

**Česká zemědělská univerzita v Praze**

**Fakulta životního prostředí**

**Katedra prostorových věd**



**Možnosti detekce lesních biotopů Natura  
2000 pomocí dat Sentinel-2**

**Bakalářská práce**

Vedoucí práce: doc. Ing. Petra Šímová, Ph.D.

Bakalant: Kristýna Nalezinková

© 2021 ČZU v Praze BAKALÁŘSKÁ PRÁCE

Kristýna Nalezinková



Česká zemědělská univerzita v Praze  
Fakulta životního prostředí

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Autorka práce:	Kristýna Nalezinková
Studijní program:	Geografické informační systémy a dálkový průzkum Země v životním prostředí
Vedoucí práce:	doc. Ing. Petra Šimová, Ph.D.
Garantující pracoviště:	Katedra prostorových věd
Jazyk práce:	Čeština
Název práce:	<b>Možnosti detekce lesních biotopů Natura 2000 pomocí dat Sentinel-2</b>
Název anglicky:	<b>Possibilities of detection of Natura 2000 forest habitats using Sentinel-2 data</b>
Cíle práce:	Česká republika disponuje velmi prostorově i tematicky podrobným mapováním přírodě blízkých biotopů, pořízeným zejména pro účely naplňování směrnice 92/43/EEC, o ochraně přírodních stanovišť, volně žijících živočichů a planě rostoucích rostlin. Tato Vrstva mapování biotopů (VMB) je od dob svého vzniku průběžně aktualizována terénním průzkumem, přičemž aktualizace celého území ČR trvá 12 let. Nabízí se tak využití volně dostupných družicových dat pro pomoc s mapováním, na druhou stranu ale přesnost klasifikace do podrobných tříd odpovídajících VMB může být problematická. Rámcovým cílem bakalářské práce je zjistit, zda a do jaké míry lze s využitím dat družic Sentinel-2 detekovat vybrané typy lesních biotopů v modelovém území. Formulace dílčích cílů je úkolem řešitelky.
Metodika:	Rámcové body postupu jsou následující: (1) Výběr modelového území a explorativní analýza kategorií lesních biotopů VMB v něm. (2) Příprava trénovacích a validačních dat na základě VMB. (3) Řízená klasifikace pomocí alespoň dvou machine learning algoritmů, popř. doplnkově neřízená klasifikace. (4) Vyhodnocení přesnosti klasifikace (chybová matice, porovnání s VMB). (5) Interpretace a diskuse výsledků. K analýzám a prezentaci výsledků řešitelka dle vlastního uvážení využije programy ArcGIS, R, SNAP a prostředí Google Earth Engine. Postupy zautomatizuje v podobě skriptů a/nebo work-flow modelů (např. Model Builder).

Doporučený rozsah práce: 30 - 60 stran

Klíčová slova: dálkový průzkum Země, Sentinel-2, klasifikace, random forest, CART, Vrstva mapování biotopů, Google Earth Engine

Doporučené zdroje informací:

1. Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., Strasser, T., vanden Borre, J., Toon, S., & Michael, F. (2015). Remote sensing for mapping natural habitats and their conservation status - New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation*, 37.
2. Grabska, E., Frantz, D., & Ostapowicz, K. (2020). Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. *Remote Sensing of Environment*, 251, 112103.
3. Chytrý, M., Kučera, T., Kočí, M., Grulich, V., & Lustyk, P. (2010). Katalog biotopů České republiky. Druhé vydání.
4. Nomura, K. & Mitchard, E.T.A. 2018. More than meets the eye: using Sentinel-2 to map small plantations in complex forest landscapes. *Remote Sensing*, doi:10.3390/rs10111693.
5. Schmidt, J., Fassnacht, F. E., Neff, C., Lausch, A., Kleinschmit, B., Förster, M., & Schmidlein, S. (2017). Adapting a Natura 2000 field guideline for a remote sensing-based assessment of heathland conservation status. *International Journal of Applied Earth Observation and Geoinformation*, 60, 61–71.

Předběžný termín obhajoby: 2020/21 LS - FŽP

Elektronicky schváleno: 23. 3. 2021  
**doc. Ing. Petra Šimová, Ph.D.**  
Vedoucí katedry

Elektronicky schváleno: 24. 3. 2021  
**prof. RNDr. Vladimír Bejček, CSc.**  
Děkan

### **Čestné prohlášení**

Prohlašuji, že svou bakalářskou práci " Možnosti detekce lesních biotopů Natura 2000 pomocí dat Sentinel-2 " jsem vypracovala samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autorka uvedené bakalářské práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 31. 3. 2021

---

### **Poděkování**

Ráda bych touto cestou poděkovala vedoucí mé bakalářské práce doc. Ing. Petře Šimové, Ph.D. za její cenné rady, trpělivost a věnovaný čas.

Tato bakalářská práce byla řešena v rámci projektu TAČR SS01010046 – Možnosti aktualizace mapování biotopů NATURA 2000 pokročilými metodami dálkového průzkumu Země (SentiMAP).

# Možnosti detekce lesních biotopů Natura 2000 pomocí dat Sentinel-2

## Abstrakt

Tato bakalářská práce je zaměřena na klasifikaci lesních biotopů Natura 2000 pomocí multispektrálních snímků družice Sentinel-2. Jejím cílem bylo zjistit, jaké podrobnosti lesů dané Katalogem biotopů lze dosáhnout pomocí řízené klasifikace v Google Earth Engine. Dílčí cíle této práce byly: S jakou přesností se rozliší les od ostatních biotopů a s jakou přesností se odliší různé třídy lesů mezi sebou na první a druhé hierarchické úrovni.

Z Vrstvy mapování biotopů byly vybrány jen ty lesy, které byly aktualizovány od roku 2015 a zároveň měly celkem větší rozlohu než 1 km<sup>2</sup>. Lesy, které splňovaly kritéria, byly listnaté. Konkrétně se jednalo o mokřadní olšiny, lužní lesy, dubohabřiny a acidofilní doubravy. Z nich byly vytvořeny trénovací data, které vstupovaly do řízené klasifikace pomocí algoritmu random forest a CART.

Klasifikace dokázala rozlišit les od ostatních biotopů s 97% přesností u obou algoritmů. U zjišťování přesnosti klasifikace lesů na první hierarchické úrovni dané Katalogem biotopů ČR (Chytrý et al., 2010) bylo dosaženo celkové přesnosti 52,2 % pomocí algoritmu random forest. Nejlépe byly určeny lužní lesy a acidofilní doubravy s téměř 75% přesností. Přesnost klasifikace lesů na druhé hierarchické úrovni dané Katalogem biotopů ČR (Chytrý et al., 2010) dosáhla přesnosti klasifikace 43,5 % pomocí algoritmu random forest. Klasifikace pomocí algoritmu CART byla přibližně o 11 % horší u určování první i u druhé hierarchické úrovně lesů.

**Klíčová slova:** dálkový průzkum Země, Sentinel-2, klasifikace, random forest, CART, Vrstva mapování biotopů, Google Earth Engine

# Possibilities of detection of Natura 2000 forest habitats using Sentinel-2 data

## Abstract

This bachelor thesis is focused on habitat classification based on Habitat Catalogue of the Czech Republic (Chytrý et al., 2010) using multispectral images of the Sentinel-2 satellite. Its aim was to find out, what forest details given by the Habitat Catalogue can be achieved using the supervised classification in the Google Earth Engine. The partial aims of this bachelor thesis were: with what precision will the forest be distinguish from other biotops and with what precision will distinguish different classes of forests from each other to the first and the second hierarchical levels.

From the habitat mapping layer were selected only those forests with area larger than 1 km<sup>2</sup> and were updated latest in 2015. To those criteria correspond deciduous forests. Specifically, these were alder carrs, alluvial forests, oak-hornbeam forests and acidophilous oak forests. From these were created training data which enter the supervised classification using “random forest” and “CART” algorithm. Classification achieves to recognize forests from other habitats with 97% accuracy at both algorithms same. The analysis of the forest classification accuracy at 1st hierarchical level shows 52,2 % overall accuracy by using “random forest”. The best result was achieved for alluvial forests and acidophilous oak forests which were identified with 75% accuracy. The forest classification accuracy at 2nd hierarchical level shows 43,5 %. The CART algorithm shows 11 % worse result at both hierarchical levels.

**Keywords:** remote sensing, Sentinel-2, classification, random forest, CART, habitat mapping layer, Google Earth Engine

## Obsah

Seznam zkratk použitých v textu .....	10
1 Úvod.....	11
2 Cíle práce .....	12
3 Literární rešerše .....	13
3.1 Natura 2000 a Vrstva mapování biotopů.....	13
3.2 Vybrané technické aspekty využití DPZ v ochraně přírody.....	14
3.2.1 Sentinel-2.....	14
3.2.2 Klasifikace .....	16
3.2.3 Spektrální chování objektů .....	20
3.3 Využití DPZ v ochraně přírody .....	21
3.3.1 Klasifikace lesů.....	22
4 Metodika .....	24
4.1 Zájmové území a použitá data.....	24
4.2 Tvorba trénovacích a validačních dat.....	28
4.3 Klasifikace obrazu v Google Earth Engine .....	29
4.4 Výpočet přesnosti klasifikace.....	30
5 Výsledky .....	31
5.1 Přesnost klasifikace při rozlišování lesů a kategorie ostatní .....	31
5.1.1 RF.....	31
5.1.2 CART.....	31
5.2 Přesnost klasifikace lesů na první hierarchické úrovni .....	32
5.2.1 RF.....	32
5.2.2 CART.....	33
5.3 Přesnost klasifikace lesů na druhé hierarchické úrovni.....	34
5.3.1 RF.....	34
5.3.2 CART.....	36



6	Diskuse.....	38
7	Závěr .....	41
8	Seznam literatury .....	43
	Přílohy .....	48

## **Seznam zkratek použitých v textu**

CART – Classification and regression trees

DPZ – Dálkový průzkum Země

EVL – Evropsky významné lokality

GEE – Google Earth Engine

MZCHÚ – Maloplošná zvláště chráněná území

NDVI – Normalizovaný diferenční vegetační index

NIR – Near infrared – blízké infračervené záření

OA – Overall accuracy – celková přesnost

PA – Producer's accuracy – přesnost zpracovatele

RF – Random forest

SWIR – Shortwave infrared – krátkovlnné infračervené záření

UA – User's accuracy – přesnost uživatele

VMB – Vrstva mapování biotopů

# 1 Úvod

Natura 2000 je soustava chráněných území, kterou musí vymezit každý členský stát Evropské unie. Natura 2000 zabezpečuje ochranu druhů živočichů, rostlin a typů přírodních stanovišť, které jsou z evropského pohledu nejcennější, vzácné, ohrožené nebo endemické. Aby tato stanoviště mohla být určena, musely se zmapovat všechny biotopy v České republice (MŽP, ©2008–2020). Díky tomu vznikla Vrstva mapování biotopů, která poskytuje informace o rozložení biotopů v celé ČR. Tyto biotopy jsou podrobně popsány v Katalogu biotopů České republiky (Chytrý et al., 2010). A jelikož se tato vrstva musí udržovat v aktuální podobě, začala aktualizace. Každý rok se znovu zmapuje 1/12 území České republiky. Celá aktualizace trvá 12 let. A právě proto se nabízí dálkový průzkum Země, který by tuto aktualizaci mohl zrychlit.

Dálkový průzkum Země poskytuje informace o současném stavu Země, i o probíhajících změnách. Multispektrální snímky družice Sentinel-2 jsou volně dostupné a jsou cenným nástrojem k mapování krajinného pokryvu, monitoringu vegetace nebo klasifikaci biotopů, protože mají 3 pásma v červeném okraji viditelného spektra. Tyto snímky jsou vhodné i k monitoringu lesů (Žubrietovský et al., 2017). Druhově pestré lesy jsou stále více ohrožovány například odvodňováním nebo vysazováním smrkových monokultur (Chytrý et al., 2010).

V této práci jsem zjišťovala, jaké podrobnosti lesů dané Katalogem biotopů (Chytrý et al., 2010) lze dosáhnout pomocí řízené klasifikace snímků družice Sentinel-2. K tvorbě trénovacích dat jsem použila Vrstvu mapování biotopů, která posléze vstupovala do řízené klasifikace. Ke klasifikaci jsem použila klasifikátory random forest a classification and regression trees.

## 2 Cíle práce

Tato bakalářská práce má za cíl zjistit, jaké úrovně podrobnosti lesů dané Katalogem biotopů České republiky (Chytrý et al., 2010) lze dosáhnout pomocí řízené klasifikace multispektrálních snímků družice Sentinel-2. Klasifikace proběhla na snímcích z roku 2018 ve vybraném zájmovém území. Ke klasifikaci byly využity dva různé algoritmy – random forest a classification and regression trees.

Dílčí cíle:

1. Zjistit, s jakou přesností byl odlišen les od ostatních biotopů.
2. Zhodnotit, s jakou přesností byly odlišeny lesy mezi sebou podle první hierarchické úrovně dané Katalogem biotopů ČR (Chytrý et al., 2010).
3. Zjistit, s jakou přesností byly klasifikovány lesy podle druhé hierarchické úrovně, tedy podle nejpodrobnějšího dělení lesů v Katalogu biotopů ČR (Chytrý et al., 2010).

### 3 Literární rešerše

#### 3.1 Natura 2000 a Vrstva mapování biotopů

Vzhledem k tomu, že na území členských států Evropské unie se stav přírodních stanovišť neustále zhoršuje a stále více druhů ve volné přírodě je vážně ohroženo, je nezbytné přijímat opatření na jejich ochranu na úrovni celého Společenství (Směrnice 92/43/EHS). Tato stanoviště je nutné klasifikovat, aby měla prioritu při včasném provádění opatření na jejich ochranu (Směrnice 92/43/EHS). Na základě toho vznikla síť Natura 2000, jejímž cílem je zabezpečit ochranu živočichů, rostlin a přírodních stanovišť, které jsou z evropského pohledu nějak cenné, ohrožené nebo vzácné. Dalším účelem je zamezit činnostem, které by mohly tato stanoviště vážně poškodit nebo narušit, a podniknout pozitivní kroky v rámci jejich zachování a obnovy, a tím zajistit jejich ochranu (Evropská komise, ©1995-2021)

Natura 2000 se skládá z Evropsky významných lokalit (EVL) a Ptačích oblastí (PO), které musely být vymezeny každým členským státem Evropské unie. Jako podklad k určení těchto oblastí vznikla Vrstva mapování biotopů, která přináší informaci o rozložení biotopů v celé České republice. Všechny biotopy, které jsou zaznamenané ve Vrstvě mapování biotopů, jsou popsány v Katalogu biotopů České republiky (Chytrý et al., 2010).

První mapování, kterého se účastnilo více než 750 mapovatelů, probíhalo od roku 2000 do roku 2005 (Härtel et al., 2009). Mapování proběhlo na dvou úrovních – kontextové mapování a podrobné mapování. Kontextové mapování mělo za cíl zjistit výskyt a rozlohu všech přírodních biotopů v celé ČR a eventuálně odhadnout území pro podrobné mapování. Nemapovala se území, kde proběhlo nebo mělo probíhat podrobné mapování, souvisle zastavěné plochy, zemědělské a lesnické kultury bez přírodních hodnot a území devastované těžbou (Guth, 2002). Výsledkem kontextového mapování byly vymezené a zakreslené přírodní biotopy takřka na celém území ČR s informací o rozloze a stavu.

Cílem podrobného mapování bylo vymežit a zakreslit přírodní biotopy, zjistit údaje o zachovalosti, reprezentativnosti, zpracovat textové a grafické výstupy (Guth, 2002). Mapovatel určil biotop na nejnižší hierarchické úrovni podle Katalogu biotopů (Chytrý et al., 2001), vymežil hranice biotopu, u mozaikovitých segmentů určil

procentuální zastoupení biotopů, doplnil sledované údaje, zapsal výskyt cévnatých rostlin a udělal fotodokumentaci (Härtel et al., 2009).

Vrstva mapování biotopů se musí udržovat v aktuální formě (Směrnice 92/43/EHS). Nutností bylo naplánovat aktualizaci této vrstvy. Aktualizace začala v roce 2006 s tím, že každý rok se zaktualizuje 1/12 území České republiky. Celá ČR je celá znovu zmapována za 12 let (Lustyk, 2016). Obsah aktualizace je dán směrnicí 92/43/EHS, o ochraně přírodních stanovišť, volně žijících živočichů a planě rostoucích rostlin. Mapovatelé ověřují výskyt a stav biotopů a zakreslí nebo pozmění jejich hranice (Härtel et al., 2009). U jednotlivých biotopů zaznamenají příčinu změny biotopu, pokud ke změně došlo, reprezentativnost biotopu, prostorovou a věkovou strukturu stromového a keřového patra, množství mrtvého dřeva, míru antropogenní degradace, management, hodnocení biotopu v regionálním kontextu, hodnocení typických druhů a hodnocení struktury a funkce (Lustyk, 2016).

Podkladem pro aktualizaci je Katalog biotopů (Chytrý et al., 2010), Příručka hodnocení biotopů (Filippov et al., 2008), Metodika aktualizace biotopů (Lustyk, 2016), tištěné ortofoto mapy 1:10 000 se zákresem hranic aktualizčních okrsků, segmentů, MZCHÚ, EVL a I. zónami na území národních parků. Dále jsou třeba tištěné lesnické mapy 1:10 000, výpis z databáze segmentů vrstvy mapování biotopů, závěrečné zprávy z předchozích mapování a v neposlední řadě jsou potřeba formuláře pro terénní zápis.

## **3.2 Vybrané technické aspekty využití DPZ v ochraně přírody**

Terénní pozorování jsou důležitým zdrojem informací pro hodnocení stavu ochrany přírody, jsou ale časově a finančně nákladná. Vzhledem k tomu, že jsou finanční zdroje omezené, je zapotřebí nákladově efektivní přístup využívající nejlepší dostupné metody monitorování (Mücher et al., 2010). Tím může být pozorování pomocí dálkového průzkumu Země, zejména multispektrální a hyperspektrální techniky, protože poskytují konzistentní informace v čase a prostoru a doplňují pozorování v terénu (Valentini et al., 2015).

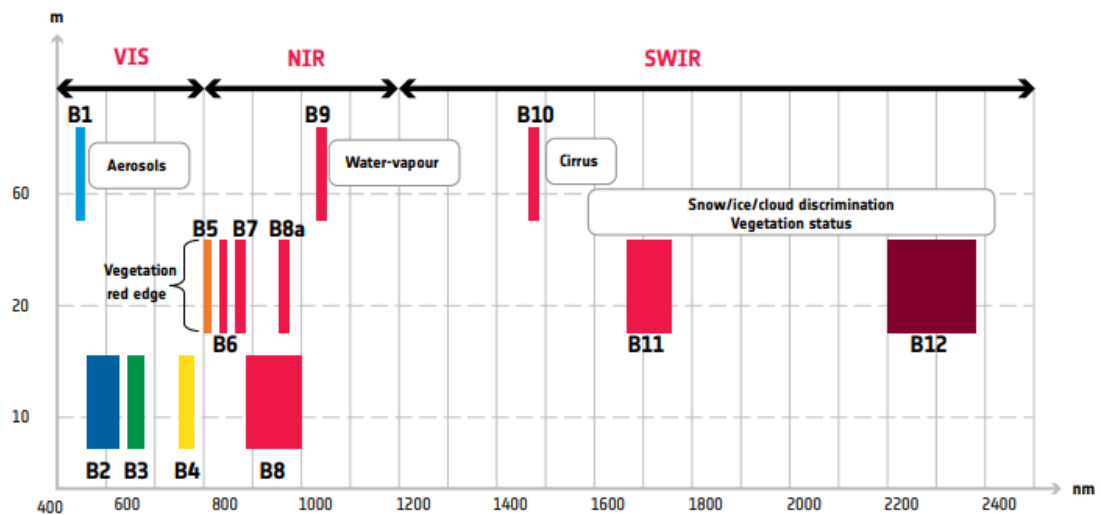
### **3.2.1 Sentinel-2**

Účinným kandidátem pro mapování stanovišť soustavy Natura 2000 by mohly být multispektrální snímky (Valentini et al., 2015), proto byla ke klasifikaci obrazu využita data družice Sentinel-2.

Sentinel-2 je misí programu Copernicus, tvoří ji 2 stejné družice – Sentinel-2A, která startovala v roce 2015 a Sentinel-2B, která byla vypuštěna na orbitu v roce 2017. Samotná družice prolétne nad stejným místem jednou za 10 dní. Ale díky tomu, že tyto družice jsou umístěny s posunem 180°, vznikají snímky jednoho místa jednou za 5 dní na místech vyskytujících se na rovníku. Tato doba se zkracuje směrem od rovníku. Snímky z míst na úrovni Evropy jsou přibližně jednou za 3 dny (Žubrietovský et al., 2017). Tyto družice jsou na heliosynchronní dráze, což znamená, že družice mají stálý úhel mezi rovinou orbity a spojnicí Země – Slunce, z toho plyne, že každé místo je snímáno vždy za stejných podmínek osvětlení (Kropáček et al., 2020).

Sentinel-2 poskytuje multispektrální snímky. Nese optický senzor MSI snímající v 13 pásmech v rozsahu 443–2190 nanometrů. Senzor MSI používá pushbroom scanner a je první, který má 3 pásma v červeném okraji viditelného spektra, tzv. red edge, tudíž je ideální k mapování krajinného pokryvu a jeho změn, k monitoringu vegetace, klasifikaci lesů a k sledování zdravotního stavu vegetace (Žubrietovský et al., 2017). Pushbroom scanner funguje tak, že sbírá řady obrazových dat napříč orbitálním řádkem a využívá pohyb družice dopředu (ESA, ©2000–2021). Šířka snímaného záběru družicí je 290 km (Žubrietovský et al., 2017).

Prostorové rozlišení závisí na konkrétním spektrálním pásmu a je znázorněno na obrázku 1. Čtyři kanály jsou v prostorovém rozlišení 10 m, konkrétně: 490 nm (B2), 560 nm (B3), 665 nm (B4), 842 nm (B8). Dalších šest kanálů je v prostorovém rozlišení 20 m: 705 nm (B5), 740 nm (B6), 783 nm (B7), 865 nm (B8a), 1610 nm (B11) a 2190 nm (B12). A disponuje třemi kanály v prostorovém rozlišení 60 m: 443 nm (B1), 945 nm (B9) a 1375 nm (B10) (ESA, ©2000–2021).



Obrázek 1: Prostorové rozlišení spektrálních pásem družice Sentinel-2 a jejich vlnové délky (ESA, ©2000–2021).

Data jsou volně dostupná ve formě tzv. dlaždic po jednoduché registraci na <https://scihub.copernicus.eu/>, a to ve dvou úrovních zpracování dat – level-1C (L1C) a level-2A (L2A). Zpracování úrovně L1C zahrnuje radiometrické a geometrické korekce včetně ortorektifikace v projekci WGS84. Dále jsou vypočteny hodnoty TOA (Top-Of-Atmosphere) (ESA, ©2000–2021), to je odrazivost z horní vrstvy atmosféry (Kropáček et al., 2020).

Snímky úrovně zpracování L2A obsahují navíc atmosférické korekce (Žubrietovský et al., 2017). Poskytují hodnoty odrazivosti BOA (Bottom-Of-Atmosphere), které jsou odvozené z přidružených produktů úrovně zpracování L1C (ESA, ©2000–2021) a udávají spektrální hodnoty povrchu, tzv. hodnoty pod atmosférou. Obě úrovně jsou poskytovány ve formě dlaždic s rozměrem 100 x 100 km (Žubrietovský et al., 2017).

### 3.2.2 Klasifikace

Ke klasifikaci byla využita data družice Sentinel-2. Klasifikace obrazu je proces, který přiřazuje určitou informaci jednotlivým prvkům. Na základě rozhodovacích pravidel, která jsou pojmenovaná jako klasifikátory, lze prvky v obraze umístit do jednotlivých kategorií. Jedny z nejvíce využívaných klasifikátorů jsou ty, které jsou založené na spektrálním projevu objektů, protože různé objekty mají jiné spektrální chování. Jsou označovány jako bodové neboli per-pixel klasifikátory. Klasifikaci dělíme na řízenou a neřízenou (Dobrovolný, 1998).



### **3.2.2.1 Neřízená klasifikace a její algoritmy**

U neřízené klasifikace je obraz nejdříve rozdělen do různých kategorií tzv. spektrálních tříd, zpracovatel jim přidává informaci až po provedení klasifikace a stanou se z nich tedy tzv. třídy informační (Dobrovolný, 1998). Před spuštěním neřízené klasifikace je nutné jen určit klasifikátor a počet různých kategorií, do kterých se má obraz podle spektrálního chování objektů roztrdit.

#### *K-means*

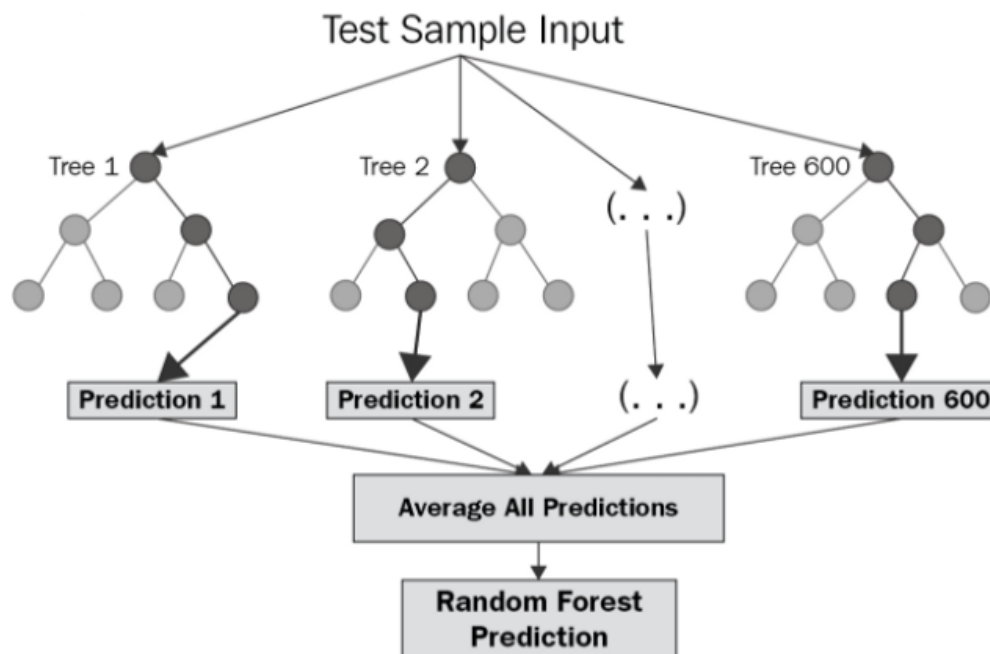
K-means je nejjednodušší shlukovací algoritmus, u kterého se na začátku musí určit finální počet shluků. Někdy je možné určit i počáteční polohu jejich středů, nebo si algoritmus tuto polohu určí sám. Dále je každý pixel přiřazen do shluku, ke kterému má nejbližší. Z klasifikace jsou pak přepočítány středy shluků a celý proces se opakuje tolikrát, kolik bylo určeno iterací na začátku klasifikace, nebo se ukončí dříve, pokud se pixely nijak výrazně nemění (Dobrovolný, 1998).

### **3.2.2.2 Řízená klasifikace a její algoritmy**

Naopak u řízené klasifikace se musí předem určit tzv. trénovací plochy. To jsou místa, o kterých si je zpracovatel jistý, do jakých tříd spadají. Následně jsou pro každou třídu vytvořeny statistické spektrální charakteristiky, s kterými je porovnáván každý pixel. A ten je na základě předem zvoleného klasifikátoru určen do správné třídy (Dobrovolný, 1998). Jeden z mnoha klasifikátorů je například algoritmus random forest.

#### *Random forest*

Random forest (RF) je algoritmus strojového učení (machine learning algorithm), vyvinutý Breimanem (Breiman, 2001), který se skládá z dílčích rozhodovacích stromů tzv. decision trees. RF využívá iterativní pytlování, tzv. bagging (Chemura et al., 2018). Jednotlivé stromy obsahují uzly, ve kterých se strom větví. Každý uzel obsahuje otázku na základě, které jsou data rozdělena (Klashka & Kotrč, 2004). Na začátku klasifikace je určen počet rozhodovacích stromů a každý jednotlivý strom předpoví třídu. Jak je zobrazeno na obrázku 2, předpovězené třídy jednotlivých stromů se zprůměrují, a ta s třída s nejvyšším počtem hlasů se stane předpovědí modelu (Yiu, 2019).



Obrázek 2: Každý rozhodovací strom něco předpoví, a na základě těchto výsledků vznikne finální předpověď' (Chaya Bakshi, 2020).

### *Classification and regression trees*

Classification and regression trees (CART) začíná hledáním nejlepšího dělení pro každý prediktor. V každém uzlu se najde nejlepší dělení a vzniknou tak další dva uzly. Dále se vybere nejlepší dělení daného uzlu pro jednotlivé pozorování – to, které nejvíce snižuje nepřesnost stromu. Pak se pozorování z uzlu rozdělí do dalších dvou, dle předchozího kroku. Tyto kroky se opakují, dokud uzel není označen jako finální. (Masaila et al., 2013). Vzniklý velký strom se na rozdíl od algoritmu random forest ještě „prořezává“, to znamená, že se některé uzly zpětně odstraňují. Při konstrukci se uzel stane konečným, když obsahuje méně pozorování, než zvolená mez nebo všechna pozorování v uzlu patří do jedné třídy (Klashka & Kotrč, 2004).

### *Chybová matice*

Po provedení klasifikace obrazu je také nutné zhodnotit její přesnost. Jedním z nástrojů k hodnocení přesnosti klasifikace je chybová matice, která je v angličtině nazývána *confusion matrix* nebo *error matrix*. Vstupují do ní data, která nejsou shodná s trénovacími daty. Chybová matice je čtvercová a porovnává data odpovídající skutečnosti s výsledkem klasifikace. Počet řádků a sloupců matice je stejný jako počet klasifikovaných tříd (Dobrovolný, 1998). Matice udává informaci, jak často je určitá třída detekována správně, a jak často je určena za jinou klasifikační třídu (Ruuska et

al., 2018). Sloupce matice vyjadřují hodnoty získané klasifikací, řádky vyjadřují hodnoty odpovídající skutečnosti. Na hlavní diagonále je tedy počet správně zařazených pixelů do dané třídy, a mimo hlavní diagonálu je počet chybně zařazených pixelů.

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negatives (TN)	False Positives (FP) Type I error
	Positive +	False Negatives (FN) Type II error	True Positives (TP)

Obrázek 3: Ukázka chybové matice, kde se rozlišují dvě třídy – negative a positive. „True negatives“ udává počet správně zařazených pixelů spadajících do třídy negative a „true positives“ udává počet správně klasifikovaných pixelů spadajících do třídy positive. Hodnota „false positives“ udává počet pixelů, který měl být klasifikován jako negative, ale byl špatně zařazen do třídy positive. Hodnota „false negatives“ udává počet pixelů, který byl špatně klasifikován jako negative, ale patří do třídy positive (NoteBook Share, ©2019-2020).

Z údajů chybové matice se pak může například vypočítat přesnost zpracovatele (producer's accuracy), přesnost uživatele (user's accuracy) a celková přesnost klasifikace (overall accuracy) (Barsi et al., 2018; Dobrovolný, 1998).

Přesnost zpracovatele udává, kolik procent pixelů, které měly být zařazeny do určité třídy, byly do této třídy skutečně zařazeny. Vyjadřuje tedy pravděpodobnost, že určitá třída je klasifikována správně. Udává poměr mezi počtem správně klasifikovaných pixelů (hodnota na hlavní diagonále) a počtem pixelů ve všech třídách uvažovaného řádku (tedy suma řádku) (Barsi et al., 2018; Dobrovolný, 1998). Dle obrázku 3 se přesnost zpracovatele pro třídu označenou jako negative vypočítá jako  $TN/(TN + FP)$ . Pro třídu označenou jako positive se vypočítá jako  $TP/(TP + FN)$ .

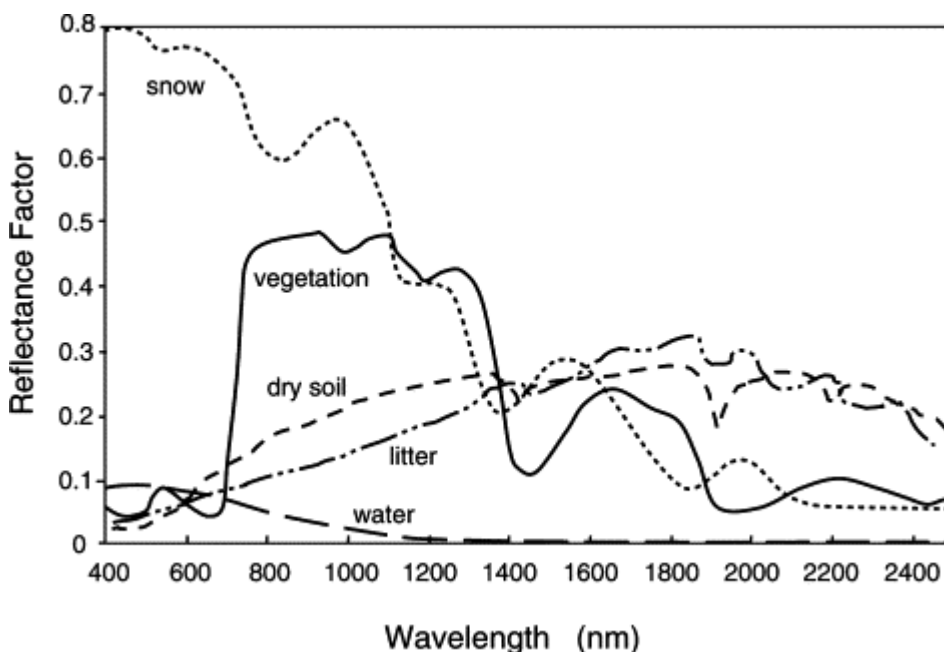
Přesnost uživatele udává, jaká je pravděpodobnost, že pixel zařazený do dané třídy, tuto třídu skutečně reprezentuje (Dobrovolný, 1998). Vyjadřuje poměr mezi počtem správně klasifikovaných pixelů do dané třídy (zase hodnota na hlavní

diagonále) a celkový počet pixelů, které byly klasifikovány jako určená třída (suma sloupce). Podle obrázku 3 se přesnost uživatele pro třídu označenou jako negative vypočítá jako  $TN/(TN + FN)$  a pro třídu označenou jako positive jako  $TP/(TP + FP)$ .

Celková přesnost klasifikace udává poměr všech správně klasifikovaných pixelů (suma hodnot na hlavní diagonále) a celková suma klasifikovaných pixelů (Červená, 2018). Podle obrázku 3 se vypočítá jako  $(TN + TP)/(TN + FN + FP + TP)$ .

### 3.2.3 Spektrální chování objektů

Nejčastěji využívané klasifikátory jsou založené na spektrálním projevu objektů, protože se povrchy objektů chovají různě. Sluneční záření dopadající na zemský povrch je zčásti pohlceno a zčásti odrazeno zpátky do atmosféry. A na zemi, když se záření odráží, podléhá určitým přeměnám. Množství odraženého záření můžeme označovat jako spektrální odrazivost, kterou lze určit jako poměr odraženého záření a intenzity záření dopadajícího na určité vlnové délce (Dobrovolný, 1998). Spektrální odrazivost je ovlivněna druhem látky či objektu, jeho fyzikálním stavem nebo stavem jeho okolí (Dobrovolný, 1998). Jak je uvedeno na obrázku 4, odrazivost objektů se mění v různých vlnových délkách, díky tomu lze rozlišit, zda se například jedná o vegetaci, holou půdu nebo vodu.



Obrázek 4: Spektrální odrazivost různých povrchů (Heute, 2004).

Ve viditelných pásmech je odrazivost zdravé vegetace relativně nízká, protože většina světla je absorbována listovými pigmenty. Chlorofyl silně absorbuje energii v modré a červené vlnové délce, a naopak v zelené části spektra se odráží. Proto se zdravá vegetace jeví zeleně (Dobrovolný, 1998; Humboldt State University, ©2014).

V blízké infračervené části spektra (NIR) se spektrální chování zdravé vegetace projevuje výrazným nárůstem odrazivosti, a to díky buněčné struktuře listů, konkrétně houbovitému mezofylu. Zdravou vegetaci lze proto snadno identifikovat podle vysoké odrazivosti v NIR a nízké odrazivosti ve viditelném záření (Humboldt State University, ©2014). A právě rozdílné odrazivosti vegetace ve viditelné červené části spektra a NIR se využívají k výpočtu některých vegetačních indexů (Dobrovolný, 1998), jako je například normalizovaný diferenční vegetační index (NDVI). Ten vyjadřuje hodnotu fotosyntetické aktivity, která je podmíněna přítomností chlorofylu. Index NDVI nabývá hodnot od -1 do 1. Hodnoty okolo 0,5 udávají velmi hustou vegetaci (Spadoni et al., 2020).

Odrazivost v krátkovlnné infračervené (SWIR) části spektra souvisí s obsahem vody ve vegetaci a její strukturou. Voda má silné absorpční pásy kolem 1,45, 1,95 a 2,50 mikrometrů (Dobrovolný, 1998). Mimo tyto absorpční pásy v oblasti SWIR se odrazivost listů zvyšuje, když se obsah vody v listu snižuje (Humboldt State University, ©2014).

Na základě spektrálního chování lze například rozlišit les listnatý od lesa jehličnatého (Immitzer et al., 2016). Listnaté stromy mají totiž vyšší odrazivost v blízké infračervené oblasti ve srovnání s jehličnany (Humboldt State University, ©2014). Spektrální odrazivost je ovlivněna buněčnou strukturou, listovými pigmenty a obsahem vody v listech. Na základě toho lze pak například identifikovat rostliny, které jsou stresované nebo nemocné (Cârlan et al., 2020).

### **3.3 Využití DPZ v ochraně přírody**

V současné době je k dispozici široká škála dat z dálkového průzkumu Země, a to z jak komerčních, tak i nekomerčních zdrojů. Jedním z nekomerčních zdrojů jsou data družice Sentinel-2 programu mise Copernicus (Žubrietovský et al., 2017). Satelitní mapování poskytuje objektivní, aktuální informace jak o současném stavu Země, tak o probíhajících změnách (Weigand et al., 2020). Data DPZ se dají využít v mnoha oblastech, jako je například rozlišení tříd krajinného pokryvu, druhů lesních

porostů, odrůd plodin, detekci nepůvodních a invazních druhů, monitorování sucha a nebezpečí požáru, odhad biomasy a uhlíku (Adamo et al., 2016), identifikace klimatických změn a mnoho dalších (Baamonde et al., 2019). Například při zvažování chráněných stanovišť jsou vyžadovány právě informace o jejich kvalitě a změnách v průběhu času, tyto informace mohou přispět k rozvoji, hodnocení strategií a politik ochrany (Tarantino et al., 2016).

Monitorování sucha a nebezpečí požáru, odhad biomasy a uhlíku může být zjištěno pomocí vegetačních indexů. Jedním z těchto indexů je NDVI, na základě jeho anomálií v průběhu času se mohou například pozorovat klimatické výkyvy. Pomocí klasifikace obrazu mohou být rozlišovány třídy krajinného pokryvu. Dalším využitím klasifikace může být například detekce rozsahu povodní.

### **3.3.1 Klasifikace lesů**

Data z mise Sentinel-2 umožňují odvozovat přesné a včasné mapy dřevin v lesích, a protože mají 3 pásma v červeném okraji spektra a mimořádně úzké blízké infračervené pásmo (NIR), jsou vhodné k identifikaci vegetačních společenstev (Adamo et al., 2016). Mapování druhů stromů poskytuje podrobný popis lesních ekosystémů. To je nezbytné pro jejich ochranu a udržitelné obhospodařování lesů.

Lesy byly klasifikovány v rámci mnoha studií. Jeden z prvních výzkumů s využitím družice Sentinel-2 (Immitzer et al., 2016) se mimo jiné zabýval i klasifikací druhů stromů. Byl rozlišován smrk, borovice, modřín, jedle, buk, dub a ostatní listnaté stromy. Zájmové území se nachází v Bavorsku. Za použití klasifikátoru random forest bylo dosaženo 63,5% celkové přesnosti klasifikace. U smrku hodnoty přesností dosahovaly 85 % u přesnosti zpracovatele a 73 % u přesnosti uživatele. Borovice však byla klasifikována s přesností mezi 20 a 35 % s tím, že ale většina byla špatně klasifikována jako smrk, protože mají podobné spektrální vlastnosti – jejich křivky spektrální odrazivosti se téměř překrývaly napříč všemi spektrálními kanály. Listnaté druhy stromů měly přesnost mezi 40 a 55 %.

Další studie, která se věnovala klasifikaci druhů stromů a využila k tomu snímky družice Sentinel-2, rozlišovala dub, borovici, modřín, habr, jedli, douglasku, břízu, buk, platan, smrk a olši v Polských Karpatech (Grabska et al., 2020). Nejlepší přesnosti klasifikace bylo dosaženo pomocí algoritmu SVM (support-vector machine) a dalších

proměnných, jako je například digitální model terénu. Klasifikace nakonec dosáhla celkové přesnosti 85 %.

Také Persson et al. (2018) rozlišovali ve Švédsku jednotlivé druhy stromů mezi sebou pomocí řízené klasifikace za použití algoritmu random forest. Byly od sebe rozlišeny smrky, borovice, modřín, břízy a duby s celkovou přesností necelých 90 %.

## 4 Metodika

### 4.1 Zájmové území a použítá data

Zvolené zájmové území se nachází přibližně mezi městy Nymburk, Hradec Králové, Heřmanův Městec a Suchdol u Kutné Hory. Přibližně ve středu zájmového území se setkávají hranice třech krajů, a to Královehradeckého, Pardubického a Středočeského. Na obrázku 5 je vybrané území zvýrazněno v rámci celé České republiky. Toto území bylo vybráno tak, aby se rozkládalo v jedné dlaždici družice Sentinel-2. Rozloha území činí 1124 km<sup>2</sup> s tím, že rozsah nadmořských výšek se pohybuje přibližně mezi 175 a 475 m n. m.



Obrázek 5: Zvýrazněné zájmové území na podkladu ortofota. (ČÚZK)

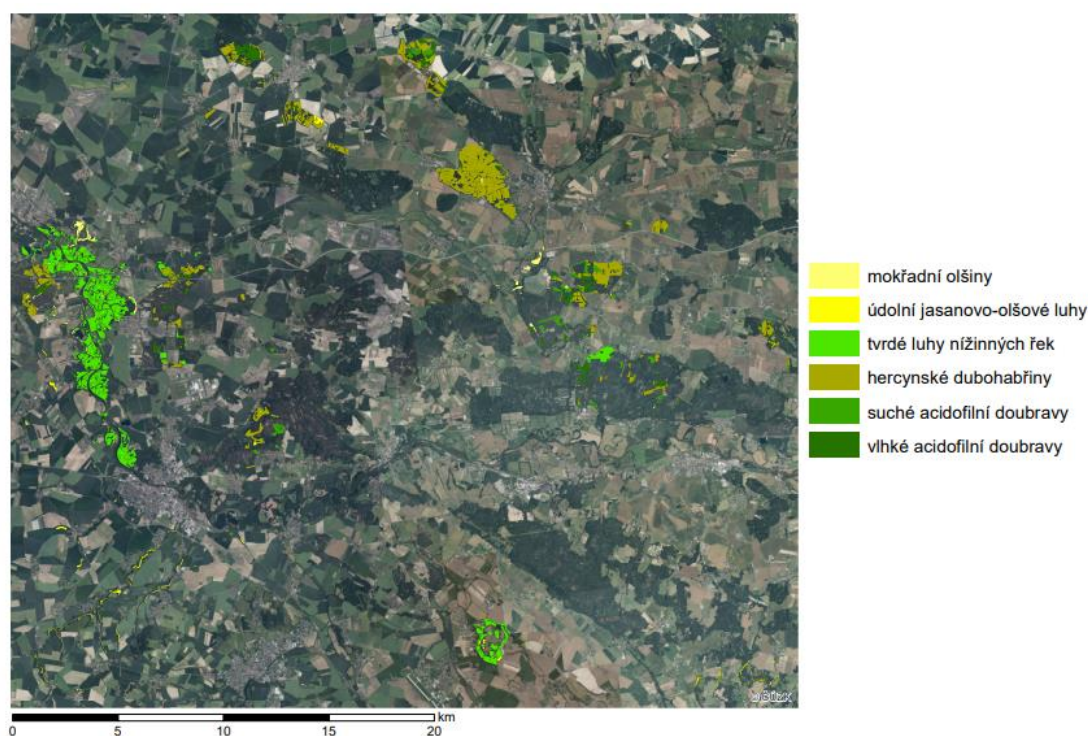
Dále byla použita vrstva mapování biotopů, která byla dostupná po přihlášení na webu poskytování dat od AOPK ČR, konkrétně na <https://data.nature.cz>. Z vrstvy mapování biotopů jsem vybrala jen ty lesy, které byly aktualizovány od roku 2015 včetně. A to kvůli tomu, že se družice Sentinel-2A dostala na oběžnou dráhu až v roce 2015. Snímky jsou tedy dostupné až od tohoto roku. Vyfiltrovaly se tedy ty třídy lesů, které byly aktualizovány v letech 2015, 2016, 2017 a 2018. Z těchto lesů jsem dále vyfiltrovala jen ty, které měly rozlohu nad 1 km<sup>2</sup>. Nachází se zde tedy nakonec celkem 38,1837 km<sup>2</sup> lesů, které odpovídají kritériím. Konkrétně se jedná o třídy lesů L1 (mokřadní olšiny), L2.2 (údolní jasanovo-olšové luhy), L2.3 (tvrdé luhy nížinných řek), L3.1 (hercynské dubohabřiny), L7.1 (suché acidofilní doubravy) a L7.2 (vlhké



acidofilní doubravy). Rozložení vybraných lesů v zájmovém území je zobrazeno na obrázku 6 a jejich rozloha v km<sup>2</sup> je zobrazena v tabulce 1.

Kód biotopu	Název biotopu	Rozloha v km <sup>2</sup>	Počet polygonů
L1	Mokřadní olšiny	1,14226	101
L2.2	Údolní jasanovo-olšové luhy	1,107996	86
L2.3	Tvrdé luhy nížinných řek	11,31776	398
L3.1	Hercynské dubohabřiny	19,077514	865
L7.1	Suché acidofilní doubravy	2,175187	105
L7.2	Vlhké acidofilní doubravy	2,999549	131

Tabulka 1: Rozlohy jednotlivých tříd lesů aktualizovaných od roku 2015 v zájmovém území, které byly vybrány ke klasifikaci.



Obrázek 6: Rozložení lesů, které by aktualizovaný v letech 2015–2018, v zájmovém území (zdroj podkladových dat: AOPK ČR, ČÚZK).

*Mokřadní olšiny* se rozkládají převážně mezi 150 a 400 m n. m. v zamokřených plošinách, v širokých nivách řek, na okrajích rybníků a na lesních močálech. Dominantním stromem je olše lepkavá, místy se může vyskytovat bříza pýřitá. Mokřadní olšiny jsou v ohrožení kvůli odvodňování a následné výsadbě smrku (Chytrý et al., 2010).

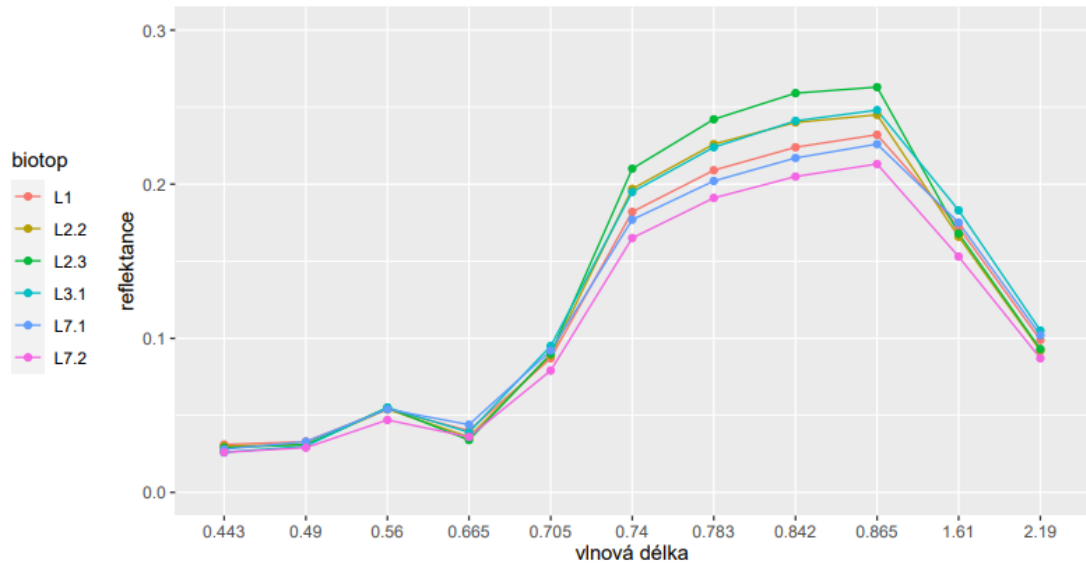
V *lužních lesích* se vyskytují stromy, které snášejí dočasné zamokření půdy. Jsou tedy tvořeny zejména olšemi, jasanem, jilmem, dubem, vrbami a topoly. Podkategorií lužních lesů jsou údolní jasanovo-olšové luhy a tvrdé luhy nížinných řek (Chytrý et al., 2010), které se vyskytují v zájmovém území. *Údolní jasanovo-olšové luhy* se vyskytují podél vodních toků takřka na území celé České republiky. Dominantním stromem je opět olše lepkavá nebo jasan ztepilý s příměsí dalších listnatých stromů. Ohrožení těchto luhů opět plyne ze změn vodního režimu, mýcení porostů nebo výsadby monokultur (Chytrý et al., 2010). Další podtřídou lužních lesů jsou *tvrdé luhy nížinných řek*. Ty se rozkládají v dolních částech toků, jako je například dolní Povltaví, dolní Poohří, dolní Dyje nebo niva Labe od Jaroměře po Litoměřice. Dub letní, jasan ztepilý nebo jilm jsou dominantními stromy. Tvrdé luhy jsou také ohroženy narušováním vodního režimu krajiny, jako jsou např. regulace toků a odvodňování pozemků (Chytrý et al., 2010).

*Hercynské dubohabřiny* se v zájmovém území vyskytují nejhojněji, ale jsou ohroženy převodem na jehličnaté kultury a spontánní sukcesí, při které se porost zahušťuje, to vede k ústupu vzácných a ohrožených druhů světlomilných rostlin. Dominuje zde habr obecný, dub zimní, dub letní a často je přimísena lípa srdčitá (Chytrý et al., 2010).

Další a poslední skupinou, která se nachází v zájmovém území, jsou acidofilní doubravy. To jsou druhově chudé lesy s dominujícími duby, k nimž se občas přimíchávají druhy bříz, dále pak jedle nebo borovice. Podtřídou acidofilních doubrav jsou suché a vlhké acidofilní doubravy. *Suché acidofilní doubravy* se hojně vyskytují ve středních, západních a severních Čechách. Dominantním stromem je dub zimní, méně často pak dub letní. Místy je přimísena bříza bělokorá a borovice lesní. Naopak *vlhké acidofilní doubravy* mají jako dominantní strom dub letní, méně často dub zimní. Dále je přimísena například bříza bělokorá, borovice lesní nebo jedle bělokorá. Tyto porosty jsou ohroženy zejména kvůli tomu, že nejsou obhospodařovány stejně jako v minulosti. Tím je myšlena například lesní pastva a hrabání steliva. Kvůli tomu se zde šíří habr, buk, bříza či olše a na místech, kde došlo k odvodnění, se rozšířily ostružiníky a vysoké trávy. Dalším problémem je, že jsou duby nahrazovány smrky (Chytrý et al., 2010).

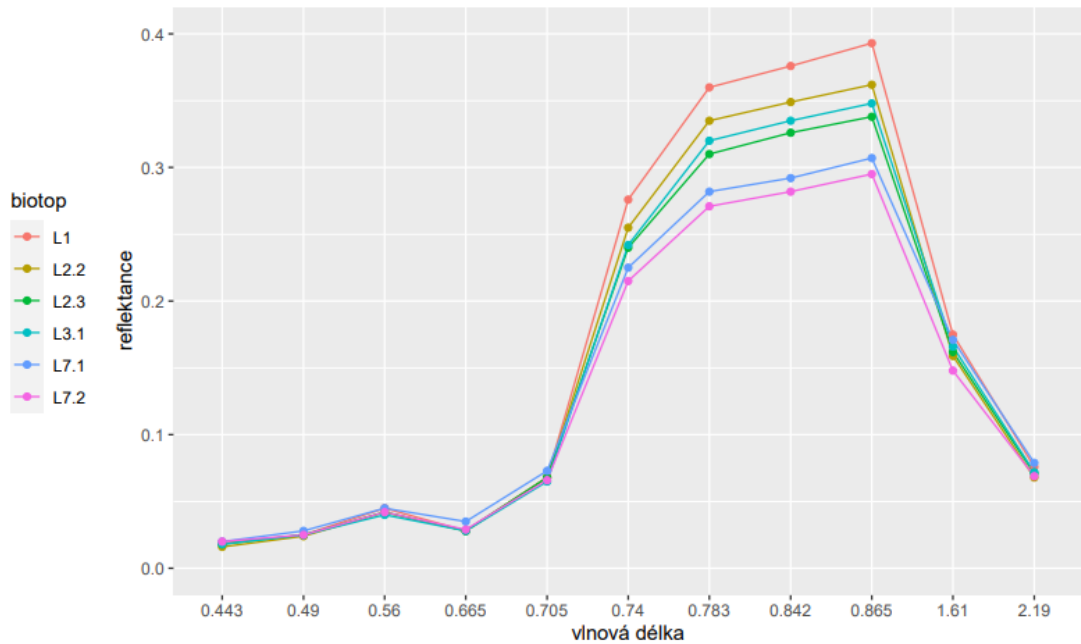
Jako multispektrální snímky byly využity snímky družice Sentinel-2 úrovně zpracování L2A. Vybrala jsem snímky z dubna 2018 a července 2018, protože právě v těchto měsících měly třídy lesů alespoň trochu rozdílné křivky odrazivosti. Tyto křivky pro vybrané měsíce jsou zobrazeny v grafu 1 a 2.

### Spektrální charakteristika duben



Graf 1: Spektrální odrazivost vybraných tříd lesů v dubnu. Ve viditelné části spektra se křivky odrazivosti téměř překrývají. V pásmech „red edge“ a blízkém infračerveném záření se překrývají jen biotopy L2.2 (údolní jasanovo-olšové luhy) a L3.1 (hercynské dubohabřiny).

### Spektrální charakteristika červenec

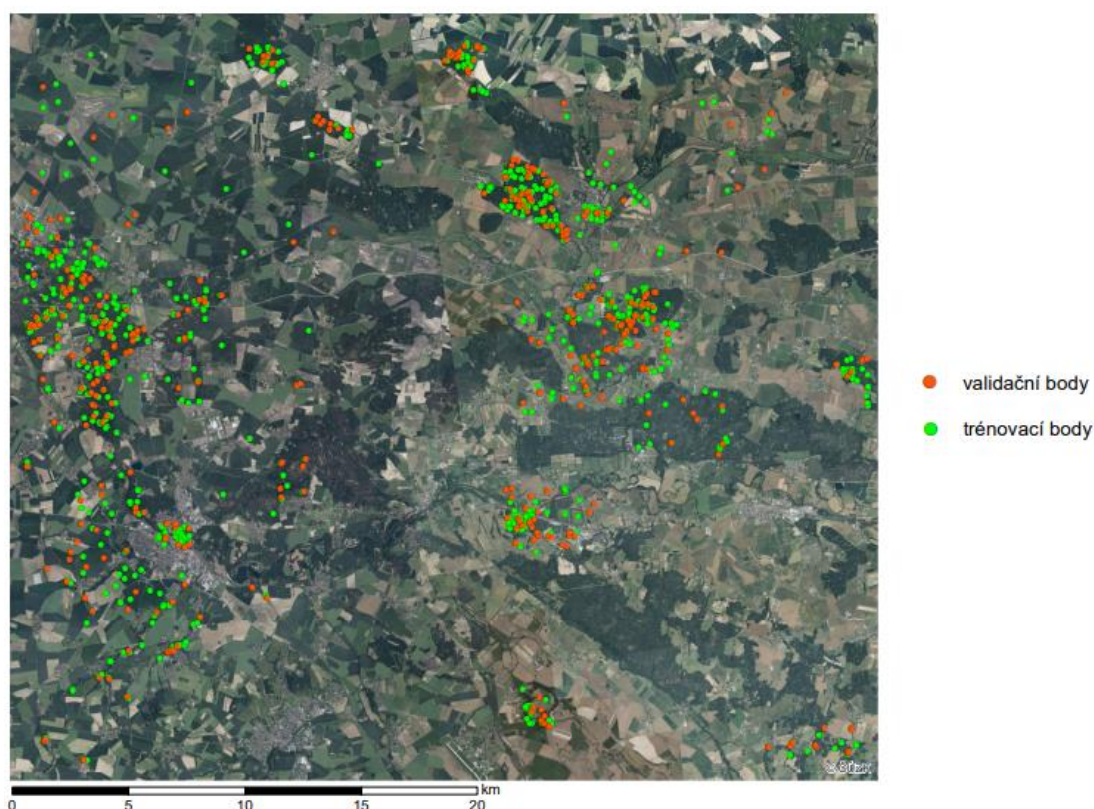


Graf 2: Spektrální odrazivost vybraných tříd lesů v červenci. Ve viditelné části spektra jsou křivky odrazivosti takřka úplně překryté, naopak v pásmech „red edge“ a blízkého infračerveného záření se křivky nepřekrývají.

## 4.2 Tvorba trénovacích a validačních dat

Protože okrajové pixely většinou udávají smíšenou spektrální informaci a mohly by pak negativně ovlivnit výsledek klasifikace (Dobrovolný, 1998), odstranila jsem pomocí funkce *Buffer* v softwaru ArcMap 10 metrů okrajů vyfiltrovaných tříd. Do vybraných tříd jsem vytvořila náhodné body tak, aby v každé třídě byl stejný počet bodů. Do každé třídy jsem tedy vygenerovala po 150 bodech. Trénovací body do třídy ostatní jsem vytvořila na podkladu RGB snímku s tím, že umístění bodů jsem vybírala tak, aby byly pokryty všechny jiné možnosti, než je les. Nakonec do třídy ostatní připadlo 304 různých bodů.

Tyto body jsem dále rozdělila tak, aby dvě třetiny bodů byly trénovací a jedna třetina validační. Třídy lesů se tedy rozdělily po 100 trénovacích bodech a 50 validačních. U kategorie ostatní připadlo do trénovacích dat 200 bodů a do validačních 104 bodů. K tomuto rozdělení jsem použila python skript dostupný na webových stránkách (Support ESRI, 2020), který náhodně vybere určitý počet bodů z vybrané třídy. Rozdělení bodů na trénovací a validační je znázorněno na obrázku 7.



Obrázek 7: Rozdělení bodů na trénovací a validační body (zdroj podkladových dat: AOPK ČR, ČÚZK).

### 4.3 Klasifikace obrazu v Google Earth Engine

Google Earth Engine (GEE) je cloudová platforma, která umožňuje uživatelům spouštět geoprostorové analýzy v planetárním měřítku na infrastruktuře Google pro řešení mnoha problémů, jako je například monitoring odlesňování, sucha, katastrof, vodního hospodářství, klimatu a ochrany životního prostředí. Existuje několik způsobů interakce s platformou. Jedním z nich je code editor, který je prostředím pro psaní a spuštění skriptů (Gorelick et al., 2017).

Ke klasifikaci v GEE jsem jako podklad použila publikovaný skript v rámci studie, ve které byla provedena řízená klasifikace pomocí klasifikátoru random forest (Nomura & Mitchard, 2018). V rámci klasifikace byla rozlišována palma olejná, kaučukovník, areka obecná, les, holá půda, voda a ostatní nelesní biotopy. Klasifikace byla provedena 1000krát s různými trénovacími a validačními body. Medián celkové přesnosti klasifikace se pohyboval okolo 95 % (Nomura & Mitchard, 2018). Skript jsem upravila podle mých potřeb. Výsledný skript je uveden v příloze.

GEE umožňuje uživatelům udělat ze všech shromážděných snímků kompozit na základě hodnoty mediánu pro každý pixel (Nomura & Mitchard, 2018). Do klasifikace tedy vstupují snímky z dvou období z roku 2018, ze kterých jsem vytvořila kompozity. První kompozit je z období od 15. dubna do 30. dubna a druhý kompozit je v rozmezí od 1. července do 15. července. Snímky, ze kterých jsem vytvořila kompozity, měly maximálně 5 % mraků. Tyto kompozity jsem ořízla podle hranic zájmového území, které jsem mimo jiné nahrála dříve do GEE. Mezi dalšími vrstvami, které jsem nahrála do GEE, byly trénovací body a hranice vybraného zájmového území.

Dále do klasifikace vstupovaly jen ta pásma, která jsem vybrala na základě spektrální odrazivosti. Z kompozitu z dubnových snímků vstupovala do klasifikace pásma B5, B6, B7, B8, B8A a B12 a z kompozitu z července vstupovala pásma B5, B6, B7, B8 a B8A. K těmto pásmům jsem přidala ještě hodnoty NDVI, kde NIR je pásmo B8 a R je pásmo B4. Normalizovaný diferenční vegetační index byl vypočítán podle vzorce:  $NDVI = (NIR - RED)/(NIR + RED)$ .

Když jsem měla připravené kompozity a vstupní pásma do klasifikace, přepsala jsem v atributové tabulce trénovacích a validačních dat tak, aby biotopy byly uloženy jako hodnoty od 0. Tento krok jsem udělala, protože biotopy v atributové tabulce byly uloženy jako kód biotopu, například jako L1 a L2.2.

Následně jsem použila algoritmus random forest, ten je stále populárnější hlavně díky tomu, že zlepšuje přesnost klasifikace oproti jiným jednodušším metodám. Do tohoto klasifikátoru jsem nastavila počet stromů na 100, protože čím víc stromů, tím víc se snižuje míra chyb. Jako další klasifikátor jsem vybrala CART, u kterého bylo ponecháno defaultní nastavení. Od algoritmu random forest se liší tím, že se na konci algoritmu ještě „prořezává“ (Masaila et al., 2013). Konečný klasifikovaný snímek jsem vyexportovala, abych ho mohla využít k dalším analýzám, jako je například výpočet chybové matice v softwaru ArcMap.

#### 4.4 Výpočet přesnosti klasifikace

Jelikož žádná klasifikace není označena za ukončenou bez určení přesnosti (Dobrovolný, 1998), byla vypočtena chybová matice a následně celková přesnost klasifikace, přesnost zpracovatele a uživatele.

Nejdříve jsem zjišťovala, jak přesně klasifikátor rozlišil všechny lesy od třídy ostatní. K tomu byly použity validační body, které jsem dříve vytvořila. Hodnoty z rastru jsem přepsala do bodů pomocí funkce *Extract values to points* v softwaru ArcMap. Díky tomu byla v atributové tabulce každého bodu skutečná třída lesa a třída lesa, která byla přiřazena klasifikátorem. Následně jsem vytvořila další tabulku pomocí funkce *Frequency*, která obsahovala hodnoty udávající, kolik pixelů představující určitou třídu, bylo zařazeno, do jaké třídy. Tyto hodnoty představovaly údaje, které jsou pak obsaženy v chybové matici, akorát nejsou správně seřazené. Tato tabulka vstupovala do funkce *Pivot table*, která data seřadila do jiné a konečné tabulky jako chybovou matici.

Abych určila přesnost klasifikace, jak byly rozděleny lesy mezi sebou podle první a druhé hierarchické úrovně, zjišťovala jsem, jak odpovídá klasifikovaný snímek vybraným lesům VMB. Do funkce *Tabulate area* v softwaru ArcMap jsem nahrála klasifikovaný snímek a polygonovou vrstvu lesů VMB aktualizovanou mezi lety 2015 a 2018, kterou jsem dříve vytvořila pro tvorbu trénovacích a validačních dat. Tímto krokem jsem získala tabulku, která ukazovala, kolik pixelů bylo zařazeno správně dle VMB a kolik naopak špatně a do jaké třídy. Vznikla tedy chybová matice, na základě které jsem pak dopočítala celkovou přesnost klasifikace, přesnost zpracovatele a přesnost uživatele.



## 5 Výsledky

### 5.1 Přesnost klasifikace při rozlišování lesů a kategorie ostatní

Nejdříve byla zhodnocena přesnost, jak dokázaly oba algoritmy rozdělit data do třídy ostatní a do třídy les. Použití dat z družic Sentinel-2 ke klasifikaci dokázalo u obou algoritmů rozlišit všechny lesy od třídy ostatní takřka dokonale. Do této chybové matice a zjišťování přesnosti vstupovaly validační body. Celková přesnost klasifikace dosáhla u algoritmu random forest 97,8 % a u algoritmu CART 97,3 %.

#### 5.1.1 RF

Pomocí algoritmu random forest bylo dosaženo celkové přesnosti 97,8 %. Přesnost zpracovatele dosáhla 97,7 % u třídy les a 98,1 % u třídy ostatní. Přesnost uživatele měla hodnoty 99,3 % pro les a 93,5 % pro třídu ostatní. Celkem bylo špatně zařazeno jenom 9 pixelů ze 404. Tyto údaje jsou zaznamenány v tabulce 2.

RANDOM FOREST		klasifikovaná data			
		les	ostatní	∑ řádku	PA
skutečnost	les	293	7	300	97,7%
	ostatní	2	102	104	98,1%
	∑ sloupce	295	109	404	
	UA	99,3%	93,6%		
	OA	97,8%			

Tabulka 2: Klasifikovaná data pomocí algoritmu random forest dosáhla celkové přesnosti 97,8 %. Ze 404 validačních bodů špatně zařazeno jenom 9 bodů.

#### 5.1.2 CART

Jak je uvedeno v tabulce 3, s klasifikátorem CART bylo dosaženo celkové přesnosti klasifikace 97,3 %. Přesnost zpracovatele u třídy les dosáhla 98,3 %, u třídy ostatní 94,2 %. Přesnost uživatele dosáhla 95,1 % u třídy ostatní a 98 % u třídy les. Počet špatně zařazených pixelů se oproti klasifikaci pomocí algoritmu random forest zvýšil o 2 na celkových 11 špatně zařazených pixelů.

CART		klasifikovaná data			
		les	ostatní	∑ řádku	PA
skutečnost	les	295	5	300	98,3%
	ostatní	6	98	104	94,2%
	∑ sloupce	301	103	404	
	UA	98,0%	95,1%		
	OA	97,3%			

Tabulka 3: Klasifikovaná data pomocí algoritmu CART dosáhla celkové přesnosti 97,3 %. Ze 404 bodů bylo špatně zařazeno 11 bodů, tedy o dva body více než u klasifikace random forest.

## 5.2 Přesnost klasifikace lesů na první hierarchické úrovni

K hodnocení přesnosti pomocí vrstvy mapování biotopů, bylo nejdříve využito členění lesů na první hierarchické úrovni dle Katalogu biotopů. Do chybové matice vstupovaly ty lesy, které splňovaly původní kritéria. Byly to mokřadní olšiny, lužní lesy, dubohabřiny a acidofilní doubravy. U tohoto dělení už byly hodnoty přesnosti horší. Celková přesnost u snímku klasifikovaným pomocí algoritmu random forest dosáhla 52 % a u snímku, který byl klasifikovaný pomocí algoritmu CART, tato přesnost dosáhla necelých 41 %.

### 5.2.1 RF

Celková přesnost klasifikovaného snímku algoritmem random forest se podle tabulky 4 snížila na 52,2 %. Přesnost zpracovatele byla nejvyšší u třídy lužních lesů, označená kódem L2, s hodnotou přesnosti 74,9 % a třída acidofilních doubrav, označená kódem L7, dosáhla přesnosti 72,2 %. Mokřadní olšiny (L1) dosáhly přesnosti zpracovatele 57,1 %. Nejnižší hodnoty byly dosaženy u třídy dubohabřin (kód L3), konkrétně 24,7 %.

Přesnost uživatele dosáhla hodnot přibližně mezi 13 a 69 % s tím, že právě nejnižší přesnost 13,9 % byla dosažena u třídy mokřadních olšin (kód L1), třída acidofilních doubrav (L7) dosáhla 35,3 %. Třídy lužní lesy (L2) a dubohabřin (L3) měly téměř stejné přesnosti, pohybovaly se přibližně mezi 66 a 69 %.

RF	Přesnost zpracovatele [%]	Přesnost uživatele [%]	Celková přesnost [%]
L1	57,1%	13,9%	52,2%
L2	74,9%	66,5%	
L3	24,7%	69,0%	
L7	72,2%	35,3%	

Tabulka 4: Přesnost zpracovatele, přesnost uživatele a celková přesnost pro snímek klasifikovaný algoritmem random forest. Celková přesnost dosáhla 52,2 %. Hodnoty u přesnosti zpracovatele se pohybují přibližně mezi 25 a 75 %. Přesnost uživatele dosáhla hodnot mezi 13 a 69 %.

Acidofilní doubravy mají sice přes 70 % přesnosti zpracovatele, ale přesnost uživatele dosáhla pouhých 35 %. Příčina tak nízké přesnosti uživatele je dle tabulky č. 5, že více než polovina pixelů, která byla zařazena do třídy acidofilních doubrav, patřila ve skutečnosti do třídy dubohabřin (L3). To je také jeden z důvodů, proč je právě u dubohabřin tak nízká přesnost zpracovatele. Přibližně dvě třetiny pixelů



dubohabřin byly mylně zařazeny do acidofilních doubrav (L7) a do lužních lesů (L2), do každé z těchto kategorií po jedné třetině.

RANDOM FOREST		klasifikovaná data					
		L1	L2	L3	L7	ostatní	∑ řádku
skutečnost	L1	288 500	118 700	33 400	63 600	1 400	505 600
	L2	783 800	6 364 500	617 500	667 500	68 300	8 501 600
	L3	787 000	2 932 900	2 267 000	3 086 900	107 200	9 181 000
	L7	213 500	157 300	368 400	2 079 100	63 200	2 881 500
	∑ sloupce	2 072 800	9 573 400	3 286 300	5 897 100	240 100	21 069 700

Tabulka 5: Chybová matice pro snímek klasifikovaný pomocí algoritmu random forest.

### 5.2.2 CART

Klasifikace provedená pomocí algoritmu CART dosáhla celkové přesnosti 40,7 %, to je uvedeno v tabulce 6 a je to přibližně o 11 % horší výsledek než u klasifikace provedené pomocí algoritmu random forest. Nejlepší přesnosti zpracovatele byly dosaženy u lužních lesů (L2) s 50 % a acidofilních doubrav (L7) s 59 %. Mokřadní olšiny (L1) dosáhly přesnosti zpracovatele 40 % a dubohabřiny (L3) jen 26 %.

Přesnost uživatele měla nejvyšší hodnoty u kategorie lužních lesů (L2) s 58 % a kategorie dubohabřin (L3) s 52 %. Acidofilní doubravy (L7) dosáhly přesnosti uživatele 27 % a úplně nejnižší přesnost uživatele měla kategorie mokřadních olšin (L1) s necelými 8 %.

CART	Přesnost zpracovatele [%]	Přesnost uživatele [%]	Celková přesnost [%]
L1	40,3%	7,9%	40,7%
L2	50,1%	58,2%	
L3	25,8%	52,2%	
L7	59,1%	27,3%	

Tabulka 6: Přesnost zpracovatele, přesnost uživatele a celková přesnost pro snímek klasifikovaný pomocí algoritmu CART. Celková přesnost klasifikovaného snímku dosáhla 40,7 %. Přesnost zpracovatele se pohybuje přibližně mezi 25 a 59 %. Přesnost uživatele má rozpětí hodnot přibližně mezi 8 a 58 %.

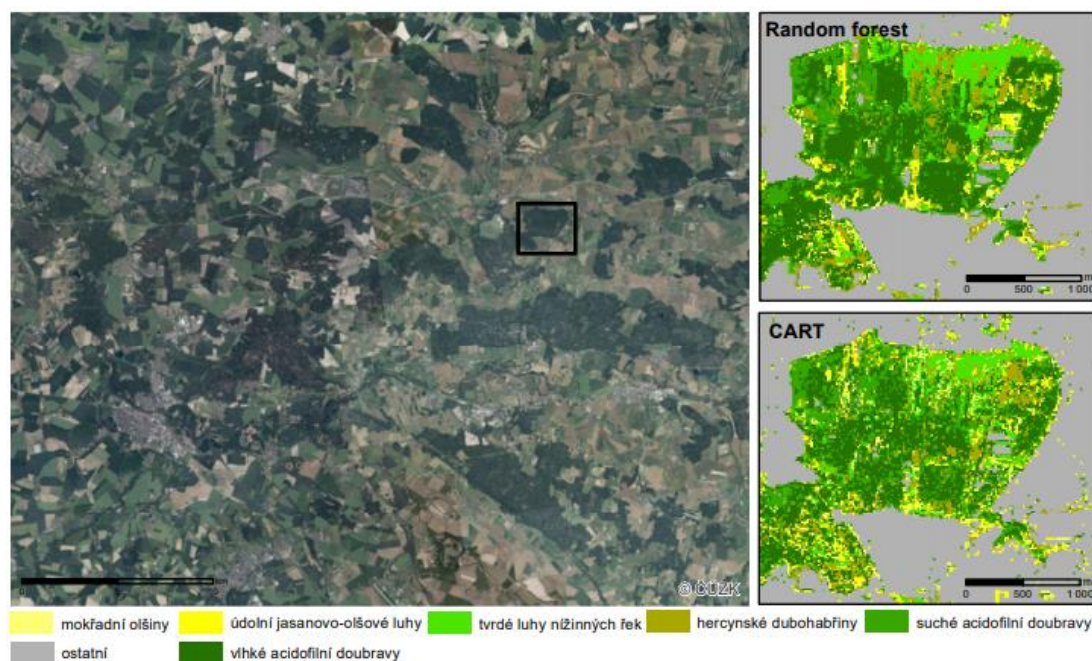
Nejlepších přesností dosáhla třída lužních lesů, avšak správně klasifikováno bylo pouze 50 %. Dle tabulky 7 byly do této třídy nejvíce mylně zařazeny pixely z kategorie dubohabřin.

CART		klasifikovaná data					
		L1	L2	L3	L7	ostatní	Σ řádku
skutečnost	L1	204 000	150 000	60 800	88 800	2 000	505 600
	L2	1 070 300	4 261 300	1 567 700	1 555 800	46 500	8 501 600
	L3	1 044 900	2 522 300	2 372 000	3 167 100	74 700	9 181 000
	L7	264 200	382 300	542 000	1 811 300	67 100	3 066 900
	Σ sloupce	2 583 400	7 315 900	4 542 500	6 623 000	190 300	21 255 100

Tabulka 7: Chybová matice pro klasifikovaná data algoritmem CART.

### 5.3 Přesnost klasifikace lesů na druhé hierarchické úrovni

Dále jsem zjišťovala přesnost klasifikace při rozdělení lesů vrstvy mapování biotopů na druhé hierarchické úrovni, tedy na ty nejpodrobnější třídy, které se v zájmovém území nacházejí. Konkrétně se jedná o mokřadní olšiny, údolní jasanovo-olšové luhy, tvrdé luhy nížinných řek, hercynské dubohabřiny, suché acidofilní doubravy a vlhké acidofilní doubravy. Celková přesnost klasifikovaného snímku pomocí algoritmu random forest dosáhla 43,5 %, kdežto u snímku klasifikovaného pomocí algoritmu CART tato přesnost dosáhla jen 32 %. Rozdíl mezi klasifikacemi je znázorněn pomocí dvou výřezů na obrázku 8.



Obrázek 8: Rozdíly mezi klasifikací random forest a CART (zdroj podkladových dat: AOPK ČR, ČÚZK).

#### 5.3.1 RF

V tabulce 8 je uvedeno, že klasifikace provedená pomocí algoritmu random forest měla pro rozdělení do nejpodrobnějších tříd dle Katalogu biotopů celkovou přesnost 43,5 %. Přesnost zpracovatele dosahuje nejvyšších hodnot u tvrdých luhů

nížinných řek (L2.3) s 64,4 % a u mokřadních olšin (L1) s 57,1 %. Vlhké acidofilní doubravy (L7.2) dosahují přesnosti zpracovatele 47,8 %. U údolních jasanovo-olšových (L2.2) a suchých acidofilních doubrav (L7.1) se přesnost pohybuje mezi 35 a 38 %. Nejnížší přesnosti bylo dosaženo s 24,7 % u třídy hercynských dubohabřin.

Přesnost uživatele je nejvyšší u hercynských dubohabřin (L3.1) s 69 % a u tvrdých luhů nížinných řek (L2.3) s necelými 67 %. Suché a vlhké doubravy (L7.1 a L7.2) mají tuto přesnost mezi 21 a 24 %. Mokřadni olšiny (L1) a údolní jasanovo-olšové luhy (L2.2) dosáhly uživatelských přesností mezi 7 a 14 %.

RF	Přesnost zpracovatele [%]	Přesnost uživatele [%]	Celková přesnost [%]
L1	57,1%	13,9%	43,5%
L2.2	35,2%	7,8%	
L2.3	64,4%	66,8%	
L3.1	24,7%	69,0%	
L7.1	37,6%	21,4%	
L7.2	47,8%	23,6%	

*Tabulka 8: Přesnost zpracovatele, přesnost uživatele a celková přesnost pro klasifikovaný snímek pomocí algoritmu random forest. Celková přesnost klasifikace dosáhla 43,5 %. Přesnost zpracovatele se pohybuje přibližně mezi 24 % a 65 %. Přesnost uživatele má rozpětí hodnot větší, pohybuje se přibližně mezi 7 % a 69 %.*

Nejvyšších hodnot u obou přesností dosáhla třída tvrdých luhů nížinných řek (L2.3). Tyto hodnoty se pohybovaly právě mezi 64,4 a 66,8 %. Dle tabulky 9 byla nejčastěji tato třída špatně klasifikována jako údolní jasanovo-olšové luhy (L2.2). A naopak byly do této třídy špatně přiřazovány pixely, které byly ve skutečnosti hercynské dubohabřiny.

Nejhůře dopadla u obou přesností třída údolní jasanovo-olšové luhy (L2.2), tato třída měla přesnost zpracovatele 35 %. To znamená, že ze všech pixelů patřících ve skutečnosti do této třídy, bylo správně zařazeno jenom 35 %. Přesnost uživatele dosáhla jen necelých 8 %, protože do této třídy byly mylně klasifikovány i ty pixely, které ve skutečnosti náležejí do jiných tříd. Skoro jedna polovina pixelů, která byla klasifikována jako L2.2 ve skutečnosti patřila třídě L2.3 a přibližně jedna třetina, která spadla do této třídy, ve skutečnosti patřila do třídy hercynských dubohabřin (L3.1).

RANDOM FOREST		klasifikovaná data							Σ řádku
		L1	L2.2	L2.3	L3.1	L7.1	L7.2	ostatní	
skutečnost	L1	288 500	85 300	33 400	33 400	16 500	47 100	1 400	505 600
	L2.2	95 300	154 700	104 400	34 000	15 700	30 700	4 500	439 300
	L2.3	688 500	911 700	5 193 700	583 500	287 800	333 300	63 800	8 062 300
	L3.1	787 000	676 300	2 256 600	2 267 000	901 900	2 185 000	107 200	9 181 000
	L7.1	46 400	45 400	69 500	149 400	432 300	366 900	40 200	1 150 100
	L7.2	167 100	111 900	115 900	219 000	363 800	916 100	23 000	1 916 800
	Σ sloupce	2 072 800	1 985 300	7 773 500	3 286 300	2 018 000	3 879 100	240 100	21 255 100

Tabulka 9: Chybová matice pro snímek klasifikovaný pomocí algoritmu random forest.

### 5.3.2 CART

Výsledek klasifikovaného snímku pomocí algoritmu CART dosáhla celkové přesnosti 32 %. Je tedy o 11,5 % horší než klasifikace, která byla provedena pomocí algoritmu random forest. Dle tabulky 10 byla přesnost zpracovatele nejvyšší u mokřadních olšin (L1) a u vlhkých acidofilních doubrav (L7.2) s 40,3 %. Zbytek tříd dosáhl přesnosti zpracovatele mezi 37 a 25 %.

Přesnost uživatele dosáhla nejvyšších hodnot u tvrdých luhů nížinných řek (L2.3) s 59 % a u hercynských dubohabřin (L3.1) s 52,2 %. U dalších tříd se přesnosti uživatele pohybovali přibližně mezi 5 a 19 %.

CART	Přesnost zpracovatele [%]	Přesnost uživatele [%]	Celková přesnost [%]
L1	40,3%	7,9%	32,0%
L2.2	27,9%	5,3%	
L2.3	36,7%	59,0%	
L3.1	25,8%	52,2%	
L7.1	32,2%	14,2%	
L7.2	40,3%	19,2%	

Tabulka 10: Přesnost zpracovatele, přesnost uživatele a celková přesnost pro snímek klasifikovaný pomocí algoritmu CART. Celková přesnost klasifikace je 32 %. Přesnost zpracovatele dosáhla hodnot přibližně mezi 25 a 40 %. Přesnost uživatele se pohybuje přibližně mezi 5 a 59 %.

Nejhůře byla rozpoznatelná třída údolních jasanovo-olšových luhů (L2.2). Tato třída byla podle tabulky 11 nejčastěji špatně řazena do mokřadních olšin (L1) a tvrdých luhů nížinných řek (L2.3). Z celkové hodnoty zařazených pixelů do této třídy byla jedna polovina pixelů náležící ve skutečnosti třídě tvrdých luhů nížinných řek.

CART		klasifikovaná data							
		L1	L2.2	L2.3	L3.1	L7.1	L7.2	ostatní	Σ řádku
skutečnost	L1	204 000	93 500	56 500	60 800	25 500	63 300	2 000	505 600
	L2.2	87 400	122 500	86 800	59 500	23 300	54 200	5 600	439 300
	L2.3	982 900	1 094 200	2 957 800	1 508 200	784 700	693 600	40 900	8 062 300
	L3.1	1 044 900	811 600	1 710 700	2 372 000	1 093 400	2 073 700	74 700	9 181 000
	L7.1	85 700	58 000	74 800	162 900	370 800	359 400	38 500	1 150 100
	L7.2	178 500	122 800	126 700	379 100	307 800	773 300	28 600	1 916 800
	Σ sloupce	2 583 400	2 302 600	5 013 300	4 542 500	2 605 500	4 017 500	190 300	21 255 100

Tabulka 11: Chybová matice pro klasifikovaný snímek pomocí algoritmu CART.

## 6 Diskuse

Data dálkového průzkumu země jsou pro klasifikaci lesa používány již mnoho let, ale mapování lesů dané Katalogem biotopů České republiky (Chytrý et al., 2010), dále jen Katalog, zatím v České republice nemá obdoby. V této práci bylo zjišťováno, s jakou přesností se pomocí řízené klasifikace lze odlišit les od ostatních biotopů. Jak se odliší vybrané lesy dané Katalogem mezi sebou na první a druhé hierarchické úrovni. Vybrané lesy podle první hierarchické úrovně dané Katalogem jsou mokřadní olšiny, lužní lesy, dubohabřiny a acidofilní doubravy. Vybrané lesy podle druhé hierarchické úrovně jsou mokřadní olšiny, údolní jasanovo-olšové luhy, tvrdé luhy nížinných řek, hercynské dubohabřiny, suché acidofilní doubravy a vlhké acidofilní doubravy.

Ke klasifikaci byly použity trénovací body, které byly rozloženy tak, aby v každé třídě byl stejný počet trénovacích bodů. K určování přesnosti u odlišení lesa a kategorie ostatní bylo využito validačních dat. Rozložení těchto bodů je příčinou výsledné téměř 98% přesnosti. Při vizuálním porovnávání klasifikovaných snímků s neklasifikovaným snímkem bylo zřejmé, že takto vysoká přesnost je zkrácena právě rozmístěním validačních bodů. Zejména pomocí klasifikátoru CART byla někdy pole klasifikována jako les, to je zobrazeno na obrázku 10. Klasifikace těchto lesních biotopů algoritmem CART se i dále ukázala jako méně efektivní, než je algoritmus random forest. To může být zapříčiněno tím, že se stromy v algoritmu CART na konci ještě „prořezávají“, a to na rozdíl algoritmus random forest nedělá. Po vizuální kontrole klasifikace random forest se nenašly žádné velké nepřesnosti, byly krásně rozlišeny i stromy podél řek.



Obrázek 9: Vlevo výřez z klasifikace pomocí algoritmu CART, vpravo ortofoto stejného území. Tento klasifikátor někdy mylně určil pole jako les (zdroj podkladových dat: AOPK ČR, ČÚZK).

Výsledky přesností při rozlišování lesů mezi sebou dané Katalogem podle první hierarchické úrovně byly lepší u klasifikátou random forest s celkovou přesností přes 50 %. Nejlépe od sebe šly rozlišit lužní lesy a acidofilní doubravy, 75 % pixelů bylo správně zařazeno do těchto tříd. Tyto třídy měly rozdílné spektrální křivky a skládají se z rozdílných druhů listnatých stromů. V lužních lesích dominuje olše a v acidofilních doubravách dub. Ostatní třídy mezi sebou byly zaměňovány více, protože například u mokřadních lesů dominuje olše stejně jako u lužních lesů.

Rozlišování lesů mezi sebou podle druhé hierarchické úrovně jsou mnohem méně uspokojivé. Bylo dosaženo celkové přesnosti přes 40 %. Nejlépe byly klasifikovány tvrdé luhy nížinných řek, kde bylo správně klasifikováno téměř 65 %. Tato kategorie lesů měla v dubnovém kompozitu nejvyšší odrazivost ze všech lesů.

Za účelem zvýšení přesnosti klasifikace jsem také zkoušela řízenou klasifikaci pomocí jiných trénovacích dat, a to podle rozlohy dané třídy. Data byla tvořena tak, aby i v nejméně početné třídě bylo 150 bodů a pak se podle rozlohy počet bodů zvyšoval. Tato cesta se ukázala jako neefektivní, protože se při spuštění klasifikace v Google Earth Engine ukázala chyba hlásící přerušení z důvodu překročení 5000 prvků. S touto chybou jsem se setkala již na začátku mé bakalářské práce, protože jsem původně chtěla zpracovávat území, které se rozkládá přes celý snímek družice Sentinel-2. Kvůli této chybě jsem byla nucena území zmenšit na to stávající.

Další pokus pro tvorbu trénovacích dat byl takový, že do každého polygonu byl náhodně vložen jeden bod. V nejméně početné třídě nakonec bylo přibližně jen 30 bodů a v té nejpočetnější bylo kolem 330 bodů. Takto rozložená trénovací data se také ukázala jako neefektivní, protože některé třídy nebyly rozpoznány vůbec.

Příčinou nízkých přesností může být i zvolené zájmového území. Po vyfiltrování lesů, které byly aktualizovány po roce 2015 včetně a zároveň měly rozlohu celkem větší než 1 km<sup>2</sup>, se v zájmovém území nacházeli jen lesy s dominancí listnatých stromů. Jehličnaté lesy, které nejsou nově zaktualizovány mnohdy neodpovídají skutečnosti, a proto nebyly přidány k trénovacím datům. K získání vyšších přesností by mohlo být dosaženo zvolením území, kde by se v hojném zastoupení nacházely i ty kategorie lesů, kde dominují jehličnaté stromy.

Porovnání s jednou z prvních studií využívající snímky družice Sentinel-2 ke klasifikaci druhů stromů v Bavorsku (Immitzer et al., 2016) ukázalo, že tato studie má

lepší celkovou přesnost přibližně o 10 %, než má bakalářská práce. Lepší výsledky studie mohou být zapříčiněny tím, že od sebe byly rozlišovány jednotlivé druhy stromů, konkrétně buk, dub, borovice smrk a další.

Při porovnání s klasifikací druhů stromů v Polských Karpatech pomocí algoritmu SVM (Grabska et al., 2020), vyšly výsledky mé bakalářské práce hůře. To může být zapříčiněno tím, že v mé práci jsem vycházela z Vrstvy mapování biotopů, kde jsou kategorie lesů dané Katalogem smíšené, to znamená, že například v kategorii lužních lesů je sice dominantní jeden druh, ale jsou tam přimísěny i další jiné druhy stromů. Naopak Grabska et al. (2020) od sebe odlišovala jednotlivé druhy stromů, například borovici, smrk, buk, dub a další a dosáhla celkové přesnosti klasifikace přibližně 85 %.

Také Persson et al. (2018) rozlišovali ve Švédsku jednotlivé druhy stromů mezi sebou pomocí řízené klasifikace za použití algoritmu random forest. Byly od sebe opět odlišovány smrky, borovice, modříny, břízy a duby s celkovou přesností necelých 90 %. V této studii naopak od mé práce byly použity ke klasifikaci snímky ze 4 různých období a do klasifikace vstupovalo celkem 40 pásem.



## 7 Závěr

Tato bakalářská práce byla zaměřena na rozpoznávání lesů dané Katalogem biotopů ČR pomocí řízené klasifikace multispektrálních snímků družice Sentinel-2. Jejím cílem bylo zjistit, jaké podrobnosti lesů dané Katalogem biotopů ČR (Chytrý et al., 2010) bude dosaženo. Ke klasifikaci byly použity snímky z roku 2018 z období dubna a července. Trénovací data byla vytvořena z Vrstvy mapování biotopů. Řízená klasifikace byla provedena v Google Earth Engine. Jako klasifikátor byl použit random forest a classification and regression trees.

Nejdříve bylo zjišťováno, jak přesně se odliší les od okolní krajiny. Ukázalo se, že se les od ostatních biotopů rozlišil téměř dokonale s celkovou přesností 97,8 % pomocí klasifikátoru random forest a s 97,3 % pomocí klasifikátoru CART. Takto vysoká přesnost je ale do jisté míry ovlivněna rozmístěním validačních bodů, protože po vizuálním srovnání klasifikovaného obrazu se satelitním snímkem bylo vidět, že místy byla klasifikována například pole jako les, a to zejména u klasifikátoru CART.

Dalším cílem bylo zjistit, s jakou přesností budou rozlišeny lesy mezi sebou podle první hierarchické úrovně, tedy třídy mokřadní olšiny, lužní lesy, dubohabřiny a acidofilní doubravy. Klasifikace pomocí algoritmu random forest se v tomto případě ukázala jako efektivnější než algoritmus CART. Klasifikátor random forest dosáhl celkové přesnosti 52,2 %. Nejlépe byly klasifikovány lužní lesy a acidofilní doubravy. Téměř 75 % pixelů, které ve skutečnosti patřily do těchto tříd, tam byly i přiřazeny tímto klasifikátorem. Pomocí klasifikátoru CART byly také nejlépe určeny lužní lesy a acidofilní doubravy. Jejich přesnosti však nepřesáhly ani 60 %.

Jako poslední cíl bylo určení přesnosti u odlišení lesů mezi sebou podle druhé, tedy nejpodrobnější, hierarchické úrovně. V zájmovém území se jednalo o mokřadní olšiny, údolní jasanovo-olšové luhy, tvrdé luhy nížinných řek, hercynské dubohabřiny, suché acidofilní doubravy a vlhké acidofilní doubravy. Vyšší celkové přesnosti klasifikace dosáhl opět algoritmus random forest s 43,5 %. Nejlépe byly klasifikovány tvrdé luhy nížinných řek, kde bylo z pixelů, které tuto třídu ve skutečnosti představují, správně určeno necelých 65 %. Horší byly výsledky pomocí algoritmu CART, kde celková přesnost dosáhla 32,2 %.

Výsledky řízené klasifikace ukazují potenciál využití multispektrálních snímků k detekci lesních biotopů soustavy Natura 2000. Avšak rozlišování listnatých lesů

mezi sebou je problematické, protože mají podobnou spektrální odrazivost a nejsou jednodruhové. Řízená klasifikace tedy dokázala rozlišit mezi sebou alespoň některé třídy lesů na první hierarchické úrovni. Z těchto výsledků je patrné, že data družice Sentinel-2 by mohla být pomocným nástrojem k mapování biotopů. Tyto snímky by jistě mohly pomoci s rozlišováním základních biotopů, ale zcela nenahradí podrobné terénní mapování.

## 8 Seznam literatury

- Adamo, M., Tarantino, C., Tomaselli, V., Veronico, G., Nagendra, H., & Blonda, P. (2016). Habitat mapping of coastal wetlands using expert knowledge and Earth observation data. *Journal of Applied Ecology*, 53(5), 1521–1532. <https://doi.org/10.1111/1365-2664.12695>
- Agentura ochrany přírody a krajiny. Aktualizace základního mapování biotopů dostupné z: <https://data.nature.cz/sds/7>
- Baamonde, S., Cabana, M., Sillero, N., Penedo, M. G., Naveira, H., & Novo, J. (2019). Fully automatic multi-temporal land cover classification using Sentinel-2 image data. *Procedia Computer Science*, 159, 650–657. <https://doi.org/10.1016/j.procs.2019.09.220>
- Bakshi Ch., Random Forest Regression. [cit 8.3. 2021] dostupné z: Retrieved March 8, 2021, from <https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>
- Barsi, Á., Kugler, Z., László, I., Szabó, G., & Abdulmutalib, H. M. (2018). Accuracy Dimensions in Remote Sensing. <https://doi.org/10.5194/isprs-archives-XLII-3-61-2018>
- Breiman, L. (2001). Random forests. *Random Forests*, 1–122. <https://doi.org/10.1201/9780429469275-8>
- Cârlan, I., Mihai, B. A., Nistor, C., & Große-Stoltenberg, A. (2020). Identifying urban vegetation stress factors based on open access remote sensing imagery and field observations. *Ecological Informatics*, 55, 101032. <https://doi.org/10.1016/j.ecoinf.2019.101032>
- Červená, L. (2018). Hodnocení přesnosti klasifikací. <https://www.natur.cuni.cz/geografie/geoinformatika-kartografie/ke-stazeni/projekty/moderni-geoinformacni-metody-ve-vyuce-gis-kartografie-a-dpz/hodnoceni-presnosti-klasifikace/>
- Dobrovolný, P. (1998). Dálkový průzkum Země, Digitální zpracování obrazu (1998. vyd.). MU Brno.
- ESA (©2000–2021) User Guides – Sentinel-2 MSI – Level-1C Product - Sentinel Online - Sentinel. [cit 9.3. 2021] dostupné z: (n.d.). Retrieved March 9, 2021, from <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/product-types/level-1c>

- ESA (©2000–2021), Sentinel-2 - Missions - Instrument Payload - Sentinel Handbook - Sentinel. (n.d.). Retrieved March 6, 2021, from. [cit 8.3. 2021] dostupné z: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload>
- ESA (©2000–2021), Spatial – Resolutions – Sentinel-2 MSI - User Guides - Sentinel Online - Sentinel. (n.d.). Retrieved March 6, 2021, from <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>
- ESA (©2000–2021), User Guides – Sentinel-2 MSI – Level-2 Processing – Sentinel Online – Sentinel [cit 9.3. 2021] dostupné z: . (n.d.). Retrieved March 9, 2021, from <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/processing-levels/level-2>
- ESA (©2000–2021), Colour Vision for Copernicus. The story of Sentinel-2. ESA Bulletin, 155, 88. <http://esamultimedia.esa.int/multimedia/publications/ESA-Bulletin-161/offline/download.pdf>
- Evropská Komise – Základní informace – Natura 2000 – Síť chráněných oblastí EU. . (n.d.). [cit 1.3. 2021] dostupné z Retrieved March 7, 2021, from [https://ec.europa.eu/environment/basics/natural-capital/natura2000/index\\_cs.htm](https://ec.europa.eu/environment/basics/natural-capital/natura2000/index_cs.htm)
- Gibson, R., Danaher, T., Hehir, W., & Collins, L. (2020). A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote Sensing of Environment*, 240, 111702. <https://doi.org/10.1016/j.rse.2020.111702>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Grabska, E., Frantz, D., & Ostapowicz, K. (2020). Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. *Remote Sensing of Environment*, 251, 112103. <https://doi.org/10.1016/j.rse.2020.112103>

- Guth J. (2002) Metodiky mapování biotopů soustavy Natura 2000 a Smaragd., AOPK ČR, Praha
- Härtel Handrij, Lončáková Jarmila & Hošek Michael [eds.]: Mapování biotopů v České republice. Východiska, výsledky, perspektivy. 1. vydání. Praha: Agentura ochrany přírody a krajiny ČR, 2009. 125 s. ISBN 978-80-87051-36-8
- Huete, A. R. (2004). Remote Sensing for Environmental Monitoring. In Environmental Monitoring and Characterization (pp. 183–206). Elsevier Inc. <https://doi.org/10.1016/B978-012064477-3/50013-8>
- Humboldt State University, Vegetation and EMR. [(n.d.). Retrieved March 7.3., 2021] dostupné z: , from [http://gsp.humboldt.edu/OLM/Courses/GSP\\_216\\_Online/lesson2-1/vegetation.html](http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson2-1/vegetation.html)
- Chemura, A., Mutanga, O., Odindi, J., & Kutuwayo, D. (2018). Mapping spatial variability of foliar nitrogen in coffee (*Coffea arabica* L.) plantations with multispectral Sentinel-2 MSI data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138, 1–11. <https://doi.org/10.1016/j.isprsjprs.2018.02.004>
- Chytrý, M., Kučera, T., Kočí, M., Grulich, V., & Lustyk, P. (2010). Katalog biotopů České republiky. Druhé vydání.
- Immitzer, M., Vuolo, F., Atzberger, C., Sarathi Roy, P., & Thenkabail, P. S. (2016). remote sensing First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. <https://doi.org/10.3390/rs8030166>
- Klashka J, Kotrč E. (2004). Klasifikační a regresní lesy [cit 10.3.2021] dostupné z: <https://www.statspol.cz/robust/robust2004/klaschka.pdf>
- Kropáček J., Moravec D., Komárek J. (2020). Dálkový průzkum Země I
- Learn And Code Confusion Matrix With Python. [cit 12.3. 2021] dostupné zRetrieved March 12, 2021:, from <https://www.nbshare.io/notebook/626706996/Learn-And-Code-Confusion-Matrix-With-Python/>
- Lustyk P. (2016) Metodika aktualizace vrstvy mapování biotopů, AOPK ČR, Praha

- Masaila, A., Karlova, U., Matematicko-fyzik, P., & Studijn, M. (2013). Regresní stromy.
- Mücher, C. A., Kooistra, L., Vermeulen, M., Haest, B., Spanhove, T., Delalieux, S., vanden Borre, J., & Schmidt, A. M. (2010). Object identification and characterization with hyperspectral imagery to identify structure and function of Natura 2000 habitats. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-4/C7, January, 1–6. <http://www.isprs.org/proceedings/XXXVIII/4-C7/pdf/Muecher-102.pdf>
- Ministerstvo životního prostředí (©2008–2020) Natura 2000 [cit 20.2.2021] dostupné z: [https://www.mzp.cz/cz/natura\\_2000](https://www.mzp.cz/cz/natura_2000)
- Nomura, K., & Mitchard, E. T. A. (2018). remote sensing More Than Meets the Eye: Using Sentinel-2 to Map Small Plantations in Complex Forest Landscapes. MDPI. <https://doi.org/10.3390/rs10111693>
- Perrson M., Lindberg E., Reese H. (2018) Tree Species Classification with Multi-Temporal Sentinel-2 Data, MDPI <https://www.mdpi.com/2072-4292/10/11/1794>
- Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., & Mononen, J. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behavioural Processes*, 148, 56–62. <https://doi.org/10.1016/j.beproc.2018.01.004>
- Směrnice rady 92/43/EHS, o ochraně přírodních stanovišť, volně žijících živočichů a planě rostoucích rostlin
- Spadoni, G. L., Cavalli, A., Congedo, L., & Munafò, M. (2020). Analysis of Normalized Difference Vegetation Index (NDVI) multi-temporal series for the production of forest cartography. *Remote Sensing Applications: Society and Environment*, 20, 100419. <https://doi.org/10.1016/j.rsase.2020.100419>
- Valentini, E., Taramelli, A., Filipponi, F., & Giulio, S. (2015). An effective procedure for EUNIS and Natura 2000 habitat type mapping in estuarine ecosystems integrating ecological knowledge and remote sensing analysis. *Ocean and Coastal Management*, 108, 52–64. <https://doi.org/10.1016/j.ocecoaman.2014.07.015>

- Weigand, M., Staab, J., Wurm, M., & Taubenböck, H. (2020). Spatial and semantic effects of LUCAS samples on fully automated land use/land cover classification in high-resolution Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102065.  
<https://doi.org/10.1016/j.jag.2020.102065>
- Yiu T. (2019) *Towards Data Science. Understanding Random Forest* March [cit 9.3. 2021] dostupné z: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
- Žubrietovský, L., Švábová, L., & Šváb, O. (2017). Sentinel 2 – DATOVÉ SPECIFIKACE, Základní informace, Geografické pokrytí.  
<http://collgs.czechspaceportal.cz/sentinel-2-datove-specifikace/>

# Přílohy

## Příloha 1: Skript

```
/*
Zdroje:
- tutorial https://developers.google.com/earth-engine/tutorials/community/classify-maizeland-ng
- paper "More than meets the eye: using Sentinel-2 to map small plantations
in complex forest landscapes" by Keiko Nomura and Edward TA Mitchard.
*/
// Vstupní data (hranice zájmového území, trénovací data, validační data)
var roi = ee.FeatureCollection("users/spyrtylc/uzemi1/uzemi_1"),
    trainingSample = ee.FeatureCollection("users/spyrtylc/uzemi1/rovnomerneTrenovaciU1"),
    testingSample = ee.FeatureCollection("users/spyrtylc/uzemi1/rovnomerneValidacniU1");

var areaName = 'Zajmove uzemi';
var year = '2018';
var startDate = '2018-04-15';
var endDate = '2018-04-30';
var randomSeed = 0;
var startDate2 = '2018-07-01';
var endDate2 = '2018-07-15';

// Zobrazeno v konzoli
print('Area' + " " + areaName + ' ' + year);
var checkbox1 = ui.Checkbox('S2 composit', true);
var checkbox2 = ui.Checkbox('NDVI', false);
var checkbox3 = ui.Checkbox('Texture', false);
var checkbox4 = ui.Checkbox('ClassifiedRF', true);
var checkbox5 = ui.Checkbox('ClassifiedCart', true);

checkbox1.onChange(function(checked) { Map.layers().get(0).setShown(checked); });
checkbox2.onChange(function(checked) { Map.layers().get(1).setShown(checked); });
checkbox3.onChange(function(checked) { Map.layers().get(2).setShown(checked); });
checkbox4.onChange(function(checked) { Map.layers().get(3).setShown(checked); });
checkbox4.onChange(function(checked) { Map.layers().get(4).setShown(checked); });

print(checkbox5);
print(checkbox4);
print(checkbox3);
print(checkbox2);
print(checkbox1);

// Sentinel-2 data
var s2 = ee.ImageCollection('COPERNICUS/S2_SR')
    .filterDate(startDate, endDate)
    .filterBounds(roi);
var s2_2 = ee.ImageCollection('COPERNICUS/S2_SR')
    .filterDate(startDate2, endDate2)
    .filterBounds(roi);

// Zobrazení dat vhodných snímků
var list = ee.List(s2.aggregate_array("system:time_start")).map(function(d) { return ee.Date(d); });
print('S2 images of the area during the study period',list);

var list = ee.List(s2_2.aggregate_array("system:time_start")).map(function(d) { return ee.Date(d); });
print('S2 images of the area during the study period',list);

// Zobrazení procentuální oblačnosti ve vybraných snímcích
var getCloudScores = function(img){
```



```

    var value = ee.Image(img).get('CLOUDY_PIXEL_PERCENTAGE');
    return ee.Feature(null, {'score': value});
  };

  var s2clouds = s2.map(getCloudScores);
  print ('cloud score', ui.Chart.feature.byFeature(s2clouds));
  var s2clouds2 = s2_2.map(getCloudScores);
  print ('cloud score', ui.Chart.feature.byFeature(s2clouds));

  // Vybrání snímků s maximální oblačností snímku 5 %
  var s2image = ee.ImageCollection('COPERNICUS/S2_SR')
    .filterDate(startDate, endDate)
    .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 5))
    .sort('CLOUDY_PIXEL_PERCENTAGE')
    .filterBounds(roi)
    .map(function(img){
      var t = img.select(['B1','B2','B3','B4','B5','B6','B7','B8','B8A',
'B9','B11','B12']).divide(10000); //Rescale to 0-1
      var out = t.copyProperties(img).copyProperties(img,['system:time_start']);
      return out;
    })

  var s2image_2 = ee.ImageCollection('COPERNICUS/S2_SR')
    .filterDate(startDate2, endDate2)
    .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 5))
    .sort('CLOUDY_PIXEL_PERCENTAGE')
    .filterBounds(roi)
    .map(function(img){
      var t = img.select(['B2','B3','B4','B5','B6','B7','B8','B8A','B11','B12']).divide(10000);
      var out = t.copyProperties(img).copyProperties(img,['system:time_start']);
      return out;
    });

  //Tvorba kompozitu z nejméně oblačných snímků, přejmenování pásem a zobrazení kompozitů
  var s2composit =
s2image.reduce(ee.Reducer.median()).select(['B1_median','B2_median','B3_median','B4_median','B5
_median','B6_median','B7_median','B8_median','B8A_median','B9_median',
'B11_median','B12_median'],['aerosol','blue','green','red','red1','red2','red3','nir','red4','h2o','swir1',
'swir2']);
  var s2ROI = s2composit.clip(roi)
  var s2composit_2 =
s2image_2.reduce(ee.Reducer.median()).select(['B2_median','B3_median','B4_median','B5_median',
B6_median','B7_median','B8_median','B8A_median'],['blue2','green2','red2',
'vred1','vnir1','vnir2','vnir3','vnir4']);
  var s2ROI_2 = s2composit_2.clip(roi)

  var vizParams = {bands: ['red','green','blue'], min: 0, max: 0.3};
  Map.addLayer(s2ROI, vizParams, 'S2 composit');
  Map.addLayer(s2ROI_2, {bands: ['red2','green2','blue2'], min: 0, max: 0.3});
  Map.addLayer(s2ROI,{bands: ['nir','red','green'], min: 0, max: 0.3},'CIR')

  // Tvorba NDVI
  var red = s2ROI.select('red');
  var nir = s2ROI.select('nir');
  var ndvi = nir.subtract(red).divide(nir.add(red)).rename('NDVI');

  var ndviParams = {min: -1, max: 1, palette: ['blue','white','green']};
  Map.addLayer(ndvi, ndviParams, 'NDVI', false);

```

```

// Přidání pásem k ke kompozitu
var s2final = ee.Image(s2ROI).addBands(ndvi).addBands(s2ROI_2);

// Vybraná pásma, která se použijí ke klasifikaci
var bands = ['red1','red2','red3','nir','red4', 'swir2','NDVI','NDVI_stdDev',
'vred1','vnir1','vnir2','vnir3','vnir4'];
var s2classification = s2final.select(bands);

// Zobrazení trénovacích dat v mapě
Map.addLayer(trainingSample);

// Přepsání hodnot zapsaných v atributové tabulce z kódu biotopu na čísla od 0
var trainingSampleClass = trainingSample.remap(['L1 (100)', 'L2.2 (100)', 'L2.3 (100)', 'L3.1
(100)', 'L7.1 (100)', 'L7.2 (100)', 'ostatní'], [0, 1, 2, 3, 4, 5, 6], 'BIOTOP_SEZ');

print (trainingSample)
print (trainingSampleClass)

// Získání hodnot pro klasifikaci
var training = s2classification.sampleRegions({
  collection: trainingSampleClass,
  properties: ['BIOTOP_SEZ'],
  scale: 10,
});

// Klasifikace

// Random forest s nastaveným počtem stromů na 100
var trainedRF = ee.Classifier.smileRandomForest({numberOfTrees: 100,})
  .train(training, 'BIOTOP_SEZ', bands);

// CART klasifikátor s defaultním nastavením
var trainedCart = ee.Classifier.smileCart().train(training, 'BIOTOP_SEZ', bands);

// Klasifikace vybraných snímků
var classifiedRF = s2classification.classify(trainedRF);
var classifiedCart = s2classification.classify(trainedCart);

// Vyrobení palety barev pro zobrazení
var palette =['4EEE94',// L1
'00EE00',// L2.2
'9ACD32',// L2.3
'CDC673',// L3.1
'FFFF00',// L7.1
'008000',// L7.2
'C0C0C0',// ostatní
];

// Zobrazení klasifikovaných snímků
Map.addLayer(classifiedRF, {min: 0, max: 6, palette: palette}, 'ClassifiedRF');
Map.addLayer(classifiedCart, {min: 0, max: 6, palette: palette}, 'ClassifiedCart');
Map.centerObject(roi, 10);

// Tvorba legendy
var labels = ['L1 Mokřadní olšiny', 'L2.2 Údolní jasanovo-olšové luhy ', 'L2.3 Tvrdé luhy nížinných
řek', 'L3.1 Hercynské dubohabřiny', 'L7.1 Suché acidofilní doubravy', 'L7.2 Vlhké acidofilní
doubravy', 'ostatní'];
var add_legend = function(title, lbl, pal) {

```

```

var legend = ui.Panel({ style: { position: 'bottom-left' } }), entry;
legend.add(ui.Label({ value: title, style: { fontWeight: 'bold', fontSize: '18px', margin: '0 0 4px 0',
padding: '0px' } }));
for (var x = 0; x < lbl.length; x++){
  entry = [ ui.Label({ style: { color: pal[x], border: '1px solid black', margin: '0 0 4px 0'}, value: '■' }),
  ui.Label({ value: labels[x], style: { margin: '0 0 4px 4px' } } ) ];
  legend.add(ui.Panel(entry, ui.Panel.Layout.Flow('horizontal')));
} Map.add(legend); };

add_legend('Legend', labels, palette);

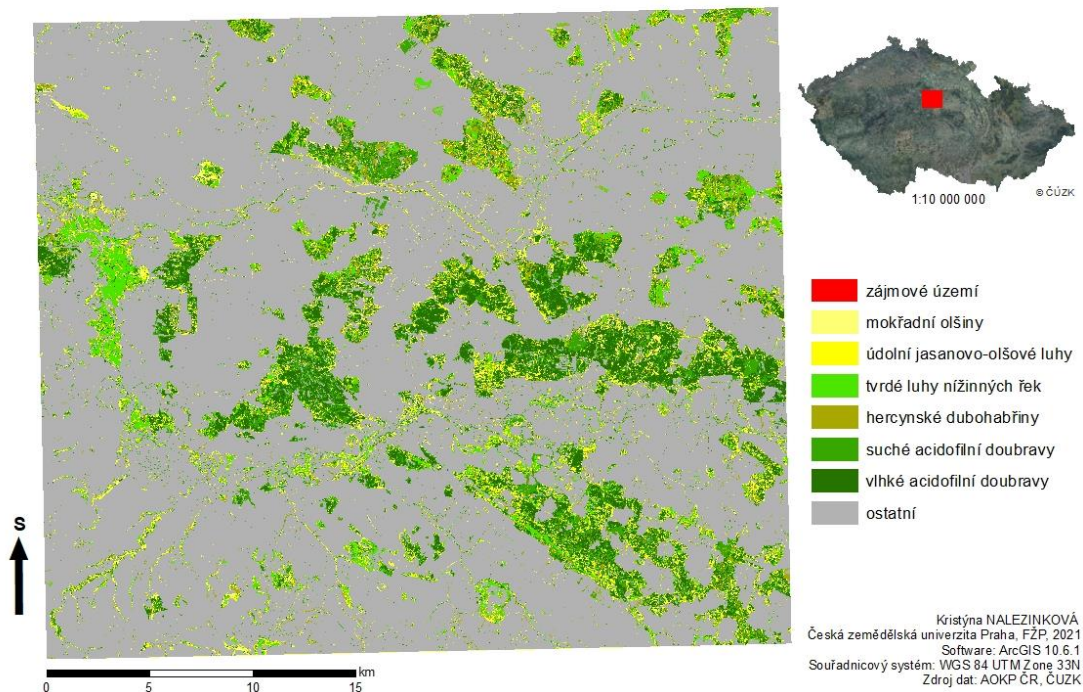
// Export klasifikovaných snímků
Export.image.toDrive({
  image: classifiedCart,
  description: 'CART',
  scale: 10,
  region: roi
});

Export.image.toDrive({
  image: classifiedRF,
  description: 'RandomForest',
  scale: 10,
  region: roi
});

```

## Příloha 2: Klasifikovaný snímek pomocí algoritmu random forest

### KLASIFIKOVANÝ SNÍMEK pomocí algoritmu random forest



### Příloha 3: Klasifikovaný snímek pomocí algoritmu CART

## KLASIFIKOVANÝ SNÍMEK pomocí algoritmu CART

