, 8(1), 51.

Cingolani A. M., Renison D., Zak M. R., Cabido M. R., 2004. Mapping vegetation in a heterogeneous mountain rangeland using landsat data: an alternative method to define and classify land-cover units. *Remote Sensing of Environment*, 92(1), 84–97.

Deng H., Runger G., 2013. Gene selection with guided regularized random forest. *Pattern Recognition*, 46, 3483–3489.

Fisher, R. A, 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7, 179-188.

Fraser R. H., Olthof I., Carrière M., Deschamps A., Pouliot D., 2011. Detecting long-term changes to vegetation in northern Canada using the Landsat satellite image archive. *Environmental Research Letters*, 6, 1–9.

Gábik J., 2010. Kreditný skóring pomocou náhodných lesov. *Diplomová práca (Ing.)*. Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky.

Geodis Brno, 2013. Hyperspektrální snímkování a tvorba map vegetace a map inventarizace lesních porostů – technická zpráva za rok 2013.

Ghulam A., Fishman J., Maimaitiyiming M., 2016. Spectral separability analysis of five soybean cultivars with different ozone tolerance using hyperspectral field spectroscopy. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, 2016, 6312-6315.

- Hejzman M., Klaudivová M., Hejmanová P., Pavlů V., Jones M., 2009. Expansion of *Calamagrostis villosa* in sub-alpine *Nardus stricta* grassland: Cessation of cutting management or high nitrogen deposition? *Agriculture, Ecosystems and Environment*, 129, 91–96.
- Hejzman M., Češková M., Pavlů V., 2010. Control of *Molinia caerulea* by cutting management on sub-alpine grassland. *Flora*, 205, 577–582.
- Hsu P., 2007. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 78–92.
- Huberty, C.J. a Olejnik S. Applied MANOVA and Discriminant Analysis. *New York: Wiley Interscience*, 2006. ISBN 0-471-46815-0.
- Chasmer L., Hopkinson C., Veness T., Quinton W., Baltzer J., 2014. A decision-tree classification for low-lying complex land cover types within the zone of discontinuous permafrost. *Remote Sensing of Environment*, 143, 73–84.
- Jakešová L., 2014. Možnosti objektově-orientované klasifikace pro určování vybraných biotopů nad horní hranicí lesa v krkonošském národním parku. *Bakalářská práce (Bc.)*. Univerzita Karlova, Přírodovědecká fakulta, Katedra aplikované geoinformatiky a kartografie.
- Jarocińska A. M., Kacprzyk M., Marcinkowska-Ochtyra A., Ochtyra A., Zagajewski B., Meuleman K., 2016. The application of APEX images in the assessment of the state of non-forest vegetation in the Karkonosze Mountains. *Miscellanea Geographica*, 20 (1).
- Kalaczka M, Sanchez-Azofeifa A. G. Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests. *CRC Press*, 2008. ISBN 9781420053418.
- 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.
- Langford Z, Kumar J., Hoffman F. M., Norby R. J., Wulschleger S. D., Sloan V. L., Iversen C. M., 2016. Mapping Arctic Plant Functional Type Distributions in the Barrow Environmental Observatory Using WorldView-2 and LiDAR Datasets. *Remote Sensing*, 8(9), 733.
- Marcinkowska A., Zagajewski B., Ochtyra A., Jarocińska, A., Raczko E., Kupková L., Stych P., Meuleman K., 2014. Mapping vegetation communities of the Karkonosze National Park using APEX hyperspectral data and Support Vector Machines. *Miscellanea Geographica*, 18(2), 23-29.
- Mureriwa N., Adam E., Sahu A., Tesfamichael S., 2016. Examining the Spectral Separability of *Prosopis glandulosa* from Co-Existent Species Using Field Spectral Measurement and Guided Regularized Random Forest. *Remote Sensing*, 8(2), 144.

Müllerová J., 2004. Use of digital aerial photography for sub-alpine vegetation mapping: A case study from the Krkonoše Mts., Czech Republic. *Plant Ecology*, 175(2), 259–272.

Neil H.T. *Applied Multivariate Analysis*. New York: Springer-Verlag, 2002. ISBN 0-387-95347-7.

Panigrahy R.K., Ray, S.S., Panigrahy S., 2009. Study on the Utility of IRS-P6 AWiFS SWIR Band for Crop Discrimination and Classification. *Journal of the Indian Society of Remote Sensing*, 37, 325-333.

Pejčoch, D. (2011): Metody řešení problematiky neúplných dat. *Forum Statisticum Slovacum*, 7, č. 7, s. 187–192.

Petropoulos G.P., Arvanitis K., Sigrimis N., 2012. Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. *Expert Systems with Applications*, 39(3), 3800–3809.

Pomahačová M., 2012. Možnosti využití DPZ při monitoringu luční vegetace a managementových zásahů v Krkonoších. *Diplomová práce (Mgr.)*. Univerzita Karlova, Přírodovědecká fakulta, Katedra aplikované geoinformatiky a kartografie.

Potůčková M., Červená L., Kupková L., Lhotáková Z., Albrechtová J., 2016. Statistical comparison of spectral and biochemical measurements on an example of Norway spruce stands in the Ore Mountains, Czech Republic. *Geoinformatics FCE CTU*, 15 (1), 69-83.

Prospere K., McLaren K., Wilson B., 2014. Plant Species Discrimination in a Tropical Wetland Using In Situ Hyperspectral Data. *Remote Sensing*, 6, 8494-8523.

Pu R., 2008. An exploratory analysis of in situ hyperspectral data for broadleaf species recognition. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, B7.

Rabasová M. 2012. Diskriminační analýza jako nástroj pro hodnocení chirurgických rizik. *Dizertační práce*. Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky.

Radoux J, Chomé G., Jacques D.C., Waldner F., Bellemans N., Matton N, Lamarche C., d'Andrimont R., Defourny P, 2016. Sentinel-2's Potential for Sub-Pixel Landscape Feature Detection. *Remote Sensing*, 8, 488.

Rao, C. R., 1948. The Utilization of Multiple Measurements in Problems of Biological Classification. *Journal of the Royal Statistical Society, Series B (Methodological)*, 10 (2), 159-203.

Reese, H., 2011. Classification of Sweden's Forest and Alpine Vegetation Using Optical Satellite and Inventory Data. *Doctoral thesis*. Swedish University of Agricultural Sciences, Faculty of Forest Sciences, Department of Forest Resource Management.

Reese H., Nyström M., Nordkvist K., Olsson H., 2014. Combining airborne laser scanning data and optical satellite data for classification of alpine vegetation. *International Journal of Applied Earth Observation and Geoinformation*, 27(A), 81-90.

Roubalová M., 2017. Klasifikace vybraných vegetačních kategorií land cover v krkonošské tundře z dat Sentinel-2a s využitím časové řady dat. *Bakalářská práce (Bc.)*. Univerzita Karlova, Přírodovědecká fakulta, Katedra aplikované geoinformatiky a kartografie.

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

Savić M., Brcanov D., Dakić S., 2008. Discriminant Analysis – Applications and Software Support. *Management Information Systems*, 3 (1), 029-033.

Suchá R., Jakešová L., Kupková L., Červená L., 2016. Classification of vegetation above the tree line in the Krkonoše Mts. National Park using remote sensing multispectral data. *AUC Geographica*, 51(1), 113–129.

Tonhauserová Z., 2013. Metoda hlavních komponent a její aplikace. *Diplomová práce (Ing.)*. Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra matematické analýzy a aplikací matematiky.

Tucker C.J., 1980. Remote Sensing of leaf water content in the near infrared. *Remote sensing of Environment*, 10, 23–32.

Ullmann T., Schmitt A., Roth A., Duffe J., Dech S., Hubberten H-W., Baumhauer R., 2014. Land Cover Characterization and Classification of Arctic Tundra Environments by Means of Polarized Synthetic Aperture X- and C-Band Radar (PolSAR) and Landsat 8 Multispectral Imagery — Richards Island, Canada. *Remote Sensing*, 6(9), 8565-8593.

Van Aardt J., Wynne R., 2007. Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote Sensing*, 28 (2), 431–436.

Vapnik V.N. Statistical learning theory. *New York: Wiley*, 1998. ISBN 0-471-03003-1

Virtanen T., Ek M., 2014. The fragmented nature of tundra landscape. *International Journal of Applied Earth Observation and Geoinformation*, 27, 4–12.

Weisa M., Gerards R. *Feature extraction for the identification of weed species in digital images for the purpose of site-specific weed control*. Precision Agriculture '07. Wageningen Academic Publishers, 2007. ISBN 978-90-8686-024-1.

Internetové zdroje:

ARCDATA Praha, 2018. *Družice Landsat* [online] [cit. 2018-02-11]. Dostupné z:

<<https://www.arcdata.cz/produkty/geograficka-data/druzicova-data/landsat>>

Blažek L. a kolektiv, 2018. *Konkurenční schopnost podniků (Analýza faktorů hospodářské úspěšnosti)*

[online] [cit. 2018-03-14]. Dostupné z:

<https://is.muni.cz/el/1456/podzim2008/D_TRP/um/monografie4_nahled01.pdf>

Copernicus, 2016. *Družice Sentinel* [online] [cit. 2018-02-11]. Dostupné z:

<<http://copernicus.gov.cz/druzice-sentinel>>

ESA, 2018. *Sentinel Handbook – Resolution and Swath* [online] [cit. 2018-02-11]. Dostupné z:

<<https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath>>

Harris Geospatial Solutions, 2018a. *Spectral Library Resampling* [online] [cit. 2018-01-15].

Dostupné z: <<http://www.harrisgeospatial.com/docs/spectrallibraries.html#Selectin>>

Harris Geospatial Solutions, 2018b. *Calculate Confusion Matrices* [online] [cit. 2018-02-25]. Dostupné z:

<<http://www.harrisgeospatial.com/docs/CalculatingConfusionMatrices.html>>

Kalina J., Tebbens J. D., 2013. *Metody pro redukci dimenze v mnohorozměrné statistice a jejich výpočet*

[online] [cit. 2018-03-14]. Dostupné z: <<http://www.cs.cas.cz/duintjertebbens/pubs/KalinaDT.pdf>>

LANDInfo, 2018. *WorldView-2 50cm Global High-Resolution Satellite Imagery* [online] [cit. 2018-02-

11]. Dostupné z: <<http://www.landinfo.com/WorldView2.htm>>

Landsat Science, 2018. *Spectral Response of the Operational Land Imager In-Band, Band-Average Relative Spectral Response* [online] [cit. 2018-02-11]. Dostupné z:

<<https://landsat.gsfc.nasa.gov/preliminary-spectral-response-of-the-operational-land-imager-in-band-band-average-relative-spectral-response>>

Nikhil Buduma, 2015. *The Curse of Dimensionality and the Autoencoder* [online] [cit. 2018-04-2].

Dostupné z: <http://nikhilbuduma.com/2015/03/10/the-curse-of-dimensionality/>

Rencher A. C., 2002. *Methods of Multivariate Analysis, Second Edition* [online] [cit. 2018-04-2].

Dostupné z: <<https://www.ipen.br/biblioteca/slr/cel/0241>>

Správa KRNAP, 2010. *Krkonošská arкто-alpínská tundra* [online] [cit. 2018-03-28]. Dostupné z:

<<http://www.krnap.cz/krkonosska-arkto-alpinska-tundra>>

StatSoft, Inc., 2013. *Electronic Statistics Textbook. Stepwise Discriminant Analysis* [online] [cit. 2018-

03-28]. Dostupné z: <<http://www.statsoft.com/Textbook/Discriminant-Function-Analysis>>

| Seznam příloh

Příloha 1: Chybová matice klasifikace SVM datového souboru 20 pásem

Příloha 2: Chybová matice klasifikace SVM datového souboru 36 pásem

Příloha 3: Chybová matice klasifikace SVM datového souboru 44 pásem

Příloha 4: Klasifikace 20 pásem

Příloha 5: Klasifikace 36 pásem