

**Univerzita Karlova**

**Přírodovědecká fakulta**

Obor kartografie a geoinformatiky



Bc. Michal Launer

**Určení druhové skladby lesa z družicových dat**  
**Forest species determination from satellite data**

**Diplomová práce**

Vedoucí diplomové práce: Doc. Ing. Jan Kolář, CSc.

Praha, 2018

**Prohlášení:**

Prohlašuji, že jsem závěrečnou práci zpracoval/a samostatně a že jsem uvedl/a všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 14. 04. 2018

Podpis

## **POĎAKOVANIE**

Ďakujem Doc. Ing. Jánovi Kolářovi, CSc. za jeho odborné vedenie práce, za jeho cenné rady a pripomienky, ktoré mi poskytoval počas celej doby práce na diplomovej práci. Ďakujem tiež RNDr. Jakubovi Lsákovi, Ph.D. za jeho cenné rady ohľadom používania softvéru ArcGIS.

## Určení druhové skladby lesa z družicových dat

### Abstrakt

V tejto práci bolo skúmané druhové zloženie lesov z družicových snímok pomocou pixelovej klasifikácie. Výskum bol robený v 24 lokalitách lesných porastov v Usteckom, Karlovarskom, Plzeňskom a Stredočeskom kraji v Českej republike. V práci boli použité dáta z družíc Landsat-8 a zo Sentinel-2 z letného ročného obdobia a ako klasifikátor bol použitý Random Forest. Ako referenčné dáta boli použité údaje o druhovom zložení lesných porastov z mapového portálu LhpoMap.

Metóda práce spočívala v tom, že sa pomocou rozsiahlej literárnej rešerši vybral najvhodnejší klasifikátor a zvolili sa najvhodnejšie hodnoty vstupných parametrov pre dosiahnutie, čo najvyššej celkovej presnosti klasifikácie.

Praktická časť bola zameraná na tvorbu klasifikačného procesu zo softvérového hľadiska. Správnosť klasifikácií jednotlivých snímok bola overená pomocou chybových matíc.

Na základe literárnej rešerši bol pre klasifikáciu snímok použitý klasifikátor Random Forest. Hodnoty parametrov boli použité Entropy kritérium, 500 rozhodovacích stromov a ostatným parametrom boli ponechané defaultné hodnoty. Celý klasifikačný proces bol robený v softvéroch ArcMap a ArcGIS Pro s využitím jazyka Python pomocou modulu sklearn.ensemble a jeho knižníc.

Výsledky klasifikácie snímok dosahovali presnosť od 88 do 96 %.

**Kľúčové slová:** dálkový průzkum, lesní porost, druhy lesa, lesnická hospodářská mapa



## **Forest species determination from satellite data**

### **Abstract**

This thesis examines the species composition of forests from satellite images using the pixel classification. The research was done on 24 forest locations in The Ustecký Region, The Karlovarský Region, The Plzeňský Region and The Central Bohemian Region in the Czech Republic. In this thesis, data from the Landsat-8 and Sentinel-2 satellites from summer season and the Random Forest Classifier method were used. The layer of species composition of forests from map portal LhpoMap was used as reference data.

The method of work consisted of a broad literature search to select the most favourable classifier and to choose the most advantageous input parameter values to achieve the highest overall accuracy of the classification.

The practical part was focused on creating a software classification process. The accuracy of the individual image values was verified using matrix errors.

Based on the literature search, the Random Forest classifier was used to classify the images. Parameter values were used for the Entropy criterion, 500 decision trees, and the other parameters were left with default values. The entire classification process was performed in ArcMap and ArcGIS Pro software using Python programming language with the help of the sklearn.ensemble module and its libraries.

The results of the image classification achieved from 88 to 96% of accuracy.

**Keywords:** remote sensing, forest canopy, forest tree types, forestry map

# OBSAH

1	Úvod .....	14
1.1	Využitie diaľkového prieskumu zeme v lesníctve .....	14
1.2	Druhovú skladbu lesa .....	15
1.3	Vymedzenie oblasti výskumnej témy diplomovej práce.....	16
1.4	Základný popis výskumných otázok, ktoré pripadajú v úvahu pre spracovanie v diplomovej práci.....	16
2	Analýza publikovaných prác – metód .....	17
2.1	Metodika práce .....	17
2.2	Rešerš literatúry – 10 kľúčových štúdií .....	18
2.3	Druhy drevín.....	22
2.4	Pásma družíc a ich vlastnosti .....	24
2.5	Výsledky klasifikácie vybraných druhov drevín.....	27
2.6	Ročné obdobie .....	28
2.7	Klasifikátor.....	32
2.7.1	Porovnanie klasifikátorov na základe výsledkov presnosti klasifikácie za jednotlivé práce .....	33
2.7.2	Prečo iné klasifikátory nie sú vhodnejšie pre použitie v tejto práci.....	34
2.8	Parametre klasifikátora .....	34
2.9	Programovací jazyk .....	36
2.10	Záver.....	38
3	Random Forest.....	39
3.1	Popis a vývoj metódy .....	39
3.2	Algoritmus.....	40
3.2.1	Predpoklady vytvorenia rozhodovacieho stromu:.....	41
3.2.2	Rozhodovací stromový algoritmus:.....	42
3.3	Praktické použitie RF .....	42
3.4	Procesy použité v RF .....	43
3.4.1	Bootstrapping.....	43
3.4.2	Bagging .....	44
3.4.3	Information gain.....	44
3.4.4	Gini index.....	44
3.4.5	Overfitting .....	44
4	Územie a dáta .....	46

4.1	Opis územia .....	46
4.2	Opis vybraných lokalít .....	47
4.3	Referenčné dáta .....	50
4.4	Dáta pre upresnenie hraníc lokalít.....	52
4.5	Dáta z družíc.....	53
4.5.1	Landsat-8.....	54
4.5.2	Sentinel-2 .....	55
4.5.3	Použité scény.....	56
5	Postup .....	57
5.1	Predspracovanie dát .....	57
5.1.1	Georeferencovanie.....	57
5.1.2	Mozaikovanie scén.....	58
5.1.3	Formátovanie údajov .....	58
5.1.4	Príprava údajov pre použitie funkcie <i>RandomForestClassifier</i> .....	59
5.2	Klasifikačný proces .....	59
5.2.1	Nastavenie parametrov funkcie <i>RandomForestClassifier</i> .....	59
5.2.2	Nastavenie parametrov RF podľa literatúry.....	61
5.2.3	Overenie nastavenie parametrov RF .....	63
5.2.4	Opis algoritmu – programu použitý na klasifikáciu.....	70
6	Výsledky .....	77
6.1	Vyhodnotenie výsledkov .....	77
6.2	Dosiahnutá klasifikačná presnosť jednotlivých scén.....	79
6.3	Dosiahnutá presnosť na základe veľkosti pixela .....	79
6.4	Dosiahnutá klasifikačná presnosť na základe druhu stromov .....	80
6.5	Vplyv kanálov na klasifikáciu .....	81
6.6	Chybové odchýlky v identifikácii druhov stromov .....	82
6.7	Mapové výstupy .....	82
7	Diskusia .....	84
7.1	Odpovede na výskumné otázky .....	84
7.2	Porovnanie dosiahnutých výsledkov s výsledkami v iných prácach.....	85
7.3	Možnosti dosiahnutia vyššej klasifikačnej presnosti.....	86
8	Záver.....	88
9	Bibliografia .....	89
	Príloha 1: Rešerš a georeferencovanie mapových podkladov .....	94

1.1.	Tabuľkový prehľad rešerše 1.....	94
1.2.	Tabuľkový prehľad rešerše 2.....	95
1.3.	Spektrálne rozlíšenie senzorov satelitov.....	96
1.4.	Priestorové rozlíšenie senzorov satelitov .....	97
1.5.	Pásma družíc 400 – 1040 nm .....	98
1.6.	Pásma družíc 1000 - 3000 nm .....	98
1.7.	Vlastnosti lokalít a georeferencovania mapových podkladov .....	99
1.8.	Prehľad scén.....	101
1.9.	Rešerš 33 prác.....	102
	Príloha 2: Chybové matice klasifikácie snímok .....	103
	Príloha 3: Mapové výstupy klasifikácie lesných porastov .....	106
	Príloha 4: Zdrojové kódy programov.....	129
4.1.	Program – Overenie správnosti nastavenie parametra $n\_estimators$ .....	129
4.2.	Program – Overenie veľkosti trénovacej množiny na výslednú presnosť.....	131
4.3.	Program – použitý na výpočet .....	132

## Použité skratky

ARVI	– Atmospherically resistant vegetation index
B	– Blue (band) <i>modrý kanál</i>
C	– Coastal (band) <i>pobrežný kanál</i>
ČHMÚ	– Český hydrometeorologický ústav (office)
ČUZK	– Český úrad zeměměřický a katastrální (office)
DCM	– Digital canopy model (model) <i>digitálny model súvislej vrstvy korún stromov</i>
DPZ	– Remote Sensing (science) <i>Diaľkový prieskum zeme</i>
DTM	– Digital terrain model (model) <i>digitálny model terénu</i>
EVI	– Enhance vegetation index
G	– Green (band) <i>zelený kanál</i>
HMS	– Multispectral images (data) <i>multispektrálne snímky</i>
L-8	– Landsat-8 (satellite) <i>družica</i>
LDA	– Linear Discriminant Analysis (method) <i>Lineárne diskriminačné analýzy</i>
LVS	– Forest vegetation zone – <i>lesný vegetačný stupeň</i>
MD	– Mahalanobis distance (method) <i>Mahalanobisová vzdialenosť</i>
MDA	– Mean decrease in accuracy (method) <i>metóda</i>
MDFC	– Mean discriminant function coefficients (method) <i>metóda</i>
MDG	– Mean decrease in Gini (method) <i>Stredná hodnota v Giniho koeficientu</i>
NASA	– National Aeronautics and Space Administration (office) <i>Národná správa letectva a vesmíru</i>
NDVI	– Normalized difference vegetation index – <i>Normalizovaný diferenčný vegetačný index</i>
NIR1	– Near Infrared (band) <i>blízke infračervené pásmo</i>
NLC	– <i>Národné lesnícke centrum Slovenskej republiky vo Zvolene</i> (office) <i>ústav</i>
NN	– Neural network (method) <i>neurónové siete</i>
OA	– Overall accuracy – <i>celková presnosť</i>
OBIA	– Object-based image analysis (method) <i>objektová obrazová analýza</i>
OLI	– Operational Land Imager (sensor) <i>senzor družice</i>
OKC	– Overall kappa coefficient – <i>celkový kappa koeficient</i>

PCA	– Principal Component Analysis (method) <i>metóda hlavných komponent</i>
PDA	– Penalized discriminant analysis (method of linear combinations) <i>metóda lineárnych kombinácií</i>
PT	– <i>porastné typy</i>
R	– Red (band) <i>červené pásmo</i>
RE	– Red-Edge (band) <i>krajné červené pásmo</i>
RF	– Random Forest (method) <i>metóda</i>
S-2	– Sentinel-2 (satellite) <i>družica</i>
SAM	– Spectral angle mapper (method) <i>metóda</i>
SHMU	– <i>Slovenský hydrometeorologický ústav</i> (office)
SID	– Spectral information divergence (method) <i>metóda</i>
SpecTex	– HMS spectral-spatial texture image (data) <i>spektrálno-priestorový texturálny obraz</i>
S-JTSK	– Systém jednotné trigonometrické sítě katastrální (coordination system) <i>súradnicový systém</i>
SVM	– Support vector machine (method) <i>metóda</i>
SPT	– <i>štruktúrovaný porastný typ</i>
QB	– QuickBird (satellite) <i>družica</i>
UHUL	– <i>Ústav pre hospodársku úpravu lesov Brandýs nad Labem</i> (office) <i>ústav</i>
USGS	– U.S. Geological Survey (office) <i>ústav</i>
TIRS	– Thermal Infrared Sensor (sensor) <i>senzor družice</i>
VARI	– Visible atmospherically resistant index – <i>viditeľný atmosférický rezistentný index</i>
VIs	– Vegetation index – <i>vegetačné indexy</i>
VRE	– Vegetation Red Edge (band) <i>vegetačné krajné červené pásmo</i>
VW-2	– World View-2 (satellite) <i>družica</i>
WGS84	– World Geodetic System 1984 (coordination system) <i>súradnicový systém</i>
ZM 1:10 000	– <i>Základná mapa Českej republiky 1:10 000</i> (map)

## Vysvetlivky zdrojov:

[1] – odkaz na zdroj literatúry

[a] – odkaz na zdroj dát

## Zoznam tabuliek, obrázkov a grafov

### Zoznam tabuliek:

Tabuľka 1: Prehľad úrovne priestorového rozlíšenia družíc spomínaných

Tabuľka 2: Deň nástupu (medián a 5 % a 95 % kvantil) a dĺžka trvania fenofáz dubových a bukových porastov za obdobie 2013 – 2014

Tabuľka 3: Dostupnosť snímok za obdobie 2013 – 2014

Tabuľka 4: Porovnanie názvov kľúčových parametrov RF pre túto prácu medzi jazykom "R" a Python

Tabuľka 5: Druhové zloženie lesov v [ha] a [%] z celkovej plochy porastnej pôdy

Tabuľka 6: Prehľad lesných vegetačných stupňov a ich zastúpenie v lesoch ČR

Tabuľka 7: Porastné typy (PT)

Tabuľka 8: Charakter zmiešania

Tabuľka 9: Základný prehľad informácií o scénach

Tabuľka 10: Vplyv parametra `n_estimators` na výpočet presnosti [%]

Tabuľka 11: Vplyv parametra `creation` na výpočet presnosti [%]

Tabuľka 12: Vplyv zmeny parametra `min_samples_leaf` na presnosť klasifikácie [%]

Tabuľka 13: Vplyv zmeny parametra `min_samples_split` na presnosť klasifikácie [%]

Tabuľka 14: Závislosť veľkosti trénovacej množiny na výslednú presnosť

Tabuľka 15: Celková a užívateľská presnosť klasifikácii jednotlivých scén

Tabuľka 16: Vyhodnotenie celkovej klasifikačnej presnosti podľa použitých scén

Tabuľka 17: Dosiahnutá užívateľská presnosť na základe druhu stromov

Tabuľka 18: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu „dub“

Tabuľka 19: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu „buk“

Tabuľka 20: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu „smrek“

Tabuľka 21: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu "borovica"

Tabuľka 22: Vplyv kanálu na klasifikáciu

Tabuľka 23: Chybové odchýlky v klasifikácii druhov drevín

Tabuľka 24: Chybová matica scény L8\_L1\_160624

Tabuľka 25: Chybová matica scény L8\_L1\_160827  
Tabuľka 26: Chybová matica scény L8\_L1\_170830  
Tabuľka 27: Chybová matica scény L8\_L2\_160624  
Tabuľka 28: Chybová matica scény L8\_L2\_160827  
Tabuľka 29: Chybová matica scény L8\_L2\_170830  
Tabuľka 30: Chybová matica scény S2\_L1\_150704  
Tabuľka 31: Chybová matica scény S2\_L1\_150813  
Tabuľka 32: Chybová matica scény S2\_L1\_160827

### **Zoznam obrázkov:**

Obrázok 1: Rozdiely vo veľkosti smrekových porastov v porovnaní údajov z LhpoMap a z google maps v Lokalite Přísečnice  
Obrázok 2: Fenologické stanice na Slovensku  
Obrázok 3: Priebeh jarných fenofáz dubových porastov v období rokov 2000 – 2014  
Obrázok 4: Priebeh jarných fenofáz bukových porastov v období rokov 2000 – 2014  
Obrázok 5: Zobrazenie druhovej skladby lesa za rok 2014 v mapovom portály LhPoMap  
Obrázok 6: Základný princíp RF – zdroj: Random Forest Introduction  
Obrázok 7: Príklad vytvárania rozhodovacieho stromu – zdroj: Random Forest Introduction  
Obrázok 8: Proces klasifikovania atribútov – zdroj: Random Forest Introduction  
Obrázok 9: Lesné vegetačné stupne v ČR  
Obrázok 10: Dátový sklad Lho  
Obrázok 11: Úprava hraníc lokality Přísečnice podľa leteckých snímok z google maps  
Obrázok 12: Immitzer M. a kol. (2012) Figure 5  
Obrázok 13: Test korelácie kanálov  
Obrázok 14: Použitie Random Forest v aplikácii "ArcGis Pro"

### **Zoznam grafov:**

Graf č. 1: Závislosť chyby od hodnoty  $n_{estimators}$  (zdroj )  
Graf č. 2: Vplyv parametra  $n_{estimators}$  na výpočet presnosti  
Graf č. 3: Vplyv parametra  $n_{estimators}$  na výpočet presnosti údajov zo scény L8\_L1\_160827



Graf č. 4: Vplyv parametra n\_estimators na výpočet presnosti údajov zo scény  
S2\_L1\_160827

Graf č. 5: Vplyv parametra creation na výpočet presnosti

Graf č. 6: Vplyv zmeny parametra min\_samples\_leaf na presnosť

Graf č. 7: Vplyv zmeny parametra min\_samples\_split na presnosť

Graf č. 8: Závislosť veľkosti trénovacej množiny na výslednú presnosť

Graf č. 9: Prehľad dosiahnutej klasifikačnej presnosti

Graf č. 10: Vplyv kanálu na klasifikáciu

### **Zoznam mapových príloh:**

(viď Príloha 3)

1. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 1 a 2)
2. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 3, 4 a 5)
3. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 6 a 7)
4. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 8, 9 a 10)
5. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 11 a 12)
6. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 13 a 14)
7. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 15 a 16)
8. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 17 a 18)
9. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 19 a 20)
10. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 21 a 22)
11. Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016 (Lokality č. 23 a 24)
12. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 1 a 2)
13. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 3, 4 a 5)
14. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 6 a 7)
15. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 8, 9 a 10)
16. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 11 a 12)
17. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 13 a 14)
18. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 15 a 16)
19. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 17 a 18)
20. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 19 a 20)
21. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 21 a 22)
22. Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015 (Lokality č. 23 a 24)

# 1 ÚVOD

Táto kapitola je úvodom do problematiky tejto diplomovej práce. Píše sa v nej o využití diaľkového prieskumu Zeme v lesníctve a o druhovej skladbe lesa. Tiež sa v tejto kapitole vymedzujú oblasti výskumnej témy diplomovej práce a definujú sa základné výskumné otázky, ktoré pripadajú v úvahu pre spracovanie v diplomovej práci.

## 1.1 Využitie diaľkového prieskumu zeme v lesníctve

V lesníctve sa používajú metódy diaľkového prieskumu zeme (ďalej DPZ) pre rýchle získanie informácií o stave lesa, bez potreby detailného plošného terénneho prieskumu. Metódy DPZ sa v lesníctve využívajú na skúmanie druhového zloženia, vekového zloženia a zdravotného stavu lesov. Taktiež sa využívajú na monitorovanie stavu lesov po rôznych prírodných katastrofách ako sú napr. požiare a veterné smršte.

Rozvoju metód na celosvetovej úrovni prispievajú okrem pribúdajúcich algoritmov aj spôsoby získavania dát. V posledných rokoch sa k leteckému snímkovaniu, k radarovým a multispektrálnym družiciam pridali aj hyperspektrálne družice a letecké laserové skenovanie. Tiež pribudol aj nový druh nosičov pre senzory. K lietadlám a družiciam sa pridali drony.

Významný vplyv na neustále zvyšujúcu presnosť klasifikácie má aj vývoj stále výkonnejších senzorov, vďaka ktorým sa vytvárajú dáta s neustále väčším priestorovým rozlíšením.

V súčasnosti z voľne dostupných snímok sú na skúmanie vegetácie najvhodnejšie snímky z optických multispektrálnych družíc Landsat-8 a Sentinel-2. Landsat-8 sníma v 11 pásmach. Panchromatické pásmo má priestorové rozlíšenie 15 m a všetky ostatné pásma majú priestorové rozlíšenie 30 m. Družica Sentinel-2 sníma v 12 pásmach a snímky majú v jednotlivých pásmach priestorové rozlíšenie 10 až 60 m.

Z komerčných družíc najväčšie priestorové rozlíšenie majú WorldView-2, Quick-Bird, Ikonos a Spot. WorldView-2 má priestorové rozlíšenie panchromatického pásma 0,46 m a multispektrálnych pásiem 1,84 m. Quick-Bird má priestorové rozlíšenie panchromatického pásma 0,61 m. Priestorové rozlíšenie multispektrálnych pásiem pri snímkach veľkosti 450 x 450 km je 2,44 m a pri snímkach veľkosti 300 x 300 km 1,63 m. Ďalšími komerčnými družicami sú Ikonos s priestorovým rozlíšením snímok 1 - 4 m a Spot s rozlíšením 1,5 - 6 m.

Od zberu dát v rôznych podobách s rôznou rozlišovacou schopnosťou sa následne odvíja predspracovanie dát a následne vývoj metód. V prípade optických družicových dát si autor volí klasifikačné príznaky a to spektrálne, priestorové, texturálne a časové. Volí si medzi pixelovou a objektovou klasifikáciou. Taktiež je potrebné si zvoliť klasifikačný algoritmus riadenej klasifikácie, neriadenej klasifikácie alebo algoritmus strojového učenia napr. neural network (ďalej NN), Random Forest (ďalej RF), atď.

V súčasnosti sa vo svete používajú najčastejšie metódy napr. Random Forest, Linear Discriminant Analysis (ďalej LDA) a Valleyfollowing Approach.

V Českej republike v Ústave pre hospodársku úpravu lesov (ďalej UHUL) a na Slovensku v Národnom lesníckom centre (ďalej NLC) sa v DPZ v lesníctve využívajú predovšetkým voľne dostupné dáta z optických družíc Landsat a Sentinel, snímky z leteckého snímkovania a snímky z dronov. Tiež sa vo veľkej miere využívajú snímky z družíc Terra a Aqua, ktoré sú vybavené spektrorádiometrom, ktorý sníma v 36 kanáloch s rozlišovacou schopnosťou 250 m až 1 km. Snímky z komerčných družíc napr. Spot a Ikonos sa využívajú zriedkavejšie, ale ich výhodou je, že produkujú snímky s väčším priestorovým rozlíšením.

V posledných rokoch boli vydané na Slovensku dva významné diela v oblasti využívania satelitov v lesníctve. Prvé dielo nesie názov *Meniace sa Slovensko očami satelitov* [1], v ktorom autori analyzovali viacero zmien z temporálneho hľadiska. Napr. zmeny atmosférického počasia a príkonu slnečného žiarenia, zmeny stavov kozmického počasia, zmeny v poľnohospodárskej krajine, zmeny v lesnej krajine, zmeny v ionosfére, hydrosfére a deformácie zemského povrchu z analýz navigačných a gravimetrických satelitov.

Druhé dielo s názvom *Satelity v službách lesa* [2] je akosi zbierkou článkov, v ktorej autori využili rôzne metódy pre výskum v rôznych oblastiach týkajúcich sa lesného porastu. V článkoch autori rozoberali témy: klasifikácia poškodenia lesov Slovenska, porovnanie priebehu rozpadu smrečín na maloplošných chránených územiach a v hospodárskych lesoch, posudzovanie štruktúry a stability horských lesov s podporou DPZ, tvorba údajovej bázy a modelovanie fenológie lesných porastov, validácia nástupu fenologických udalostí bukových porastov, fenológiu dubových a bukových porastov v období 2000 – 2014, vplyv prízemnej vegetácie a podrastu na priebeh fenologickej krivky bukových porastov, stres dubových a bukových porastov suchom a teplom s použitím experimentálnych satelitných snímok Modis, teoretický koncept modelovania primárnej produkcie lesných porastov, stanovenie ročnej primárnej produkcie uhlíka bukových a dubových porastov na Slovensku a informačný systém na hodnotenie stavu lesa zo satelitných snímok.

## 1.2 Druhovú skladbu lesa

Les je možné deliť podľa viacerých kritérií. Vďaka týmto kritériám správcovia lesov majú lepší prehľad o lese, ktorý spravujú a vedľa s ním efektívnejšie hospodáriť. O delení lesov z rôznych hľadísk sa píše napr. v dokumente *Zpráva o stavu lesa a lesního hospodárství České republiky v roce 2016* [3].

Les sa podľa druhového zloženia delí na rovnírodý a rôznorodý les. Podľa veku sa les delí na rovnoveký a rôznoveký les. V lesníctve sa používa pojem vekový stupeň, ktorý je na porastových mapách farebne odlišný. Podľa prevládajúceho využitia, funkcie lesa sa les delí na hospodársky, ochranný a les osobitného určenia. Z pohľadu

vlastníctva sa lesy delia na štátne, súkromné, družstevné, cirkevné, mestské, obecné, neurčiteľné (nemajú určené vlastníctvo). Lesy tiež môžu vlastníť urbáre (urbár - druh pozemkovej knihy), spoločnosti a komposesoráty.

Subjektom spravujúcim lesy pre lepší prehľad druhového zloženia lesov, vekovej štruktúry lesov, veľkosti lesných plôch a ich polohu slúžia lesnícke mapy a hospodárske knihy. Plochy lesov sú od najväčších po najmenšie hierarchicky usporiadané: oddelenie, diel, porastová skupina a etáž. Jednotky plochy sú udávané v hektároch.

Les môže mať rôzne druhové zloženie. Zmiešaný les je prirodzený a preto pokrýva väčšinu územia lesných pozemkov v Českej republike. V hospodárskych lesoch veľa plôch zaberajú monokultúry, pretože boli vysádzané umelo. Ale aj napriek tomu v hospodárskych lesoch tvoria monokultúry menšinu, pretože čelia náletovým druhom drevín a tiež semennej banke pôvodného lesa, ktorá je uložená v pôde.

### **1.3 Vymedzenie oblasti výskumnej témy diplomovej práce**

Cieľom práce je nájsť vhodnú metodiku, ktorá dokáže z dát optických družíc, čo najpresnejšie klasifikovať druhové zloženie drevín na území lesov.

### **1.4 Základný popis výskumných otázok, ktoré pripadajú v úvahu pre spracovanie v diplomovej práci**

1. Dáta z ktorej multispektrálnej družice, vzhľadom na vlastností pasiem, je vhodnejšie použiť pri skúmaní druhového zloženia lesov na území ČR?
2. Použitie ktorého klasifikátora a nastavením akých hodnôt jeho parametrov je najvhodnejšie pre klasifikáciu drevín na území ČR?
3. Je možné na základe súčasných poznatkov dosiahnuť celkovú klasifikačnú presnosť nad 80 %?
4. Pri klasifikácii sa ktoré druhy drevín (dub, buk, smrek, borovica) najviac zamieňajú?

## 2 ANALÝZA PUBLIKOVANÝCH PRÁC – METÓD

V rámci rešerši bolo preskúmaných okolo 50 prác od vedcov z celého sveta, ktoré boli zamerané na skúmanie druhového zloženia lesov pomocou dát z optických družíc. V rešeršiach sa mnohokrát používali, ako zdrojové dáta kombinácie dát z multispektrálnych družíc s napr. leteckými snímkami, dát z hyperspektrálnych družíc, leteckého laserového skenovania, radarových družíc a z mnohých ďalších zdrojov.

V rámci rešerše bolo v tejto práci podrobne analyzovaných 10 literárnych zdrojov (viď. *Prílohy 1.1 a 1.2*) a to ich obsah po stránke využitia druhu družice, kanálov, metódy, druhy drevín a výsledky klasifikácie podľa jednotlivých metód.

Následne bola popísaná metodika výberu čo najvhodnejšieho klasifikátora a výberu čo najvhodnejších hodnôt pre vstupné parametre vzhľadom na dané okolnosti (druhové zloženie lesov, vlastnosti kanálov družíc, ročné obdobie, atď.).

Osobitná podkapitola bola venovaná rozhodovaniu medzi použitím jazyka „R“ a Python.

### 2.1 Metodika práce

Pre dosiahnutie čo najvyššej klasifikačnej presnosti lesného porastu bolo potrebné brať do úvahy množstvo vstupujúcich vlastností metodiky.

V prvom rade bolo potrebné určiť čo najvhodnejší postup výberu vlastností metodiky a ich hodnôt pre budúcu klasifikáciu. Pre dosiahnutie čo najvyššej presnosti klasifikácie bol zvolený nasledovný postup hodnotenia vlastností a výberu ich hodnôt.

Výber vlastností metodiky práce:

1. Druhy drevín
2. Vlastnosti pásiem družíc
3. Výsledky klasifikácie vybraných druhov drevín
4. Ročné obdobie
5. Klasifikátor
6. Parametre klasifikátora
7. Programovací jazyk

### **Vysvetlenie pojmov:**

*metodika* – myslený postup krokov pomocou, ktorých sa vyberú klasifikačné triedy, družice, kanály družíc, ročné obdobie v ktorom bude druhové zloženie lesov skúmané, klasifikátor a parametre klasifikátora

*metóda* – myslený synonymum slova klasifikačná metóda, alebo klasifikátor

*vlastnosť* – parameter pre metodiku práce

*parameter* – vstupujúca vlastnosť pre konkrétnu klasifikačnú metódu (klasifikátor)

*atribút* – vlastnosť (stĺpec) v atribútovej tabuľke shapefile vrstvy

## **2.2 Rešerš literatúry – 10 kľúčových štúdií**

V tejto podkapitole sa nachádzajú rešerše prác k danej úlohe, v ktorých sa nachádza opis územia výskumu, počet skúmaných druhov drevín, použité družice a ich kanály a použitá metóda.

Týchto 10 prác bolo vybraných z pomedzi všetkých prác z dôvodov: dosiahnutej vysokej celkovej presnosti klasifikácie (skrt. OA, angl. overall accuracy of classification), výskum stál predovšetkým na dátach multispektrálnych družíc a proces výskumu bol v článku dostatočne podrobne popísaný.

*Immitzer M. a kol. (2012) [4]* si vybrali územie výskumu v Rakúsku v Burgenlande. Jednalo sa o zmiešaný rôznorodý les. Terén sa nachádzal v predhorí v nadmorskej výške 290 až 670 m.n.m. Bolo skúmaných 10 druhov drevín.

Čo sa týka využitia kanálov autori uvádzajú, že pre vysokú koreláciu kanálov zo satelitu WV-2 (World View-2) najväčšie množstvo informácie nesie kombinácia kanálov C (Coastal), G (Green), R (Red), NIR1 (Near Infrared). Ďalej autori v práci tvrdia, že kanál RE (Red Edge) dokáže odlišovať zdravé stromy od stromov, ktoré postihli nejaké choroby a odlišuje stromy podľa veku a aj druhu. Vo viditeľnom pásme najväčšie rozdiely vidieť v zelenom a žltom pásme. Avšak všeobecne najväčšie rozdiely sú v NIR pásmach. Buk sa viac odlišuje v niektorých pásmach od ostatných drevín, ako v iných pásmach a to výrazne v pásmach NIR1 a NIR2. V pásme NIR2 na rozdiel od pásma NIR1 je snímaný povrch v o niečo vyššej vlnovej dĺžke a je odolnejší voči atmosférickým vplyvom a preto je vhodnejšie na skúmanie biomasy, podľa autorov tohto článku.

Z pohľadu vhodnosti využitia metód na základe výsledkov klasifikácie autori článku tvrdia, že metódy RF (Random Forest) a LDA (Linear Discriminant Analysis) sú vhodnejšie ako metódy MDA (mean decrease in accuracy), MDG (mean decrease in Gini), MDFC (mean discriminant function coefficients) a Wilks' Lambda pre skúmanie druhového zloženia lesov. Na dáta aplikovali iteratívne 500-krát metódu LDA. V prípade použitia LDA bola dosiahnutá vyššia klasifikačná presnosť pri použití 4 kanálov, ako v prípade použitia 8 kanálov. V ich práci porovnávajú dosiahnuté výsledky s prácami v ktorých boli použité iné klasifikátory: MLC (Maximum Likelihood), MD

(Mahalanobis Distance), SAM (Spectral Angle Mapper), SVM (Support Vector Machine), SID (Spectral Information Divergence), NN (Neural Network).

V tej istej práci celková klasifikačná presnosť s použitím RF (pri 20 iteráciách, 10 druhov drevín, objektovej klasifikácii) najlepšie vyšla pri kombinácii kanálov C, G, R, NIR1 a to 0,792. Najvyššia celková klasifikačná presnosť v prípade použitia LDA vyšla pri kombinácii kanálov B (Blue band), G, R, NIR a to 0,790. Avšak podľa spektrálnych príznakov opísaných v tomto článku autori považujú kanály NIR1, NIR2, RE, G, ako vhodné kanály pri použití vhodných metód na odlišenie vybraných druhov drevín (dub, buk, smrek a borovica). Najvyššia užívateľská presnosť vyšla pri objektovej klasifikácii 4 druhov drevín (8 kanálov, Worldview-2). Užívateľská aj výrobná klasifikačná presnosť bola vyššia, ako 90% aj u listnatých aj u ihličnatých druhov drevín. Správnosť klasifikácie dosiahla celkovú presnosť 95,9%, s Kappa indexom 0,945. Výrobná presnosť buka vyšla 98,4% (užívateľská 98%), duba 97,4%, smreka 94,9% (užívateľská 93,7%) a borovice 93,4% (užívateľská 94,6 %).

*Verlič a kol. (2014)* [5] sa vo svojom výskume zamerali na zmiešaný heterogénny mestský prírodný les Tivoli, Rožnik a Šišenski hrib (kopec) v Ľubľane v Slovinsku. Územie výskumu je kompaktné s rozlohou 459 ha a s nadmorskou výškou od 295 do 429 m.n.m. Bolo skúmaných 5 druhov drevín.

Autori tvrdia, že im vyšla vyššia celková presnosť klasifikácie použitím 8 kanálov (WV-2), ako len použitím 4 pôvodných kanálov (WV-2). V práci píše, že podľa Mean spectral signatures a podľa výsledkov z box-and-whiskers plots, by z viacerých možných kombinácií kanálov klasifikácia mala dosiahnuť najvyššiu celkovú presnosť použitím kombinácie kanálov RE, NIR1, NIR2 a G. Avšak z aplikácie metódy PCA (Principal Component Analysis) na 8 pôvodných kanáloch WV-2 vyšlo, že najvyššiu informačnú hodnotu nesú kanály RE, NIR1 a NIR2. Najvyššie množstvo informačnej hodnoty získala prvá komponenta PCA z kanálov NIR1 a NIR2.

Autori v práci porovnávali objektovú klasifikáciu (skrt. OBIA, angl. object-based image analysis) a pixelovú klasifikáciu (angl. pixel-based image analysis) s použitím klasifikátora SVM. V prípade OBIA išlo o dva hlavné procesy a to o segmentáciu a klasifikáciu. V prípade pixelovej klasifikácie sa využilo 8 pôvodných snímok pretransformovaných analýzou hlavných komponent (skrt. PCA, angl. Principal Component Analysis) v kombinácii s dátami leteckého laserového skenovania a RE NDVI (Band Red-Edge normalized difference vegetation index).

Celková presnosť klasifikácie vyšla 58% a Kappa koeficient 0,431. Procesná presnosť duba vyšla 70%, buka len 38%, smreku 80% a borovice 50%. Užívateľská presnosť duba vyšla 47%, buka 64%, smreka 69% a borovice 59%.

*Chinsu Lin a kol. (2015)* [6] si vybrali ako územie výskumu dažďový les v centrálnej časti Taiwanu. Bolo skúmaných 40 druhov drevín.

Autori použili snímky z družice QB (QuickBird) a pomocou nich vypočítali vegetačné indexy ARVI (atmospherically resistant vegetation index), VARI (visible atmospherically resistant index) a EVI (enhance vegetation index). Predpokladajú, že najvhodnejšie sú modré a zelené pásma na skúmanie druhového zloženia drevín. Klasifikácia z VIs im vyšla s nižšou klasifikačnou presnosťou, ako klasifikácia HMS snímok, pretože každý z VIs používa iba časť spektrálnych príznakov z HMS.

Autori urobili v práci pixelovú klasifikáciu metódou Maximum Likelihood Classification a následne na tie isté dáta aplikovali oddelene metódy MD, SAM, SVM, SID a NN.

Pri klasifikácii 40 druhov drevín vyšiel celkový kappa koeficient 0,58. Po klasifikácii 5 VIs vyšiel celkový kappa koeficient len 0,48. V tomto článku autori využili okrem spektrálneho príznaku aj texturálny príznak, vďaka čomu autori dosiahli pri skombinovaní HMS, VIs, SpecTex (HMS spectral-spatial texture image) OKC viac, ako 0,98. Takže je možné vidieť, že texturálny príznak môže zvýšiť hodnotu OKC.

*Carleer A. & Wolff E. (2004) [7]* si vybrali ako územie výskumu časť Brussels-Capital Sonian Forest. Skúmaných bolo 11 druhov drevín.

V tejto práci autorom vyšla najvyššia celková presnosť klasifikácie kombináciou pôvodných kanálov (R, G, B, NIR), NDVI a prvou komponentou PCA.

Autori vo svojej práci urobili pixelovú klasifikáciu s použitím klasifikátora MLC na kanáloch pôvodných HMS, NDVI a PCA. S cieľom znížiť vnútornú odchýlku tried aplikovali mean filter do všetkých pôvodných aj nových kanálov. Aplikácia mean filtra na kanále obraz vyhladí a lepšie vidieť rozdiely medzi triedami, ale horšie v rámci tried. Na elimináciu salt-and-pepper efektu bol použitý modal filter. Klasifikácia filtrovaného obrazu vyšla lepšie ako nefiltrovaného obrazu.

Celková presnosť klasifikácie vyšla 85,8 % a Kappa index 0,838. U duba vyšla užívateľská presnosť klasifikácie 87,8 %. U mladého buka vyšla užívateľská presnosť 97,7 %, u starého buka 100%, u purpurového buka 86,2%, u borovici lesnej 63,9 % a u borovice čiernej 43,5 %. Autori na záver dodávajú, že dub a starší buk sa zle odlišujú v neriadenej klasifikácii kvôli ich spektrálnej podobnosti.

*Krahwinkler & Rossmann (2010) [8]* si vybrali územie výskumu v lokalite severo-východne od mesta Arnsberg v Nemecku. Skúmaných bolo 6 druhov drevín.

Na základe výsledkov v práci autori tvrdia, že pri leteckých a družicových SPOT snímkach sú pre skúmanie buku najvhodnejšie nové kanály IR-R, IR-G, IR-B. Pre klasifikáciu dubu je najvhodnejšie použiť IR-G. Pre klasifikáciu smreku je najvhodnejšie použiť SW-IR a SW-R.

Autori vo svojej práci poukazujú na PDA (penalized discriminant analysis). PDA produkuje lineárne kombinácie kanálov, to poukazuje na to, ako pomocou komponent predictor vector contribute zdefinovať diskriminačné pravidlo. V práci ako klasifikačnú



metódu použili rozhodovací strom založený na pomere DN hodnôt jednotlivých kanálov.

Na oddelenie území bez lesa a lesa s tieňom od územia pokrytým lesom bol využitý DCM (digital canopy model), celkový jas a rozdiely IR-R, G-B.

Celková presnosť klasifikácie vyšla 78 %. Klasifikačná presnosť duba vyšla 74,3 %, buka 79,7 % a smreka 79,1 %.

*Immitzer M. a kol. (2016)* [9] publikovali svoj článok v nemčine, prezentovali ho však aj v angličtine v podobe posteru na Living Planet Symposium 2016 v Prahe *Immitzer M. a kol. (2016)*. [10]

Autori svoj výskum robili na území Ebersberg a Altötting, lokalizované východne od Mníchova v Nemecku. Les bol rôznorodý s dominanciou smreku. Skúmaných bolo 12 druhov drevín.

Autori článku na základe výsledkov klasifikácie v práci tvrdia, že dáta z družíc L-8 (Landsat-8) a S-2 (Sentinel-2) nemajú dostatočné priestorové rozlíšenie na skúmanie druhového zloženia rôznorodého lesa.

Autori vo svojej práci využili na dáta individuálnu segmentáciu použitím automatizovaných procedúr a mean shift algoritmu. Tiež využili metódu RF.

U dát z WorldView-2 vyšla OA 74,4% a Kappa index 0,691.

Pre dub vyšla spracovateľská presnosť 0,559 a užívateľská presnosť 0,589.

Pre buk vyšla spracovateľská presnosť 0,785 a užívateľská presnosť 0,664.

Pre smrek vyšla spracovateľská presnosť 94,5 % a užívateľská presnosť 86,4 %.

Pre borovicu vyšla spracovateľská presnosť 84 % a užívateľská presnosť 89,4 %.

U dát zo Sentinel-2 vyšla OA 68,1 % a Kappa index 0,611.

Pre dub vyšla spracovateľská presnosť 0,281 a užívateľská presnosť 0,450.

Pre buk vyšla spracovateľská presnosť 0,727 a užívateľská presnosť 0,706.

Pre smrek vyšla spracovateľská presnosť 90,9 % a užívateľská presnosť 77,8 %.

Pre borovicu vyšla spracovateľská presnosť 23,8 % a užívateľská presnosť 71,4 %.

U dát z Landsat-8 vyšla OA 49,4 % a Kappa index 0,380.

Pre dub vyšla spracovateľská presnosť 0,194 a užívateľská presnosť 0,261.

Pre buk vyšla spracovateľská presnosť 0,621 a užívateľská presnosť 0,526.

Pre smrek vyšla spracovateľská presnosť 64,9 % a užívateľská presnosť 56,2 %.

Pre borovicu vyšla spracovateľská presnosť 19 % a užívateľská presnosť 36,4 %.

*Bucha T. (2007)* [11] si ako výskumné územie vybral celé zalesnené územie Slovenska. Skúmaných bolo 15 druhov drevín.

Autor vo svojej práci tvrdí, že lepšie výsledky klasifikácie dosiahol použitím nového kanálu vytvoreného z DTM (digitálny model terénu) v kombinácii s poznatkami o raste jednotlivých druhov drevín v konkrétnom rozpätí nadmorských výšok, ako z prvej komponenty PCA z pôvodných HMS kanálov.

V práci autor použil metódu MLC s Bayassovým klasifikátorom.

Správnosť klasifikácie bola preverená na súbore 412 trénovacích plôch a pohybovala sa v závislosti od scény v rozpätí od 80% do 90%.

*Sothe C. a kol. (2017) [12]* vo svojej práci skúmali druhové zloženie lesov v Brazílii. Využívali ortofoto zhotovené senzorom SAAPI a dáta z družíc Landsat-8 a Sentinel-2. V práci porovnávali výsledky klasifikácie dosiahnutými pomocou klasifikátorov RF a SVM. Táto štúdia preukázala, že snímky z družíc Sentinel-2 a Landsat-8 sú vhodné pre klasifikáciu dažďových pralesov. Procesná a užívateľská presnosť dosiahli u všetkých klasifikačných experimentov rovnakú, alebo vyššiu presnosť, ako 80%.

Autori uvádzajú, že dosiahli celkovú presnosť nad 80 % u všetkých klasifikačných experimentov. Je všeobecne predpokladané, že tak vysokú celkovú presnosť mohli dosiahnuť vďaka použitiu ortofota zo senzoru SAAPI. Pozoruhodne vysokú celkovú presnosť dosiahli vďaka použitiu aj iných príznakov, ako spektrálnych. Vďaka ortofotu autori využili texturálny priemer, rozdielnosť, kontrast a jarné spektrálne pásma. V prípade snímok z Landsat-8 priniesli texturálna metrika, multičasová informácia a vegetačné indexy očakávané zvýšenie celkovej presnosti. Avšak v prípade dát zo Sentinel-2 využitie týchto príznakov neprinieslo žiadne zvýšenie celkovej presnosti.

*Marine Le Louarn a kol. (2017) [13]* skúmali druhové zloženie stromov v mestskom prostredí v časti mesta Marseille vo Francúzsku.

Využívali dáta z bi-temporálnej družice Pléiades. Klasifikačných tried bolo zvolených 7 t.j. šesť druhov drevín a jedna trieda boli ostatné druhy drevín. Ako klasifikátory použili RF a SVM. Klasifikácia s klasifikátorom RF dosiahla celkovú presnosť klasifikácie až 98,5 %.

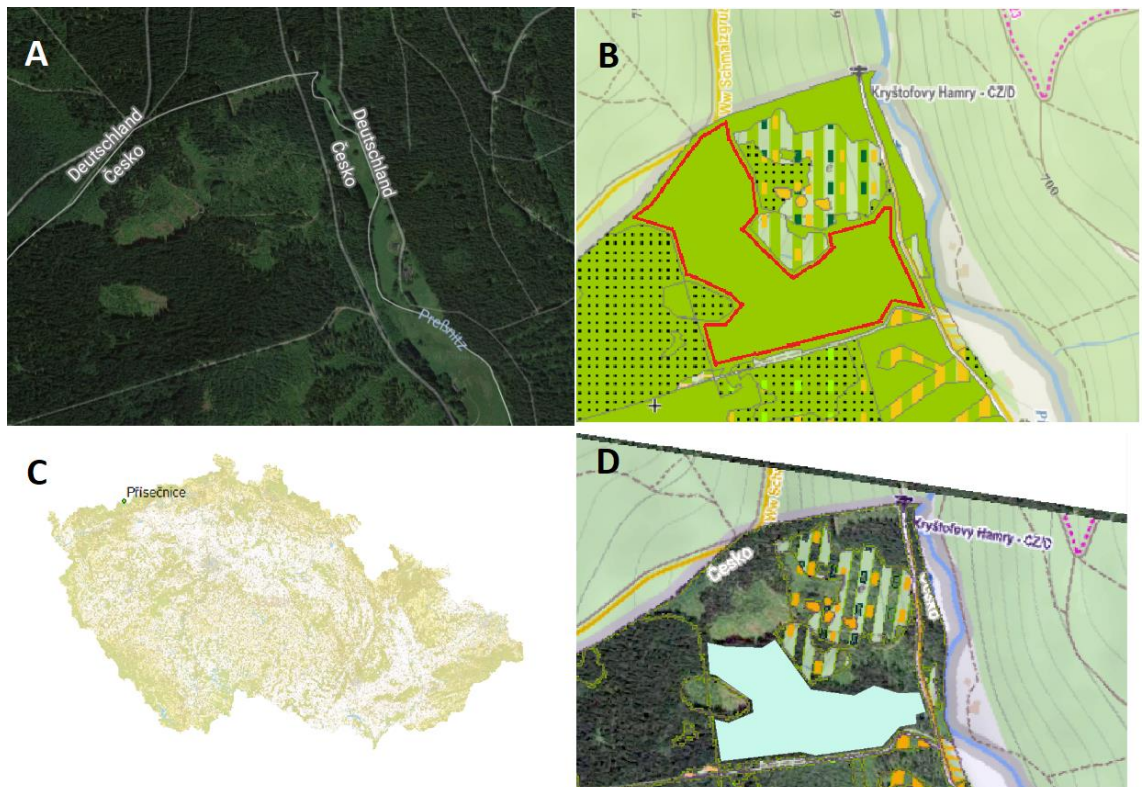
### **2.3 Druhy drevín**

Na území Českej republiky sa väčšinou nachádzajú druhovo heterogénne lesy. Na základe dát z mapovej služby LhpoMap od UHUL len určité druhy drevín tvoria samostatné homogénne kompaktné lesy. Ako pravidlo pre výber druhu dreveniny, ako triedy, bolo potrebné nájsť 6 lokalít s rozlohou každej lokality aspoň 10 ha. Do lokality s rozmermi 10 ha sa zmestí napr. 111 pixelov z družice Landsat-8, alebo 999 pixelov z družice Sentinel-2. Hodnoty sú síce zaokrúhlené smerom na dol na úrovni celých pixelov, ale výpočet neberie do úvahy tvar lokality. Takže počet pixelov v lokalitách môže byť nižší.

Lokality boli vyhľadávané na základe mapovej služby LhpoMap a overované boli na základe porovnávania s údajmi z google maps (viď. podkapitolu 5.1.1 *Georeferencovanie*). Jednotlivé lokality majú rozlohu od 10,56 do 41,59 ha. (viď. *Príloha 1.7: Vlastnosti lokalít a georeferencovania mapových podkladov*)

Iba jedna lokalita má pod 10 ha a to 7,77 ha. Ide o lokalitu smreku. Smrek je drevina, ktorá tvorí najväčšie súvislé porasty v Českej republike. Teda lokalít smreku s plochou nad 10 ha je dostatočné množstvo. Smrek teda spĺňa nami zadané kritérium pre klasifikačnú triedu. Táto lokalita je podľa mapového portálu o veľa rozsiahlejšia ako 10 ha, ale po korekcii lokality pomocou georeferencovania prinskrínu z google maps bola veľkosť tejto lokality upravená. Upravené hranice lokality kopírujú skutočnú hranicu lesa, t.j. podľa google maps. Dôvodom nekompaktnosti smrekových lesov sú prírodné katastrofy a škodcovia, ktorí výrazne znižujú veľkosti kompaktných plôch smrekových lesov.

*Obrázok 1: Rozdiely vo veľkosti smrekových porastov v porovnaní údajov z LhpoMap a z google maps v Lokalite Přísečnice*



*A – pohľad na lesný porast lokality Přísečnice (prinskrin z google maps),  
 B – pohľad na lesný porast lokality Přísečnice (prinskrin z LhpoMap),  
 C – poloha lokality Přísečnice vzhľadom k ČR (prinskrin z ArcMap),  
 D – pohľad na výber polygónu pre trénovanie a referenčné množiny lokality Přísečnice (prinskrin z ArcMap)*

*Autor: vlastné*

Hranica veľkosti plochy lokality bola stanovená na 10 ha, pretože pri hranici 5 ha by veľkosť trénovej množiny nemusela zachytiť všetky špecifiká reliéfu na ktorom je lokalita umiestená. Hranicu 15 ha by mali problém dosiahnuť lokality dubu. Takže keby hranica rozlohy lokalít bola stanovená na 15 ha, tak klasifikačných tried v tejto

práci by boli iba 3. A to pre buk, smrek a borovicu. Výber 2 tried listnatých druhov drevín a 2 tried ihličnatých druhov drevín je vhodný pre skúmanie a porovnávanie spektrálnych vlastností odrazivosti jednotlivých druhov drevín. Priebeh spektrálnej krivky každej dreveniny sa líši od priebehu spektrálnych kriviek iných drevín. Na záver je potrebné zopakovať, že lokality boli vybrané na základe dát z mapovej služby LhpoMap od UHUL a následne ich hranice a teda aj rozloha boli upravené na základe dát z google maps. Viac informácií o úpravách a definícii plochy a hraníc lokalít je možné nájsť v podkapitole 5.1.1 *Georeferencovanie*.

Dostatočne veľké lokality druhovo homogénnych lesov tvoria duby, buky, smrek a borovice. Ostatné druhy drevín nespĺnili požadované podmienky.

Preto do rešerše boli vybraté práce, v ktorých sa nachádzajú tieto 4 skúmané dreveniny. Vybrané dreveniny sú dub letný (*Quercus robur*), buk lesný (*Fagus sylvatica*), borovica sosnová (*Pinus silvestries*), smrek obyčajný (*Picea abies*). Prác ktoré obsahovali vybrané druhy drevín bolo do rešerši vybraných 7 (viď *Príloha 1.1*). Do rešerše tiež boli zaradené 3 práce, v ktorých neboli klasifikované vybrané druhy drevín (viď *Príloha 1.1*). Tieto 3 práce sú hodnotné pre hodnoty iných vlastností, ktoré boli v práci popísané nižšie.

*Immitzer M. a kol.* [4] najprv urobil vo svojej práci klasifikáciu 4 druhov drevín a následne 8 druhov drevín. Po zhodnotení výsledkov klasifikácie tvrdí, že čím je menší počet druhov drevín klasifikovaných, tým klasifikácia dosiahne vyššej klasifikačnej presnosti. Ten istý princíp potvrdzuje pohľad na tabuľkový prehľad podrobnej rešerše *Immitzera M. a kol.* [4] v jeho práci, v ktorej sa nachádza 33 prác (viď *Príloha 1.9*).

Rovnaký názor má aj *Chinsu Lin a kol.* [6], ktorí robili klasifikáciu 40 druhov drevín. Vo svojej práci tvrdia, že klasifikácia s takýmto množstvom tried nie je vhodná, pretože vďaka spektrálnej podobnosti niektorých druhov drevín, niektoré druhy drevín dosahujú nízku klasifikačnú presnosť. Autori odporúčajú buď robiť klasifikáciu na vyššej hierarchickej úrovni napr. na úrovni čeľade (lat. familia), čo by znížilo počet klasifikačných tried. Alebo druhou možnosťou pre dosiahnutie vyššej presnosti klasifikácie je klasifikovať územie po menších častiach a tým zabezpečiť zníženie počtu druhov drevín.

Toto bol dôvod prečo v tejto práci pre klasifikáciu boli vybrané 4 druhy drevín. Viac o tomto dôvode je popísané v časti práce o pozemných dátach a vlastnostiach výskumného územia v podkapitole 5.1.1 *Georeferencovanie*.

## 2.4 Pásma družíc a ich vlastnosti

Každá multispektrálna družica sníma Zemský povrch inak, ako ostatné družice. Dôvodom je konštrukcia a nastavenie vlastností senzora.

Družice sa od seba líšia v mnohých hodnotách vlastností: počtom pásiem, rôznou šírkou pásiem, rôznou dolnou a hornou hranicou pásiem, rôznym stredom

pásma a rôznym priestorovým rozlíšením pásma (viď *Prílohu 1.3: Spektrálne rozlíšenie senzorov satelitov* a *Prílohu 1.4: Priestorové rozlíšenie senzorov satelitov*).

Pre klasifikáciu snímok sa najčastejšie používajú pásma ležiace vo viditeľnej časti spektra a v blízko infračervenej časti spektra.

Z dôvodu, že pre túto prácu boli nakoniec dostupné len dáta z misií Landsat a Copernicus, tak sa rozhodovalo o tom, z ktorých družíc týchto misií sú vhodné snímky pre účel tejto práce. Na základe rešerše literatúry bolo rozhodnuté, že sa v tejto práci použijú najnovšie družice Landsat-8 a Sentinel-2. Jeden z dôvodov výberu týchto dvoch družíc bol ten, že majú nastavené niektoré kanály tak, aby zachytávali určité špecifiká vegetácie, na základe ktorých sa dajú napr. odlíšiť jednotlivé druhy drevín (viac v podkapitolách *4.5.1 Landsat-8* a *4.5.2 Sentinel-2*).

Družice Landsat-8 a Sentinel-2 majú tiež spoločné to, že disponujú senzormi, ktoré snímajú Zemský povrch v pásmach C, B, G, R a NIR. Tieto pásma u oboch družíc majú podobné vlastnosti (viď *Prílohu 1.5: Pásma družíc 400 - 1040 nm*).

V čom sú tieto 2 družice odlišné je, že družica Sentinel-2 disponuje oproti družici Landsat-8 aj 3 pásmami v krajočervenej časti spektra (tzv. Red Edge, skrt. RE), ktoré sa nachádza medzi červenou časťou spektra a blízko infračervenou časťou spektra. Rozdiel je aj v tom, že kým družica Landsat-8 sníma v blízko infračervenej časti spektra len v jednom pásme, tak Sentinel-2 sníma v tejto časti spektra v 2 pásmach.

V porovnaní s ostatnými družicami z rešerši, družice Landsat-8 a Sentinel-2 majú najpodobnejšie vlastnosti pásiem, v ktorých snímajú Zemský povrch. Dokonca Landsat-8 je oveľa podobnejšia družica Sentinel-2, ako Landsat-4, alebo Landsat-5 (viď *Prílohu 1. 5: Pásma družíc 400 - 1040 nm*, *Prílohu 1.6: Pásma družíc 1000 - 3000 nm*).

Čo sa týka priestorového rozlíšenia, družice Landsat-8 aj Sentinel-2 sa zaraďujú medzi družice so stredným priestorovým rozlíšením (viď *Tabuľka 1*). Na dáta z družíc so stredným priestorovým rozlíšením sa všeobecne preferuje použiť skôr pixelovú klasifikáciu, ako objektovú klasifikáciu. Pretože pre objektovú klasifikáciu je priestorové rozlíšenie snímky z družice so stredným priestorovým rozlíšením príliš nízke, pretože sa ťažko odlišujú jednotlivé objekty (napr. koruny stromov).

Tabuľka 1: Prehľad úrovně priestorového rozlíšenia družíc spomínaných v rešerši

Priestorové rozlíšenie	Družice
stredné (10 - 500 m)	Landsat4-5 Landsat-8 SPOT 5 Sentinel-2
vysoké (1 - 10 m)	WorldView-2 Quick-Bird Ikonos 2 RapidEye Pléiades

Zdroj: Autor

V nasledujúcich odstavcoch je možné vidieť, aké kanály preferujú autori pri klasifikácii snímok z rôznych družíc s použitím rôznych klasifikačných metód.

*Immitzer M. a kol.* [4] vo svojej práci používajú iba dáta z družice WorldView-2. Po použití všetkých ôsmich kanálov vyšla o niečo vyššia celková presnosť klasifikácie, ako len s použitím 4 najvhodnejších kanálov (C, G, R, NIR1). Určenie 4 najvhodnejších pásiem pre klasifikáciu 4 základných druhov drevín je súčasťou práce.

*Verlič A. a kol.* [5] tvrdí, že zo snímok WorldView-2 najlepšie sa dajú odlíšiť druhy drevín v pásmach RE, NIR1, NIR2 a G. Ide o dreviny dub, buk, smrek, borovica a gaštan.

*Lin C. a kol.* [6] vo svojej práci píše, že vysoko priestorové snímky zachytávajú jemné priestorové detaily potrebné pre mapovanie povrchovej vegetácie. Preto je vhodné používať napr. družice QuickBird, alebo GeoEye, ktoré poskytujú snímky s vysokým priestorovým rozlíšením 0,61 m/2,44 m a 0,41 m/1,65 m pre panchromatické a multispektrálne snímky.

*Le Louarn* [13] je presvedčený, že najlepšie je možné odlíšiť listnaté stromy od ihličnatých v NIR pásme v prípade použitia dát z družice Pléiades.

Z vyššie uvedených tvrdení je možné vidieť, že autori najčastejšie používali kanály G, R a NIR. K týmto výsledkom dospeli na základe empirických výsledkov v ich prácach. Tieto informácie sa nedajú použiť v tejto práci, pretože podmienky za akých dosiahli tieto výsledky sa v každom prípade líšia. Tieto informácie slúžia iba na porovnanie.

V práci bude využitý spektrálny príznak založený na vlnovej dĺžke, teda bude sa pracovať s pôvodnými HMS kanálmi.

## 2.5 Výsledky klasifikácie vybraných druhov drevín

Jedným z kľúčových spôsobov posudzovania vhodnosti metódy pre klasifikáciu dát určitého územia je celková klasifikačná presnosť (OA), procesná (PA) a užívateľská (UA) klasifikačná presnosť konkrétnych tried. Tieto ukazovatele presnosti boli použité aj v tejto práci.

Pri posudzovaní vhodnosti metód, výber parametrov a ich hodnôt je dôležitým faktorom dosiahnutá celková presnosť klasifikácie a tiež užívateľská a procesná presnosť jednotlivých drevín, ktoré sú skúmané. V tomto prípade dubu, buku, smreku a borovice.

Niektoré druhy drevín majú podobné spektrálne vlastnosti, ako iné druhy drevín. Napr. listnaté druhy drevín majú v určitom spektrálnom rozsahu vyššiu odrazivosť prirodzeného žiarenia, ako ihličnaté druhy drevín.

Podľa rešerše (viď *Prílohy 1.1 a 1.2*) najvyššiu celkovú presnosť dosahujú klasifikácie v 3 prácach, v ktorých nebola ani jedna zo skúmaných druhov drevín dub, buk, smrek, borovica. V dvoch prípadoch boli skúmané časti dažďového pralesa na Taiwane a v Brazílii. V treťom prípade išlo o klasifikáciu mestskej zelene v časti mesta Marseille vo Francúzsku. Klasifikácia lesných porastov v týchto prácach dosiahla hodnoty celkovej klasifikačnej presnosti od 98,4% do 100%. Autori článkov dosiahli tak vysoké hodnoty celkovej presnosti pomocou klasifikátorov MLC, SVM a RF. Na základe rešerši je teda možné tvrdiť, že je možné pomocou týchto klasifikátorov dosiahnuť uspokojivé výsledky, ak sa nastaví ich parametrom vhodné hodnoty.

V poradí ďalším článkom v ktorom klasifikácia dosahuje vysokej celkovej presnosti je práca od *Immitzera a kol.* [4] v ktorom pomocou klasifikátora RF dosiahol celkovú presnosť 95,9%. V tejto práci boli použité 4 klasifikačné triedy a to dub, buk, smrek a borovica. Procesná aj užívateľská presnosť jednotlivých druhov drevín dosiahla vysoké hodnoty. Najnižšiu procesnú presnosť dosiahla borovica a to 93,4%. Najvyššiu procesnú presnosť dosiahol buk a to 98,4%.

Ďalší článok, v ktorom klasifikácia dosiahla vysoké hodnoty presnosti napísali *Carleer A. & Wolff E.* [7]. Pomocou klasifikátora MLC dosiahli celkovú presnosť klasifikácie 85,5%. Najnižšiu užívateľskú presnosť dosiahla borovica a to 43,5%. Najvyššiu používateľskú presnosť dosiahol starší buk a to 100%.

Ďalším článkom v ktorom klasifikácia lesného porastu dosiahla vysokých hodnôt napísal *Bucha T.* [11], ktorý použitím klasifikátora MLC dosiahol celkovú presnosť klasifikácie jednotlivých scén od 80% do 90%. Pri pohľade napr. na chybovú maticu klasifikácie scén 187/26 a 187/26 vidieť, že najnižšiu užívateľskú presnosť dosiahol dub a to 79%. Najvyššiu procesnú presnosť dosiahla borovica a to hodnotu 90,1%. Celková presnosť klasifikácie týchto scén je 89,4%.

Pre zhrnutie. V prvých troch prácach síce autori dosiahli najvyššiu celkovú presnosť klasifikácie z celej rešerše, ale klasifikácia neobsahovala klasifikačné triedy dub, buk, smrek a borovica. Pre určenie najvhodnejších hodnôt pre vlastnosti metodiky boli ďalej porovnávané posledné tri práce a to od *Immitzera M. a kol.* [4], *Carleer A. & Wolff E.*

[7] a *Buchu T.* [11]. Ostatné práce použité v rešerši nedosahovali dostatočne vysoké hodnoty celkovej, užívateľskej a procesnej presnosti jednotlivých druhov drevín. Preto sa nimi ďalej nebudeme zaoberať.

## 2.6 Ročné obdobie

Ešte predtým ako začneme porovnávať klasifikátory a ich parametre pozrieme sa na to, z akého ročného obdobia použili autori snímky na klasifikáciu.

Dôvod prečo je dôležité skúmať druhové zloženie vegetácie z pohľadu ročného obdobia je, že počas vegetačného obdobia listy stromov majú rôznu koncentráciu pigmentov. Kým na jar a na jeseň hrá dôležitú úlohu začiatok, koniec a dĺžka vegetačného obdobia, v lete na vrchole vegetačného obdobia, kedy je vegetácia najviac rozvinutá je možné najlepšie rozlíšiť druhy drevín podľa množstva chlorofylu v listoch.

V jarom období je najvhodnejšie sa zamerať na rôzny začiatok vegetačného obdobia u rôznych drevín. Napr. dub oproti buku má začiatok vegetačného obdobia posunutý o približne 6 až 10 dní skôr. Záleží od konkrétneho roku a miesta. Preto v jarom období je vhodné najst' snímky z dátumu v ktorom už niektorým druhom drevín začalo vegetačné obdobie a iné druhy drevín sa ešte len pripravujú na začiatok vegetačného obdobia.

V jesennom období je najvhodnejšie rozoznať druhy drevín podľa sfarbenia listov.

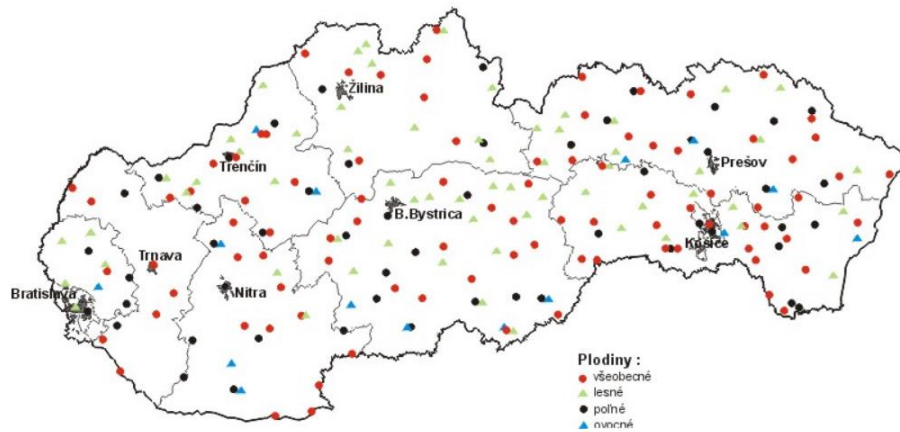
V rokoch 2006 až 2014 bol v Českej republike, na Slovensku a v Poľsku pomerne bohatý výskum zameraný na fenologické fázy lesných druhov drevín. Skúmal sa začiatok pučenia, začiatok olistovania, začiatok žltnutia a začiatok opadu listov.

*Škvareninová* [14] uvádza, že pri smreku súvisia rôzne začiatky nástupu fenofáz s teplotami vzduchu a dĺžkou trvania snehovej pokrývky.

Rozdiel medzi začiatkom fenologických fáz v konkrétnych rokoch môže byť niekedy naozaj veľký. Napr. rozdiel medzi rokmi 2013 a 2014 v prípade dubu je 22 dní. Ide o medián nameraných hodnôt za jednotlivé fenologické stanice na Slovensku. Údaje z fenologických staníc v ČR za roky 2013 a 2014 neboli nájdené, preto v práci boli použité údaje za celé SR. V roku 2014 bol časový posun medzi pučaním dubu a buku len 2 dni, preto nie je možné za každý rok zohnať dáta za Landsat-8, alebo Sentinel-2 z tohto obdobia.

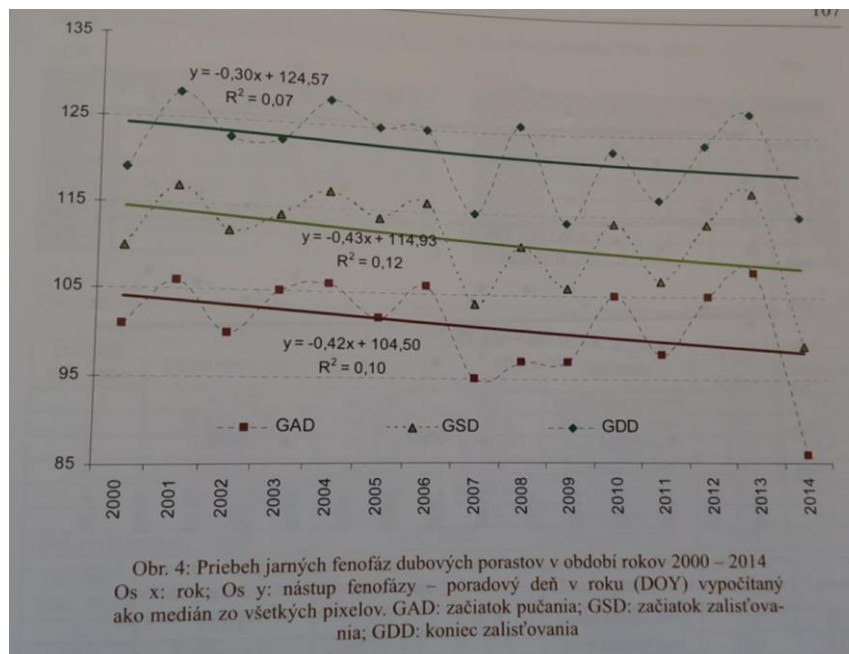


Obrázok 2: Fenologické stanice na Slovensku



Autor: SHMÚ [15]

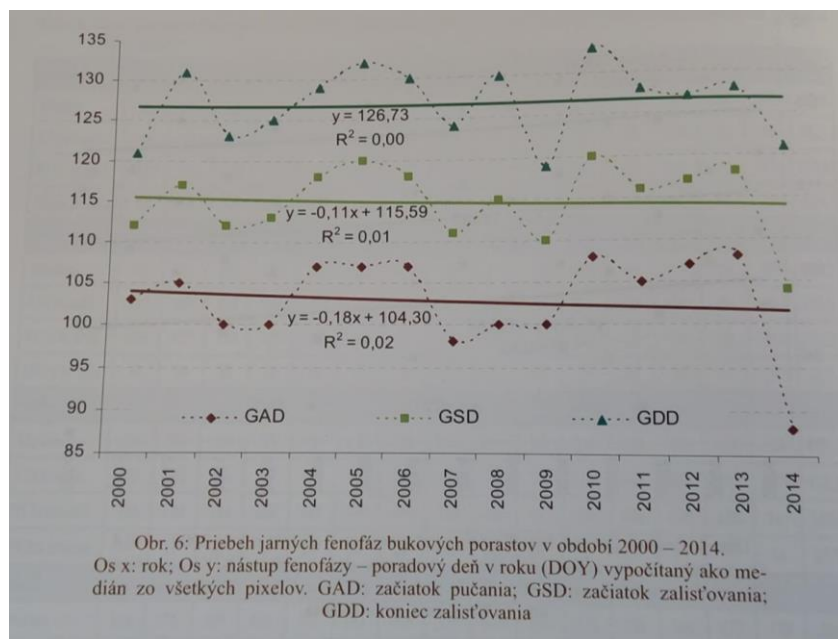
Obrázok 3: Priebeh jarých fenofáz dubových porastov v období rokov 2000 – 2014



Obr. 4: Priebeh jarých fenofáz dubových porastov v období rokov 2000 – 2014  
 Os x: rok; Os y: nástup fenofázy – poradový deň v roku (DOY) vypočítaný ako medián zo všetkých pixelov. GAD: začiatok pučania; GSD: začiatok zalistovania; GDD: koniec zalistovania

Zdroj: : Bucha T. a kol. [2]

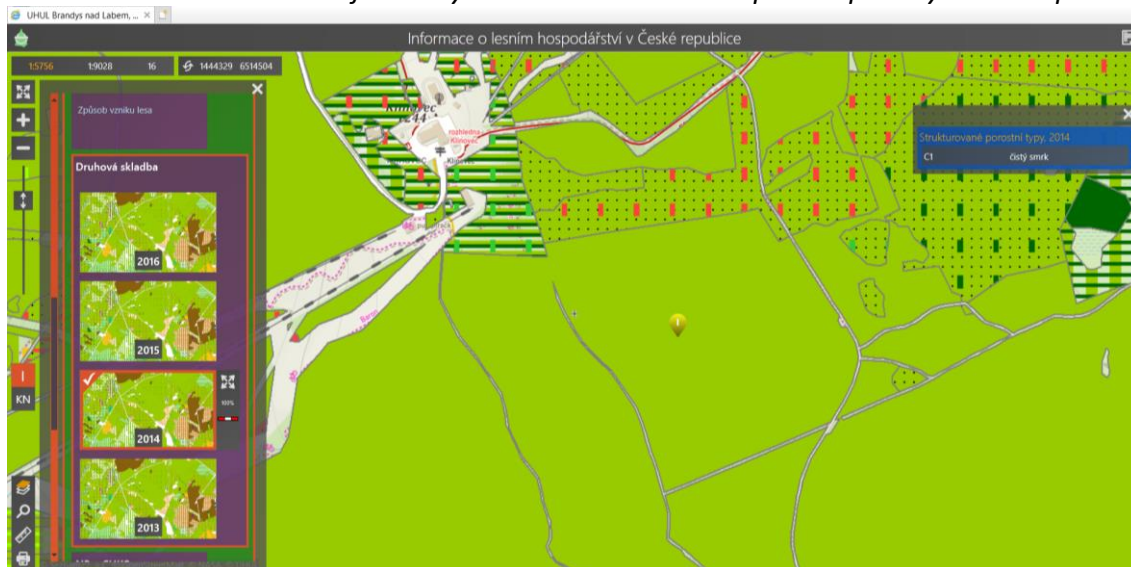
Obrázok 4: Priebeh jarných fenofáz bukových porastov v období rokov 2000 – 2014



Zdroj: : Bucha T. a kol. [2]

Dáta z rôznej literatúry sú dostupné len do roku 2014. Mapový portál LhPoMap od UHUL obsahuje dáta o druhovom zložení lesov za roky 2013, 2014, 2015 a 2016 (viď. Obrázok 5). Legenda sa nachádza v *Tabuľka 5: Druhová skladba lesa*.

Obrázok 5: Zobrazenie druhovej skladby lesa za rok 2014 v mapovom portály LhPoMap



Autor: vlastný

Tabuľka 2: Deň nástupu (medián a 5 % a 95 % kvantil) a dĺžka trvania fenofáz dubových a bukových porastov za obdobie 2013 – 2014

[poradový deň v roku]	DUB		BUK		Rozdiel [poč. dní]	
	2013	2014	2013	2014	2013	2014
GSD (začiatok zalisťovania)	–	–	–	–	–	–
Medián	118	99	118	104	0	5
<b>5% kvantil</b>	<b>115</b>	<b>93</b>	<b>116</b>	<b>96</b>	<b>1</b>	<b>3</b>
95% kvantil	123	110	126	117	3	7
Dĺžka trvania	8	17	10	21	2	4
DAD (začiatok žltnutia)	–	–	–	–	–*	–
<b>Medián</b>	<b>270</b>	<b>278</b>	<b>264</b>	<b>271</b>	<b>-6</b>	<b>-7</b>
5% kvantil	261	267	255	261	-6	-6
95% kvantil	279	288	277	281	-2	-7
Dĺžka trvania	18	21	22	20	4	-1

\* V atribúte "Rozdiel" kladné hodnoty predstavujú rozdiel v počtu dní kedy začala skôr fenologická fáza u dubu, ako u buku. V prípade záporných hodnôt je to naopak.

- jednotky v prípade Medián, 5% kvantil, 95% kvantil sú udávané v poradovom dni v roku
- jednotky v prípade dĺžky trvania sú udávané v počte dní

Autor: Bucha T. a kol. [2], upravené

Tabuľka 3: Dostupnosť snímok za obdobie 2013 – 2014

	Rozdiel medzi dubom a bukom				Dostupné snímky (oblačnosť do 20%)	
	Poradový deň		Dátum		L8	S2
	2013	2014	2013	2014		
GSD (začiatok zalisťovania)	–	–	–	–	–	–
Medián	118 - 118	99 - 104	28.4.	9.4. - 14.4.	x	x
<b>5% kvantil</b>	<b>115 - 116</b>	<b>93 - 96</b>	<b>25.4. - 26.4.</b>	<b>3.4. - 6.4.</b>	<b>NIE</b>	<b>od 2015</b>
95% kvantil	123 - 126	110 - 117	3.5. - 6.5.	20.4 - 27.4.	x	x
Dĺžka trvania	x	x	x	x	x	x
DAD (začiatok žltnutia)	–	–	–	–	–	–
<b>Medián</b>	<b>264 - 270</b>	<b>271 - 278</b>	<b>21.9. - 27.9.</b>	<b>28.9. - 5.10.</b>	<b>NIE</b>	<b>od 2015</b>
5% kvantil	255 - 261	261 - 267	12.9. - 18.9.	18.9. - 24.9. 8.10. -	x	x
95% kvantil	277 - 279	281 - 288	4.10. - 6.10.	15.10.	x	x
Dĺžka trvania	x	x	x	x	x	x

NIE – nie sú dostupné snímky z družice Landsat-8 za dané obdobie s oblačnosťou do 20 %

od 2015 – v danom období ešte nebola družica Sentinel-2 v prevádzke

Autor: vlastný

Sieť fenologických staníc a správu ich údajov, ako aj poskytovanie údajov má na starosti Český hydrometeorologický ústav (ČHMÚ) a Slovenský hydrometeorologický ústav (SHMÚ). Tieto dáta nie je možné voľne získať. Keďže nebolo možné voľne získať

dáta zo siete fenologických staníc z rokov 2015 až 2017, tak druhové zloženie lesov bolo v tejto práci skúmané len v letnom ročnom období.

*Larcher W.* [16] vo svojej knihe píše, že v prípade klasifikácie druhov drevín, by sa mali brať do úvahy tieto vlastnosti vegetácie: zelenatosť listov, percentuálna zmena mezofylu a fyziologické fotosyntetické efekty, pretože tieto faktory majú priamy, alebo nepriamy vplyv na odrazivosť vegetácie. Maximálne hodnoty fotosyntézy rôznych rastlín sú veľmi rozličné a to od 17 do 68 ( $\mu \text{ mol m}^{-2} \text{ s}^{-1}$ ). Toto indikuje, že stromy používajú energiu modrej a červenej časti svetelného spektra rôzne. To pridáva význam v skúmaní druhového zloženia lesov pomocou modrého a červeného pásma.

Pre túto prácu stačí porovnať ročné obdobia a dôvody vybraného vhodného ročného obdobia pre získanie snímok len troch vybraných prác a to *Immitzera M. a kol.* [4], *Carleer A. & Wolff E.* [7] a *Bucha T.* [11]. Všetky informácie o dátumoch a ročných obdobiach v ktorých boli získané snímky pre jednotlivé články sa nachádzajú v *Prílohe 1.1*. V nasledných 3 odstavcoch bude popis dátumu zhotovenia snímok a dôvod, prečo si autori zvolili práve snímky s takýmto dátumom.

*Immitzera M. a kol.* [4] vo svojej práci používajú iba jednu snímku a to z 10.7.2010. Autori si vybrali pre skúmanie druhového zloženia lesa letný mesiac júl, pretože si chceli vybrať snímku z vrcholu vegetačného obdobia, kedy je vegetácia najviac rozvinutá.

*Carleer A. & Wolff E.* [7] vo svojej práci použili snímky z dátumov 8.6. a 7.10.2000. Dôvodom prečo použili pre klasifikáciu jednu snímku zo začiatku leta a druhú zo stredu jesene sú rôzne spektrálne vlastnosti listov v rôznom ročnom období.

*Bucha T.* [11] vo svojej práci použil väčší počet snímok, všetky boli zhotovené v mesiacoch júl a august. Autor si snímky z týchto mesiacov vybral z toho istého dôvodu ako *Immitzera M.* a to pretože predpokladá, že jednotlivé druhy drevín je najvhodnejšie odlíšiť, keď je vegetácia najrozvinutejšia.

Na základe týchto poznatkov sa klasifikácia v rámci tejto diplomovej práce bude robiť na snímkach s dátumom letného ročného obdobia.

## 2.7 Klasifikátor

Ďalším dôležitým krokom pre výber čo najvhodnejšej metódy je zvolenie čo najvhodnejšieho klasifikátora vzhľadom na dané okolnosti (charakteristika územia, vlastnosti vybraných drevín, dostupné dáta, dostupná literatúra, atď.).

Odhladnuc od skúmaných druhov drevín v jednotlivých článkoch, v ktorých vyšla najvyššia celková presnosť klasifikácie, autori použili klasifikátory MLC, RF a SVM. V *Prílohe 1.1* a *Prílohe 1.2* vidieť, ako veľmi vplývajú na výsledok klasifikácie rôzne okolnosti bez ohľadu na výber klasifikátora.

V troch prácach, v ktorých celková presnosť dosiahla najvyššie hodnoty a zároveň v nich boli klasifikované 4 vybrané druhy drevín (dub, buk, smrek, borovica), ako

samostatné klasifikačné triedy sú práce od *Immitzera M. a kol.* [4], *Carleer A. & Wolff E.* [7] a *Buchu T.* [11]. V prvej práci bol použitý klasifikátor RF a v ďalších 2 prácach bol použitý klasifikátor MLC.

V ostatných prácach z rešerši, ktoré dosiahli nižšie hodnoty celkovej presnosti klasifikácie, boli použité klasifikátory RF, SVM, MLC, LDA a PDA. To poukazuje na význam zvolenia vhodných parametrov pre jednotlivé klasifikátory. A tejto téme bude venovaná podkapitola 2.8 *Parametre klasifikátora.*

### **2.7.1 Porovnanie klasifikátorov na základe výsledkov presnosti klasifikácie za jednotlivé práce**

V rešerši je popísaných 10 prác, v rámci ktorých autori pre klasifikáciu druhového zloženia lesov použili jeden, alebo viac z týchto klasifikátorov RF, SVM, MLC a PDA.

Nasledujúce odstavce sa venujú porovnávaniu výsledkov klasifikácie v rámci jednotlivých prác oddelene z rešerše.

Z rešerše spomenutej vyššie je možné vidieť, že v prípade klasifikácie dubu, buku, smreku a borovice najvyššiu klasifikačnú presnosť dosiahol *Immitzer M.* [4] vo svojej práci s použitím metódy RF. Vo svojej práci porovnal výsledky klasifikácie metódou RF s výsledkami klasifikácie metód LDA, MDA, MDG, MD, SAM, SVM, SID, NN a Wilks' Lambda. Klasifikácia metódou RF dosiahla v článku vyššiu celkovú presnosť aj procesnú a užívateľskú presnosť jednotlivých tried, ako výsledky v iných štúdiách ním vytvorenej rešerše (viď *Prílohu 1.9*) s použitím iných klasifikátorov.

*Immitzer M.* [9] v inej práci tiež použil klasifikátor RF s takými istými hodnotami vstupných parametrov, ale použil dáta z iných družíc a vybral si iné výskumné územie. Výsledkom bola nižšia klasifikačná presnosť (viď *Príloha 1.2*).

SVM sa spomína len v štyroch prácach z vybranej literatúry. V prácach od *Immitzer M.*, *Chinsu Lin*, *Le Louarn* a *Sothe C.*

*Immitzer M.* [4] v práci vychádzal z literárnej rešerše 33 prác. Z toho v 7 prácach bola použitá metóda SVM na dátach z AISA, Geo-Eye-1 a ADS40-SH53. V rešerši neopísal na základe akých funkcií bola urobená klasifikácia pomocou metódy SVM (napr. pomocou sigmoidu, radiálnej bázovej funkcie, alebo inej funkcie). V rešerši bol SVM použitý na oddelenie veľkého množstva tried - druhov drevín (3 až 11). SVM nie je vhodná pre klasifikáciu s veľkým množstvom tried, pretože sa stáva výpočetne veľmi náročná.

*Chinsu Lin* [6] v práci porovnáva výsledky aplikácie SVM, MLC, MD, SAM, SID, NN na rôzne kombinácie kanálov (HMS, HMS5VI, SpecTex, HMS13B). V prípade HMS najlepšie výsledky dosahujú v poradí metódy MLC, NN, SVM. Metódy MD, SAM a SID dosahujú horších výsledkov.

V práci *Sothe C.* [12] nebola skúmaná žiadna zo 4 vybraných drevín. Boli použité dáta z Landsat-8 a zo Sentinel-2. Výsledky klasifikácie klasifikátorom SVM vyšli s vyššou presnosťou, ako v prípade použitia klasifikátora RF.

*Le Louarn M.* [13] vyšla klasifikácia metódou RF s vyššou klasifikačnou presnosťou, ako v prípade SVM. Treba však dodať, že išlo o objektovú klasifikáciu a v tejto práci bude použitá pixelová klasifikácia.

### **2.7.2 Prečo iné klasifikátory nie sú vhodnejšie pre použitie v tejto práci**

Klasifikátor SVM bol vylúčený pre použitie v tejto práci, pretože vo všetkých zdrojoch literárnej rešerše dosahoval nižšiu klasifikačnú presnosť, ako RF. Dosiahnutá vyššia klasifikačná presnosť bola iba v prípade práce, v ktorej išlo o skúmanie druhového zloženia dažďových pralesov. Ďalším dôvodom je, že kým metóda RF sa v súčasnej dobe veľmi rýchlo vyvíja a je vysoko automatizovaná, tak pri použití metódy SVM je potrebné nastaviť veľké množstvo parametrov v závislosti od vlastností použitia konkrétnych dát. To znamená, že aj keď autori použili hodnoty niektorých vstupných parametrov, podľa ich vlastnej rešerše, tak je potrebné ručne nastaviť množstvo hodnôt parametrov v závislosti od DN hodnôt snímok.

Klasifikátor MLC v rešerši prác dosahoval výraznejšie nižšiu klasifikačnú presnosť, ako metódy RF a SVM. Vyššiu klasifikačnú presnosť dosiahol iba v práci, v ktorej išlo o klasifikáciu dažďového pralesa, v ktorom bolo iné druhové zloženie drevín, ako v lesoch v ČR. *Lin C.* dosiahol, tak vysokú presnosť klasifikácie vďaka využitiu kombinácie spektrálneho príznaku s texturálnym príznakom a tiež, pretože použil objektovú klasifikáciu. Objektová klasifikácia dosahuje v prácach rešerše vyššiu celkovú presnosť, ako pixelová klasifikácia. *Lin C.* mohol využiť texturálny príznak a objektovú klasifikáciu, vďaka tomu, že pracoval s dátami s vysokým priestorovým rozlíšením, konkrétne s dátami z družice Quick Bird.

V prácach rešerše boli použité aj iné klasifikátory napr. LDA a PDA. V rámci jednotlivých štúdií, aj v rámci porovnávania klasifikátorov, ktoré dosiahli najvyššie klasifikačné presnosti v jednotlivých prácach. Pomocou klasifikátorov LDA a PDA, boli dosiahnuté celkové presnosti klasifikácie s výrazne nižšou klasifikačnou presnosťou, ako v prípade použitia klasifikátorov RF, SVM a MLC.

Z vyššie uvedených dôvodov bolo v tejto práci vylúčené použitie klasifikátorov SVM, MLC, LDA a PDA. Preto v tejto práci bol použitý klasifikátor RF.

## **2.8 Parametre klasifikátora**

Najvyššiu celkovú presnosť klasifikácie, v ktorej boli skúmané 4 vybrané druhy drevín (dub, buk, smrek, borovica) dosiahol vo svojej práci *Immitzer M. a kol. (2012)* [4]. Vo svojej práci použil klasifikátor RF a preto nasledujúce odstavce sú venované voľbe čo najvhodnejších parametrov a ich hodnôt tohto klasifikátora pre dosiahnutie čo najvyššej celkovej presnosti.

Tzv. „základné“ parametre metódy RF sú vo mnohých prácach pomenované ako: kritérium (angl. *criterion*), počet stromov v lese (skrt. *ntree*) a počet náhodných premenných v každom bode (skrt. *mtry*). Ide o názvy parametrov v programovacom jazyku „R“.

*Immitzer M. a kol. (2012)* [4] v práci použili 3 základné parametre. Ako kritérium zvolili Giniho koeficient, počet stromov v lese (*ntree*) zvolili hodnoty 1 až 2500, počet náhodných premenných v každom bode (*mtry*) zvolili 1 až 5. Celková presnosť klasifikácie prudko rástla do počtu stromov 250, následne rástla len mierne. Použitím 1 až 2 náhodných vybraných vstupných premenných pre každý uzol, dosiahla presnosť klasifikácie lepšie výsledky, ako použitím 3 až 5 premenných. Pre porovnanie výsledkov s inými štúdiami autori použili predvolené nastavenia a to 500 stromov (*ntree*) a 2 premenné v každom uzle (*mtry*).

*Immitzer M. a kol. (2016)* [9] vo svojej práci v roku 2016 tiež použili 3 základné parametre a to kritérium, *ntree* a *mtry*. Hodnotu pre počet náhodných stromov (*ntree*) zvolili 500. Ostatné hodnoty parametrov neboli popísané v práci.

*Sothe a kol.* [12] okrem klasifikátora SVM, s ktorým dosiahli o pár desiatín percenta vyššie hodnoty celkovej klasifikačnej presnosti, tak autori použili vo svojej práci aj metódu RF, v ktorej nastavili parametre nasledovne: ako kritérium zvolili Giniho kritérium. Počet rozhodovacích stromov v lese (*ntree*) zvolili 500. Pôvodne však v práci vyskúšali pre parameter *ntree* zvoliť hodnoty 100, 500 a 1000. Klasifikácia pomocou hodnoty 500 dosiahla najvyššej presnosti, preto použili hodnotu 500. Odhadom *mtry* parametra, zvolená bola prednastavená hodnota, ktorá korešponduje s odmocninou celkového počtu kanálov, použitých v každom experimente. Algoritmus RF využil 2/3 tréningových vzoriek pre tréning. Na odhad klasifikačnej chyby RF algoritmus nechal približne 1/3 tréningových vzoriek, ktoré sa nazývajú aj validačné dáta.

*Marine Le Louarn a kol. (2017)* [13] vo svojej práci použili klasifikátor RF s tým, že ako kritérium zvolili Giniho koeficient, počet stromov v lese (*ntree*) zvolili hodnoty 1 až 2500, počet náhodných premenných v každom bode (*mtry*) zvolili 1 až 5. Ostatné parametre nechali nastavené defaultne. Výsledok klasifikácie bol taký, že počet stromov (*ntree*) rástol po hodnotu 500 a počet náhodných premenných v každom bode (*mtry*) sa rovnal odmocnine celkového čísla vstupných hodnôt (t.j. počet polygónov, ide totiž o objektovú klasifikáciu).

Na základe tejto rešerši v tejto práci bude použitá metóda RF s Giniho kritériom, 500 rozhodovacích stromov (*ntree*) a 2 náhodné premenné v každom uzle (*mtry*). Keďže metóda RF patrí k riadenej klasifikácii, tak táto klasifikácia bude riadená.

Popis jednotlivých parametrov sa nachádza v podkapitole 5.2.2 *Nastavenie parametrov RF podľa literatúry*.

## 2.9 Programovací jazyk

V tejto podkapitole je opísaná voľba použitia jazyka „R“ vs. Python. Následne je tu opísaná rešerš zameraná na praktický postup klasifikácie v programovacom jazyku Python. V tejto rešerši absentuje klasifikácia druhového zloženia lesov, keďže žiadne práce neboli nájdené, v ktorých by bol použitý jazyk Python pre klasifikáciu druhového zloženia lesov.

Pre aplikovanie klasifikátora RF bolo potrebné vybrať vhodný nástroj. Pri skúmaní, aké existujú možnosti, boli nájdené dva programovacie jazyky.

Programovací jazyk „R“ je interpret jazyka predovšetkým určený za účelom dolovania dát (angl. data mining) a štatistik [17] a okrem iných modulov obsahuje aj modul machineLearning [18] a [19]. Ako je vedené v „What is R?“ [17] základom jazyka „R“ je jazyk „S“, ktorý je určený na štatistické výpočty a grafické výstupy a bol vyvinutý Bell Laboratories (predtým AT&T, teraz Lucent Technologies) Johnom Chambersom a jeho kolegami. Jazyk „R“ stojí na základoch jazyka „S“ a mnohé moduly má spoločné, ale „R“ poskytuje širokú škálu štatistických nástrojov (lineárne a nelineárne modelovanie, klasické štatistické testy, časové analýzy, klasifikácia, ...) a grafických nástrojov, ktoré sú značne rozšírené. Silnou stránkou jazyka „R“ je jednoduché publikovanie výstupov vrátane matematických symbolov a rovníc. „R“ je voľný softvér vrátane zdrojových kódov (Free Software Foundation's GNU General Public License) s možnosťou kompilácie a behu pod širokou škálou operačných systémov od UNIX, LINUX, Windows až po MacOS.

Programovací jazyk Python, ako je uvedené v Applications for Python [20] je určený nie len pre webové i internetové aplikácie, ale predovšetkým pre vedecké a matematické úlohy. Pre tento jazyk sú pripravené knižnice scikit-learn Machine Learning in Python [21] a to pre verzie jazyka 2.7, alebo 3.6 a tiež pre 32 bit, alebo 64 bitové operačné systémy.

Ako vhodný nástroj pre túto prácu bol vybraný jazyk Python, nakoľko ESRI tento jazyk používa na rozšírenie funkcií pre svoje produkty. V jazyku Python je vyvinutý modul ArcPy, čo je knižnica nástrojov pre aplikáciu ArcMap a v súčasnosti už aj pre najnovší produkt ArcGIS Pro. Kým najnovšie verzie ArcMap používajú staršie verzie jazyka Python 2.7, tak najnovšie verzie ArcGIS Pro používajú najnovšie verzie Pythonu 3.6 [20].

Jazyk Python bol vybraný kvôli tomu, že sa dá v ňom pracovať v ArcMap a ArcGIS Pro. Tieto produkty od spoločnosti ESRI sú pre autora tejto práce dostupné a sú užívateľsky prijateľnejšie, ako softvéry používajúce jazyk „R“. [22]

Pri skúmaní možných riešení použitia RF bolo nájdených niekoľko možných príkladov použitia RF, ale ani jeden z nich priamo neriešil danú problematiku.

Prvý príklad uvedený v článku „Memoryerror in Supervised Random Forest Classification in Python sklearn“ [23] opisuje možnosť klasifikácie dát z Landsat-8 s 5 klasifikačnými triedami klasifikátorom RF pomocou jazyka Python (v práci je uvedený príklad programu). Hodnoty parametrov klasifikátora boli nastavené defaultne okrem



parametrov  $n\_estimators = 1000$  a  $oob\_score = True$ . Tento postup nie je možné dosiahnuť pomocou nástrojov v ArcMap a ArcGIS Pro. Uvedený program je, ale nefunkčný a úplne neodpovedá na otázky vedúce k použitiu RF v tejto práci.

Druhý príklad opisovaný v sérii článkov ESRI „Predict Seagrass Habitats with Machine Learning“ [24] hovorí o použití RF pre klasifikáciu (predikciu) podmorských rias (angl. seagrass) na základe siedmich meraných fyzikálnych a chemických veličín vody. Na základe nameraných údajov a použitiu RF bol vytvorený model, kde všade na Zemi rastú podmorské riasy. Tento príklad bol dobrý na to, že ukázal možnú previazanosť medzi jazykom Python, klasifikátorom RF a modulom ArcPy (rozšírenie jazyka Python pre prácu s geofile databázou v produktoch od ESRI). V práci boli použité hodnoty parametrov klasifikátora RF defaultne okrem parametrov  $n\_estimators = 500$ ,  $oob\_score = True$ .

Posledný príklad v článku „Random Forest Classifier Example“ [25] opisuje podrobne použitie nástroja RF a jeho ďalších analytických funkcií, ako sú napr. funkcie pre vytváranie chybovej matice, či analýzu váženého dôležitosti vstupov (v práci použité na relevantnosť určitého kanálu na klasifikáciu stromov). V tejto práci išlo o klasifikáciu druhu kvetov na základe ich vlastností. Hodnoty parametrov klasifikátora boli v tejto práci nastavené nasledovne  $bootstrap=True$ ,  $class\_weight=None$ ,  $criterion='gini'$ ,  $max\_depth=None$ ,  $max\_features='auto'$ ,  $max\_leaf\_nodes=None$ ,  $min\_impurity\_split=1e-07$ ,  $min\_samples\_leaf=1$ ,  $min\_samples\_split=2$ ,  $min\_weight\_fraction\_leaf=0.0$ ,  $n\_estimators=10$ ,  $n\_jobs=2$ ,  $oob\_score=False$ ,  $random\_state=0$ ,  $verbose=0$ ,  $warm\_start=False$ .

Na základe príkladov [23] [25] a [24] bolo rozhodnuté, že pre účely tejto práce bude použitý ArcGIS Pro do ktorého budú doinštalované potrebné knižnice. Výhoda ArcGIS Pro oproti použitiu staršej aplikácie ArcGIS Desktop je, že je v nej možné použiť najnovšie knižnice. V budúcnosti Spoločnosť ESRI plánuje, že po dokončení všetkých funkcií ArcGIS Pro bude ich hlavným produktom.

**Pre účely tejto práce bola použitá knižnica scikit-learn v.0.19.1 (Nov 21, 2017) [26].**

Tabuľka 4: Porovnanie názvov kľúčových parametrov RF pre túto prácu medzi jazykom "R" a Python

Jazyk „R“		Python	
Premenná	Vysvetlenie	Premenná	Vysvetlenie
criterion	Spôsob meranie kvality rozdelenia (gini, entropy)	criterion	Spôsob meranie kvality rozdelenia (gini, entropy)
ntree	Počet použitých stromov	n_estimators	Počet stromov v lese
mtry	Počet vlastností náhodne vybraných pri každom rozdelení	max_features	Počet vlastností s ktorými je potrebné uvažovať pri hľadaní najlepšieho riešenia
oob.prox	Použití tréningové dáta na overenie výsledku z množiny „bagging“ (False, True)	oob_score	Použití tréningové dáta na overenie výsledku z množiny „bagging“ (False, True)

## 2.10 Záver

Táto časť bude venovaná zhrnutiu najvhodnejšej metodiky práce vyplývajúcej z rešerše pre skúmanie druhového zloženia lesov na území ČR.

V tejto práci budú mať samostatné klasifikačné triedy dub, buk, smrek a borovica. V klasifikácii budú použité pásma B, G, R a NIR družíc Landsat-8 a Sentinel-2. Použité budú snímky z letných mesiacov júl a august, prípadne jún. Na výrezoch zo snímky bude urobená riadená pixelová klasifikácia. Ako klasifikátor bude zvolený RF. **Parametre použité pri klasifikácii budú zvolené: 500 rozhodovacích stromov (*n\_estimators*), 2 náhodné premenné v každom uzle (*max\_features*) a koeficient Gini (*ale nakoniec v práci na základe testov z kapitoly 5.2.3.2 bol použitý koeficient Entropy*5.2.2.2).** Programovací jazyk bude použitý Python v softvéri ArcGIS Pro.

## 3 RANDOM FOREST

### 3.1 Popis a vývoj metódy

*Tin Kam Ho* (1995) [27] (1998) [28] vo svojich prácach popisuje metódu Random Forests (skrt. RF, slov. Náhodné lesy), alebo Random Decision Forests (náhodné rozhodovacie lesy), ako metódu učenia súborov pre klasifikáciu, regresiu a iné úlohy, ktoré fungujú na princípe budovania množstva rozhodovacích stromov (*Poznámka: množstvo stromov tvorí les z toho ten názov.*) v tréningovom procese, kde výstup je trieda (klasifikácia), alebo stredná predikcia (regresia) jednotlivých stromov, ktorá vznikla z „hlasovania“ jednotlivých rozhodovacích stromov. Metóda RF napráva výsledky rozhodovacieho stromu, ak chybné vyhodnotí údaje z tréningového súboru. [29] [30]

*Poznámka: V tejto práci sa neprekladá výraz „Random Forests“ nakoľko by výraz lesy mohol byť zamieňaný z výrazom lesy označujúce súbor drevín a ktorý je v práci často spomínaný.*

Prvý algoritmus pre RF vytvoril *Tin Kam Ho* [27] pomocou metódy náhodného subpriestoru [28], ktorý je v *Ho T. K.* formulácii spôsobom implementácie prístupu k stochastickým diskrimináciám, ktorý navrhol *Eugene Kleinberg* [31].

Rozšírenie algoritmu vyvinuli *Leo Breiman* [32] a *Adele Cutler* [33] a "Random Forests" je ich ochranná známka [34]. Tiež vytvorili kód pre užívateľov tejto metódy [35]. Rozšírenie spája Breimanovu "baggingovú" myšlienku a náhodný výber vlastností, aby sa vytvorila kolekcia rozhodovacích stromov s kontrolovaným rozptylom, ktoré najprv predstavil *Ho T. K.* [27] a neskôr nezávisle od *Ho T. K.* aj *Amit Y.* a *Geman D.* [36].

Výber náhodných podmnožín príznakov je príklad metódy náhodných podpriestorov (angl. Random Subspace Method), ktorú *Ho T. K.* použil, ako implementačnú techniku pre predchádzajúcu techniku od *Kleinberg E.*

Toto používajú aj iné metódy, ktoré musia byť necitlivými na určité údaje. Pri tejto metóde je potrebné nastaviť správne parametre lebo ak sa nenastavia dostatočne presne, metóda nedosiahne požadované výsledky. Ďalším nebezpečenstvom je preparametrizovanie funkcie (overfitting – vysvetlené nižšie). Veľmi dobre to popísal vo svojej práci *Kleinberg E.* (1996) [37]

*Breiman L.* bol ovplyvnený prácou *Amit Y.* a *Geman D.* (1996) [36], ktorý predstavil myšlienku prehľadávania náhodnej podskupiny dostupných rozhodnutí pri rozdelení uzla v súvislosti s pestovaním jedného rozhodovacieho stromu. Myšlienka výberu náhodného podpriestoru od *Ho T. K.* bola tiež dôležitá pri navrhovaní náhodných lesov. V tejto metóde sa vytvára les rozhodovacích stromov. Variabilita medzi rozhodovacími stromami sa zadáva tým, že sa pred tréningom každého stromu (uzla) premietnu tréningové dáta do náhodne vybraného podpriestoru. Nakoniec

myšlienku náhodnej optimalizácie uzlov, kde rozhodovanie v každom uzle je vybrané náhodným postupom, namiesto deterministickej optimalizácie predstavil *Dietterich T. G.* (2000) [38].

*Breiman L.* (2001) [32] ako prvý aplikoval metódu „*Random Forests*“ a aj ju publikoval v časopise *Machine Learning*. Tento článok popisuje spôsob budovania lesa nekorelovaných stromov pomocou procedúry podobnej CART kombinovanej s optimalizáciou náhodných uzlov a „*bagging*“. Autor v tomto článku spája niekoľko starších zložiek s novými zložkami a používa:

1. „*Out-of-bag error*“ a „*generalization error*“, ako odhad chyby na vzorke vytvorenej metódou *bagging* (*Vyhodnocuje sa chyba na vzorke údajov, ktorá neobsahuje originálnu sadu dát so sadou dát, ktorá ich obsahuje.*)
2. Meranie variabilného významu prostredníctvom permutácie.

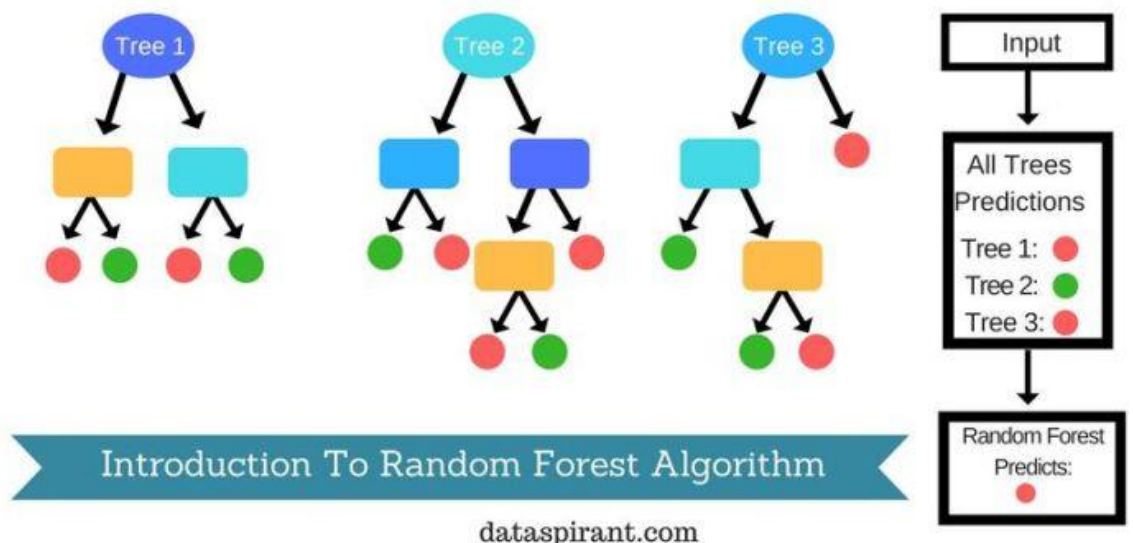
Článok tiež ponúka prvý teoretický výsledok pre metódu RF, ktorého postup berie do úvahy aj „*generalization error*“, ktorá závisí od sily rozhodovacích stromov a ich korelácie.

### 3.2 Algoritmus

Celý algoritmus vychádza zo základného princípu uvedeného v predchádzajúcej kapitole. Praktickú realizáciu opisuje *Polamuri S.* v práci „*How the random forest algorithm works in machine learning*“ [39].

„RF algoritmus je riadený klasifikačný algoritmus a ako naznačuje názov, tento algoritmus vytvára les z množstva stromov. Vo všeobecnosti čím viac stromov, tým väčší les a tým vyššia dosiahnutá presnosť.“ Výsledok potom pozostáva z „*hlasovania*“ jednotlivých rozhodovacích stromov, tak ako je znázornené na „*Obrázok 6*“

Obrázok 6: Základný princíp RF – zdroj: *Random Forest Introduction* [1]



Pri tomto postupe vzniká problém, ako mať viac tréningových údajov pre každý rozhodovací strom, keď je k dispozícii len jedna tréningová sada údajov. Tento problém je riešený metódou nazývanou „bagging“ (opísaná nižšie), ktorá z tréningového súboru vytvorí potrebný počet podsúborov, ktoré sú použité pre jednotlivé rozhodovacie stromy.

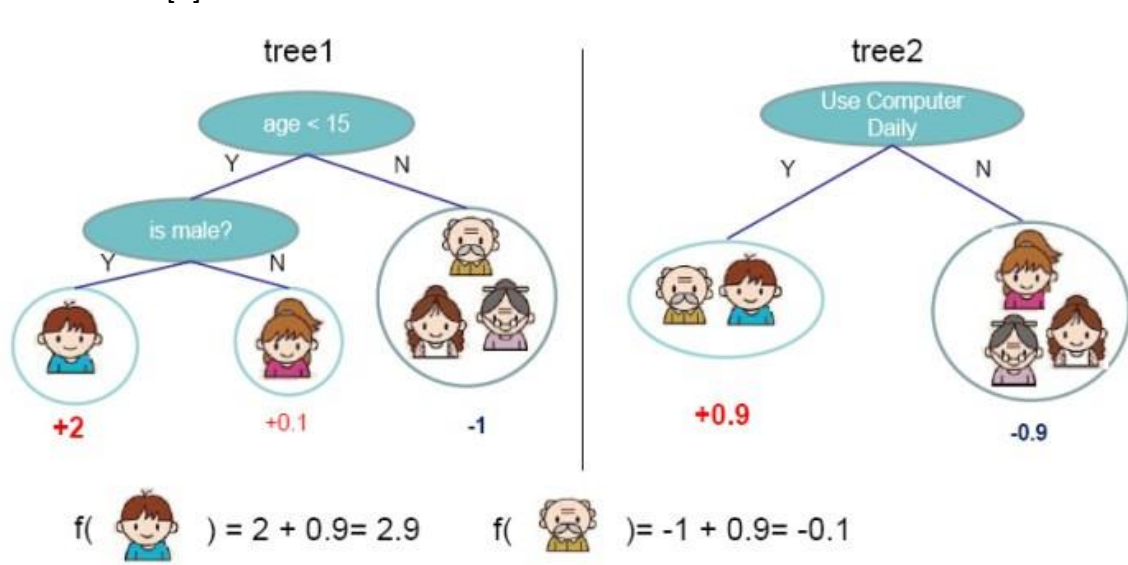
Taktiež spôsob vytvárania jednotlivých rozhodovacích uzlov v stromoch nie je rovnaký pre každý strom, ale na základe typu atribútov (spojité, diskkrétne) používa štatistické metódy nazývané „information gain“ a „gini index“ (vysvetlené nižšie). Úplne neplatí, že vo všetkých prípadoch ak sú hodnoty atribútov diskkrétne, tak je lepšie použiť „gini index“ aj keď vo väčšine prípadoch to tak je. To isté platí pre metódu „information gain“.

Polamuri S. uvádza v práci [39] algoritmus RF. Tento algoritmus sa objavuje aj v iných prácach, ale je opísaný pomocou rôznych programovacích jazykov.

### 3.2.1 Predpoklady vytvorenia rozhodovacieho stromu:

1. Všetky tréningové údaje tvoria koreň rozhodovacieho stromu.
2. Hodnoty atribútov by mali byť kategorizovateľné. Ak hodnoty sú spojité, tak je potrebné z nich vytvoriť diskretný model. (Poznámka: V našom prípade je frekvenčné spektrum súvislé v jednotlivých kanáloch.)
3. Vetvenie v rozhodovacom strome sa vytvára rekurzívne na základe hodnôt atribútov.
4. Použitie atribútov v jednotlivých rozhodovacích uzloch stromu je robené na základe štatistických metód.

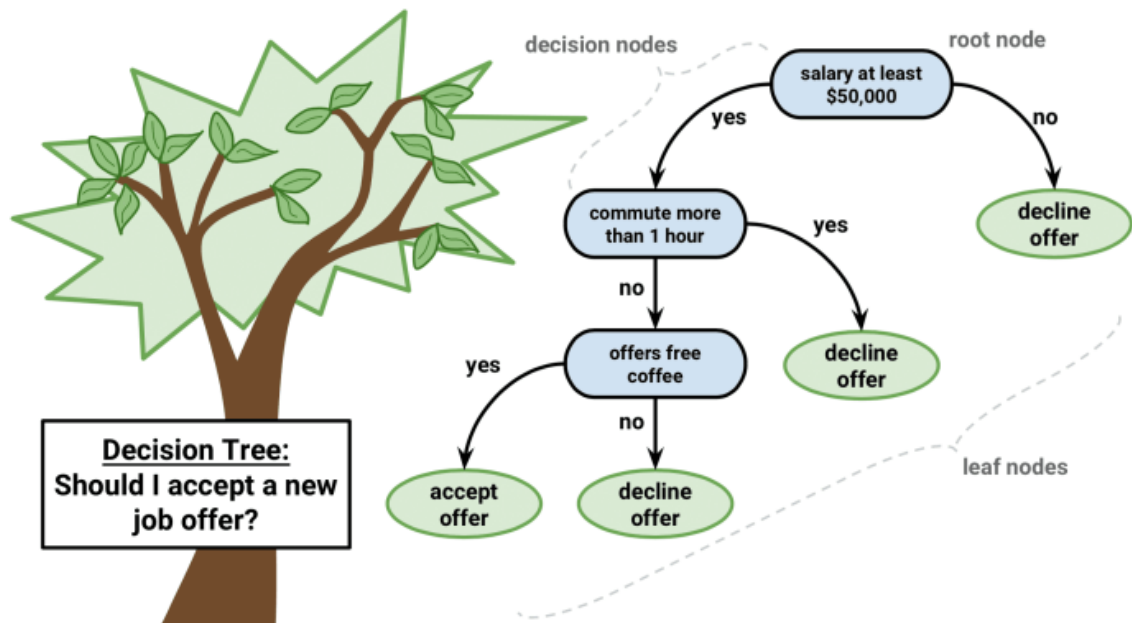
Obrázok 7: Príklad vytvárania rozhodovacieho stromu – zdroj: Random Forest Introduction [1]



### 3.2.2 Rozhodovací stromový algoritmus:

1. Umiestni sa najlepší atribút v koreni rozhodovacieho stromu.
2. Trénovacia sada sa rozdelí na podmnožiny. Podmnožiny by mali byť vytvorené tak, aby každá podmnožina obsahovala údaje s rovnakou hodnotou atribútu. (Poznámka: Podmnožiny vznikajú z predchádzajúcej podmnožiny a reprezentujú množinu atribútov, ktoré prešli cez daný rozhodovací uzol.)
3. Body 1 a 2 sa opakujú dovtedy pokiaľ sa nenájde „lístok“ pre každú vetvu v rozhodovacom strome.

Obrázok 8: Proces klasifikovania atribútov – zdroj: Random Forest Introduction [1]



K bodu 1 tejto podkapitoly: Problém, ako nájsť najlepší atribút, t.j hranicu, ktorá rozdelí údaje na skupiny sa dosiahne na základe náhodného generovania hodnoty, ktorá je hraničná a toto je následne overované pokiaľ nie je nájdená hodnota, ktorá najlepšie oddeľuje údaje. Použitie čisto len tohto spôsobu by bolo výpočtovo veľmi náročné a preto je tento spôsob kombinovaný s použitím štatistických metód. Tu je potrebné nastaviť správne parametre, aby nedošlo k preparametrizovaniu funkcie (overfitting – vysvetlené nižšie) a tým nižšej dosiahnutej presnosti.

### 3.3 Praktické použitie RF

Aplikácia metódy RF je riešená viacerými knižnicami pre programovacie jazyky. Tieto programovacie jazyky sú objektové a preto aj použitie RF je zabalené do objektu, ktorý má zadané atribúty a metódy (Poznámka: metóda je funkcia objektu v objektovom programovaní).

Postup použitia je nasledovný:

1. Vytvorí sa objekt, ktorý ako jednu metódu má metódu pre RF klasifikáciu. (Tie isté objekty sa zvyčajne používajú aj na regresiu.) Pri vytváraní objektu je možné zadať atribúty objektu, ako parametre volanej funkcie.
2. Ak neboli zadané všetky potrebné nastavenia, tak je možné ešte zadať hodnoty atribútov pre použitie RF.
3. Zavolá sa funkcia, ktorá má za úlohu naučiť sa na základe trénovacej množiny údajov. Ako parametre takejto funkcie sú minimálne dve množiny. Jedna množina sú trénovacie dáta (voláme ju množina „X“) a druhá je množina klasifikácie (voláme ju množina „Y“), ktorá obsahuje známe odpovede. Výsledkom volania funkcie je vytvorenie lesu rozhodovacích stromov, ktoré sú súčasťou objektu.

*Poznámka: V našom prípade ako množina „X“ je náhodný výber pixelov z nám známych lokalít spolu z DN hodnotami v jednotlivých kanáloch družice. Množina „Y“ je pole údajov o druhu stromov. Hodnota indexu v poli množiny „X“ je rovnaká ako hodnota indexu v poli množiny „Y“ (v „Y“ sú odpovede na „X“).*

4. Zavolá sa funkcia, ktorá na základe už vytvoreného „lesa“ priradí k validačnej množine „Xv“ množinu „Yv“.
5. Vytvorený objekt následne vie pomocou volaní rôznych jeho metód vytvárať rôzne štatistické výstupy. Dá sa vytvoriť napr. chybová matica, tabuľka vplyvu atribútov na rozhodovací proces, korelačná tabuľka medzi atribútmi množiny „X“ a pod.
6. Ak bola dosiahnutá dostatočná presnosť je možné použiť postup v bode 5 na množinu údajov u ktoré je potrebné klasifikovať.

### **3.4 Procesy použité v RF**

#### **3.4.1 Bootstrapping**

Vysvetlenie tohto výrazu je potrebné na pochopenie výrazu *Bagging* často používaného pri RF.

Ako uvádza *Gehlot M.* [40] bootstrapping rieši problém, ako vytvoriť veľké množstvo množín na základe tých istých údajov. Princíp spočíva v tom, že náhodný príklad vzniká z „n“ známych pozorovaní, kde náhodne vybrané údaje sú nahradzované z pozorovaných údajov. Takto vznikajú množiny, ktoré nie sú identické sú rovnako veľké a ale nemajú iné údaje len tie ktoré boli zozbierané.

Príklad uvádzaný v [40]: Ak máme množinu 2, 4, 6, 7, 7, tak nasledujúce možné „bootstrap“ množiny sú:

- 2, 6, 6, 7, 7
- 4, 6, 7, 7, 7
- 2, 2, 4, 6, 6
- 2, 2, 2, 4, 7
- 2, 2, 2, 2, 2
- 4, 7, 7, 7, 7

### 3.4.2 Bagging

*Gehlot M.* v práci [40] uvádza, že „*bagging* je *bootstrap aggregating* (slov. bootstrap = zoskupovanie), ktorý je meta-algoritmus strojového učenia určený na zlepšenie stability a presnosti algoritmov strojového učenia používaných v štatistickej klasifikácii a regresii. Hoci sa zvyčajne používa pri rozhodovacích metódach, môže byť použitý s akýmkoľvek typom metódy.“ *Geurts P. a kol.* (2013) [41] tvrdia, že cieľom tohto postupu je znížiť „predzaujatosť“, t.j. brať do úvahy určité atribúty z tréningovej množiny.

### 3.4.3 Information gain

*Polamuri S.* uvádza v práci [39], že metódou „*information gain* (slov. informačný zisk) je snaha odhadnúť informácie obsiahnuté v každom atribúte. Na toto sa používajú metódy obsiahnuté v teórii informácií, ktoré hovoria, že na meranie náhodnosti ,alebo neistoty náhodnej premennej „*X*“ je definovaná *Entropia*.

Príkladom binárneho klasifikačného problému s dvoma triedami pozitívnu a zápornou hodnotou je, ak sú všetky príklady hodnôt pozitívne, alebo sú všetky negatívne. V tom prípade *entropia* bude nulová. Ak je polovica záznamov pozitívna a druhá negatívna, tak entropia je veľmi vysoká.

Vypočítaním entropického merania každého atribútu môžeme vypočítať jeho zisk informácií. Výpočet informácií vypočíta očakávané zníženie entropie v dôsledku triedenia atribútu. Zisk informácií je možné vypočítať.“

V tejto práci pri použití klasifikátora RF bola použitá hodnota *Entropy* v parametre kritérium. Hodnota *Entropy* dáva lepšie výsledky pri spojitych hodnotách atribútov.

### 3.4.4 Gini index

Ako *Polamuri S.* opisuje v práci [39], že „*gini index* je metodika na meranie toho, ako často by bol náhodne vybraný prvok nesprávne identifikovaný. Znamená to, že by sa mal uprednostniť atribút s nižším indexom *gini*.“

### 3.4.5 Overfitting

*Polamuri S.* [39] uvádza že „*overfitting* je praktický problém pri vytváraní modelu rozhodovacieho stromu. Model má problém nadhodnocovania, keď algoritmus naďalej prechádza hlbšie a hlbšie, aby sa znížila chyba tréningovej množiny, ale výsledkom je



zvýšená chyba testovacej množiny, t.j. presnosť predpovede nášho modelu klesá a väčšinou sa to stane, keď sa vytvorí veľa vetiev kvôli odchýlkam a nezrovnalostiam v údajoch.“

K overfittingu dochádza v dvoch prípadoch, buď sa zastaví proces rozhodovania príliš skoro alebo príliš neskoro.

„príliš skoro“ znamená, že sa zastaví vytváranie stromu (vetvenie) skorej, ako je dosiahnutá prahová hodnota nameraných údajov. Problém je, že je ťažké určiť prahovú hodnotu a tým správny bod kde sa proces zastaví (V tejto práci tomu odpovedá parameter *max\_depth* v RF kapitola 5.2.2.6 a parameter *min\_samples\_leaf* v kapitole 5.2.2.8).

„Príliš neskoro“ znamená, že vetvenie v strome postupuje viac a viac až sa do úvahy začnú brať údaje, ktoré sú nepodstatné (napr. šum). Toto sa optimalizuje pomocou krížovej validácie (podmnožín v rámci trénovacej množiny dát), kde sa overuje či toto vetvenie zvýšilo presnosť alebo nie. Ak zvýšilo pokračuje sa vo vetvení. (V tejto práci sme použili túto validáciu zapnutím parametra *oob\_score* ako je uvedené v kapitole 5.2.2.4.)

## 4 ÚZEMIE A DÁTA

### 4.1 Opis územia

Územie Českej republiky sa nachádza vo vnútrozemí v miernom podnebnom pásme. Nachádza sa v oblasti medzi oceánskou a kontinentálnou klímou. [42]

Česká republika patrí v súčasnosti ku krajinám s vysokou mierou lesnatosti. Lesné pozemky pokrývajú v súčasnosti 2 666 376 ha, čo je cca 33,9 % z územia štátu. Rozloha lesov sa od druhej polovice 20. storočia neustále zvyšuje. [43]

V roku 2016 najrozšírenejšou drevinou bol smrek ztepilý, ktorý pokrýval 50,5 % lesného územia ČR. Ďalšími najrozšírenejšími drevinami boli borovica lesná (16,4 %), buk lesný (8,3 %), viacero druhov dubov (7,2 %) a ďalšie (vid' Tabuľka 5).

Tabuľka 5: Druhové zloženie lesov v [ha] a [%] z celkovej plochy porostnej pôdy

Dřeviny	Rok					
	2000	2010	2013	2014	2015	2016
	plocha porostní půdy ha / %					
smrk ztepilý	1 397 012	1 347 239	1 327 398	1 319 733	1 315 487	1 312 204
	54,1	51,9	51,1	50,7	50,6	50,5
jedle	23 138	25 869	27 509	28 251	28 699	29 086
	0,9	1,0	1,1	1,1	1,1	1,1
borovice	453 159	436 308	431 721	429 636	428 030	425 687
	17,6	16,8	16,6	16,5	16,5	16,4
modřín	97 170	100 761	100 917	100 749	100 283	99 917
	3,8	3,9	3,9	3,9	3,9	3,8
ostatní jehličnaté	4 586	6 352	7 048	7 755	7 846	8 063
	0,2	0,2	0,3	0,3	0,3	0,3
jehličnaté	1 975 065	1 916 529	1 894 593	1 886 124	1 880 344	1 847 961
	76,5	73,9	72,9	72,5	72,3	72,1
dub	163 761	178 466	184 180	184 441	185 044	186 718
	6,3	6,9	7,1	7,1	7,1	7,2
buk	154 791	189 998	202 638	207 595	211 835	215 535
	6,0	7,3	7,8	8,0	8,2	8,3
bříza	74 560	72 264	71 628	71 779	71 796	71 579
	2,9	2,8	2,8	2,8	2,8	2,8
ostatní listnaté	183 696	209 559	215 602	219 363	219 207	221 243
	7,1	8,1	8,3	8,4	8,4	8,5
listnaté	576 808	650 287	674 048	683 178	687 882	695 075
	22,3	25,1	25,9	26,3	26,5	26,7
Celkem bez holiny	2 551 873	2 566 816	2 568 641	2 569 302	2 568 227	2 570 036
	98,8	98,9	98,8	98,7	98,8	98,8

Zdroj: UHUL [3]

Územie ČR pokrývajú väčšinou zmiešané lesy. Z druhovo homogénnych lesov sú to dobové, bukové, smrekové a borovicové lesy. [44]

Aj z pohľadu lesných vegetačných stupňov (ďalej LVS), najviac percent územia ČR pokrývajú, z druhového hľadiska heterogénne lesné vegetačné stupne. Z nich najväčšie územie pokrýva jedlobukový LVS s 30,04 % a za ním nasledujú dubobukový (14,89 %) a smrekovobukový LVS (11,95 %). Z pohľadu druhovo homogénnych LVS najviac územia lesa pokrýva dubový VLS a to 8,31 % (vid' Tabuľka 6).

V tomto prípade sú údaje uvedené za rok 2009, pretože tabuľkový *Prehľad lesných vegetačných stupňov a ich zastúpenie v lesoch ČR* sa v dokumente *Zpráva o stavu lesa a lesního hospodářství České republiky* od roku 2010 nenachádza.

Tabuľka 6: *Prehľad lesných vegetačných stupňov a ich zastúpenie v lesoch ČR*

Lesní vegetační stupeň Forest vegetation zone			Nadmořská výška Elevation above sea level	Prům. roční teplota Average annual temperature	Roční srážky Annual precipitation	Vegetační doba Growing season
číslo code	označení prevailing species	%	m	°C	mm	dnů days
0	bory pine	3,73	azonální non-zonal			
1	dubový oak	8,31	< 350	> 8,0	< 600	> 165
2	bukodubový oak with beech	14,89	350–400	7,5–8,0	600–650	160–165
3	dubobukový beech with oak	8,41	400–550	6,5–7,5	650–700	150–160
4	bukový beech	5,69	550–600	6,0–6,5	700–800	140–150
5	jedlobukový beech with fir	30,04	600–700	5,5–6,0	800–900	130–140
6	smrkobukový beech with spruce	11,95	700–900	4,5–5,5	900–1050	115–130
7	bukosmrkový spruce with beech	5	900–1 050	4,0–4,5	1 050–1200	100–115
8	smrkový spruce	1,69	1 050–1 350	2,5–4,0	1 200–1500	60–100
9	klečový dwarf pine	0,29	> 1 350	< 2,5	> 1500	< 60

Zdroj: UHUL [45] (Údaje sú za rok 2009)

## 4.2 Opis vybraných lokalít

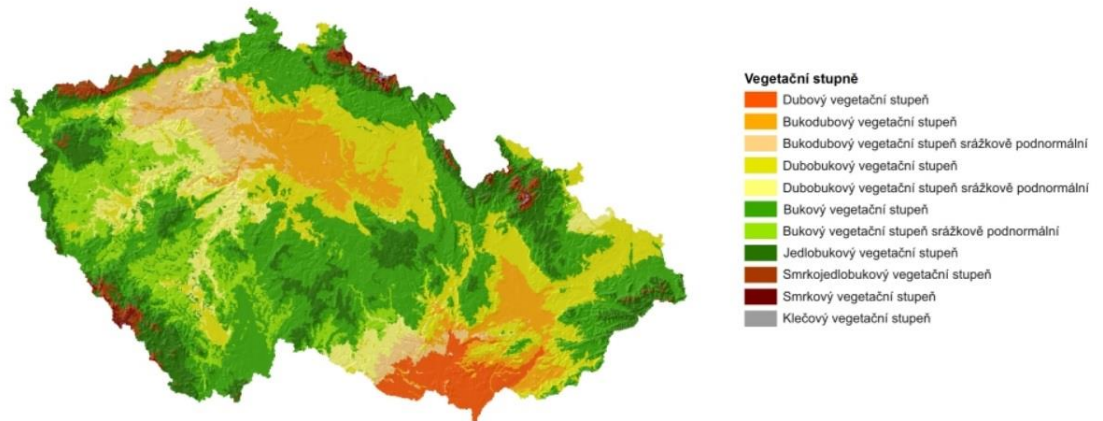
Výber lokalít bol podmienený podmienkou, aby sa všetky lokality (dubových, bukových, smrekových a borovicových porastov) nachádzali na čo najmenšom území. Dôvody boli, aby atmosférické podmienky vo všetkých lokalitách boli čo najpodobnejšie a tiež, aby sa všetky lokality zmestili na jednu scénu Landsat-8 a Sentinel-2, aby sa vyhlo mozaikovaniu snímok.

Keďže dubové a borovicové porasty sa nachádzajú pomerne rovnomerne rozložené na území ČR. Výber územia pre nájdenie vhodných lokalít sa opieral o hľadanie lokalít výskytu rozsiahlejších lesných porastov buku a smreku.

Plošne najrozsiahlejšie územia čistých lesných bukových porastov ležia na juhovýchodných svahoch Krušných hôr. Čisté lesné bukové porasty sú rozšírené ešte v oblastiach Východočeskej tabuli, na juhovýchodných svahoch Brnenskej vrchoviny a pohoria Chřiby.

Smrekové porasty sa v ČR nachádzajú v pohoriach Moravsko sliezske Beskydy, Jeseníky, Orlické hory, Krkonoše, Krušné hory, Český les, Slavkovský les a Šumava. Teda po okraji územia ČR.

Obrázok 9: Lesné vegetačné stupne v ČR



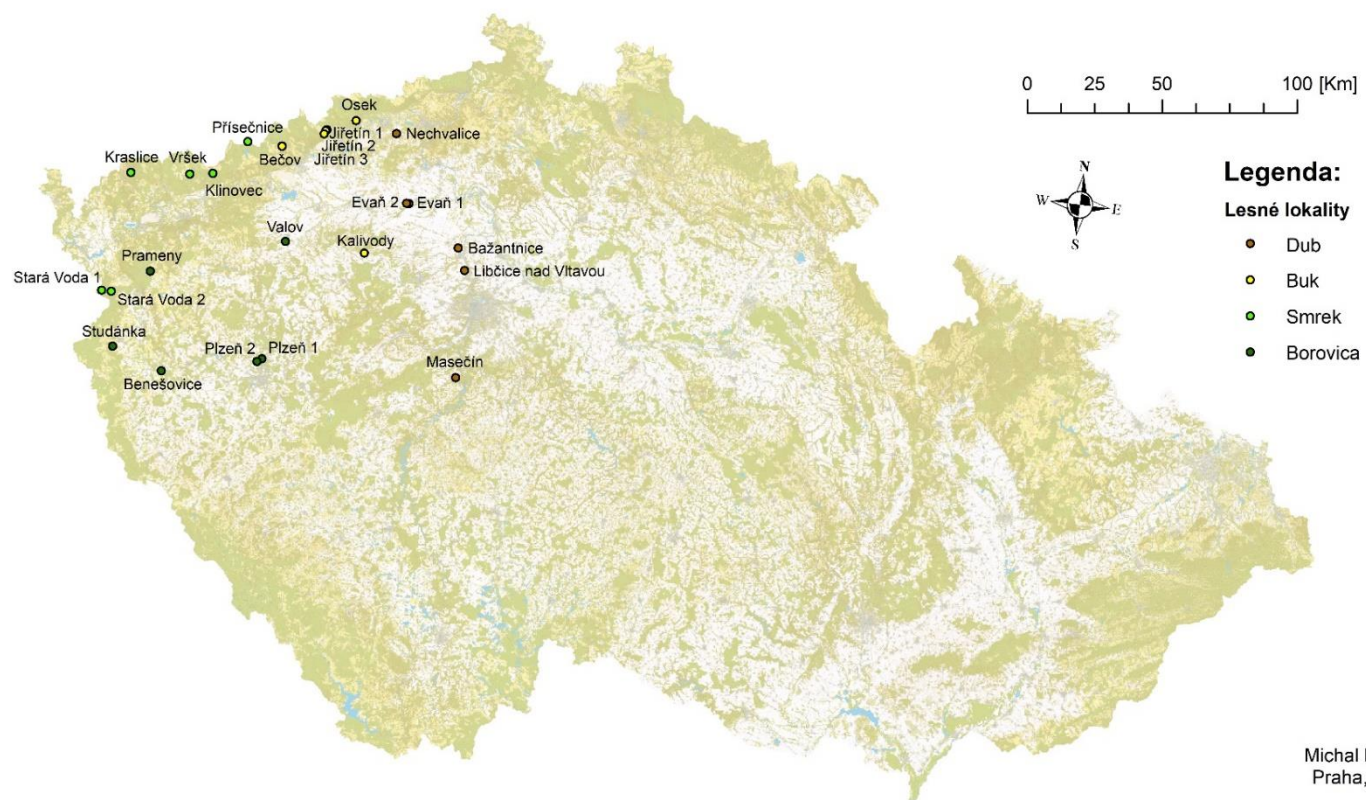
Autor: Divíšek J. a kol. [46], upravené

Z uvedených informácií vyplýva, že bolo zvolené územie, ktoré je ťažko geograficky definovať, pretože sa rozprestiera na viacerých geomorfologických celkoch. Je to územie, ktoré leží medzi mestami Cheb, Ústí nad Lebe, Prahou a Plzňou (viď Mapu č. 1).

Na tomto území bolo vybratých 24 lokalít. Podrobný prehľad lokalít sa nachádza v Prílohe 1.7: *Vlastnosti lokalít a georeferencovania mapových podkladov.*

Mapa č. 1

## Výskumné lokality



Michal Launer  
Praha, 2018

### 4.3 Referenčné dáta

Ako referenčné dáta boli zvolené dáta druhovej skladby lesa od UHUL [47], ktorých základnou jednotkou je štruktúrovaný porastný typ (ďalej SPT). SPT je zadefinovaný spojením skupín drevín (porastných typov, skrt. PT) a charakterom ich zmiešania (vid' *Tabuľka 7* a *Tabuľka 8*).

Ako referenčná vrstva boli použité plochy, ktoré obsahovali čistý štruktúrovaný porastný typ. V tejto práci sa pracovalo so 4 porastnými typmi. Prvým bol dub (kód 5, ozn. PT DB). Okrem dubu letného mohol obsahovať ojedinele aj dub zimný, dub plstnatý, dub bahenný, dub cerový, gaštan jedlý, brest horský, vāz, lipu srdcovitú, lipu veľkolistú a lipu striebornú. Druhým bol buk (kód 6, ozn. PT BK). Okrem buku lesného mohol obsahovať aj prímes javora mliečneho, javora horského a čerešne vtáčej. Tretím bol smrek (kód 1, ozn. PT SM). Tejto porastný typ obsahuje iba smrek ztepilý. Štvrtým bola borovica (kód 3, ozn. PT BO). Obsahuje iba borovicu lesnú.

Pre každý porastný typ bolo nájdených 6 lokalít, ktoré boli použité pre získanie trénovacích množín pixelov.

Mapa STP nie je najpodrobnejšia lesnícka mapa v ČR. Najpodrobnejšie sú lesné porastové mapy, ktoré okrem porastného typu obsahujú aj vek porastu. Najpodrobnejšie informácie nie sú v mapovej podobe, ale sú uložené v podobe tabuliek v hospodárskych knihách. Zatiaľ čo v mapách STP a lesníckych porastových mapách dosahujú údaje o zmiešaní porastných typov od 10 do 20 %, v hospodárskych mapách je dosahovaná presnosť udávania zmiešania porastných typov 5 %.

Voľne je dostupná len mapa STP v podobe mapovej služby v portály LhpoMap od UHUL, avšak mapová aplikácia dátového skladu Lho od UHUL poskytuje už nejaký čas vrstvu lesníckych porastových máp a tiež aj lesnícke hospodárske knihy, ale táto vrstva nie je naplnená údajmi ani do 5 % (vid' *Obrázok 10*). Dátový sklad Lho má rovnaké užívateľské rozhranie ako LhpoMap, ale ide o inú aplikáciu, ktorá je tiež verejnosti voľne dostupná. Keďže mapová aplikácia dátového skladu Lho nemá dátami naplnené vrstvy lesnícke porastové mapy a hospodárske knihy, tak v tejto práci, ako referenčná vrstva bola použitá mapová vrstva STP.





Tabuľka 8: Charakter zmiešania

Charakter smiešaní	popis	Šrafa
<b>Čisté (C)</b>	jeden porostný typ se zastoupením 90 – 100 %	Plná plocha
<b>Smiešané (S)</b>	<b>dominantní (D)</b> - jeden porostný typ se zastoupením 70 – 89,9 %	Plná plocha s černým tečkovaním
	<b>majoritní (M)</b> - jeden porostný typ se zastoupením 50 – 69,9 %	Kolmá šrafa
<b>Nesourodé (N)</b>	<b>základní (Z)</b> - jeden porostný typ se zastoupením 30 – 49,9 %	Vodorovná šrafa
	<b>přimíšené (P)</b> - jeden porostný typ se zastoupením 10 – 29,9 %	Prší
<b>Vtroušené (V)</b>	skupiny dřevin se zastoupením do 10 % nejsou zpravidla uvažovány	

Zdroj: UHUL [47]

#### 4.4 Dáta pre upresnenie hraníc lokalít

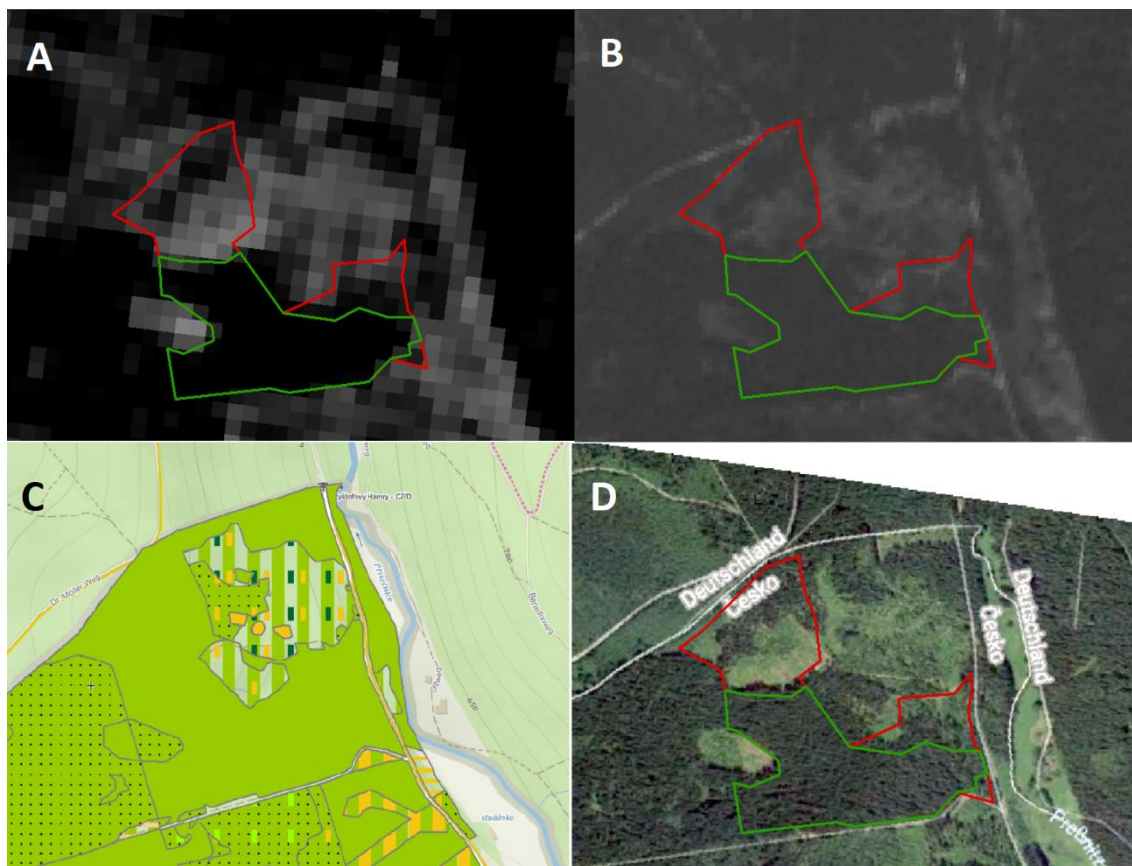
Dáta z google maps [d] boli použité pre upresnenie hranice lokalít. Aplikácia používa dáta z misií Landsat a Copernicus v prípade náhľadu na územie z väčšej vzdialenosti a v prípade náhľadu z menšej vzdialenosti používa letecké snímky. V prípade územia ČR používa letecké snímky od firmy GEODIS BRNO spol. s r.o. z roku 2017 zo zimného a letného obdobia. Aplikácia google maps je v používaní snímok z letného a zimného obdobia dosť chaotická.

V tejto práci pre upresnenie hraníc lokalít bol použitý náhľad, v ktorom boli použité letecké snímky. V tejto práci bolo rozhodnuté, že sa dajú presnejšie určiť hranice lokalít z leteckých snímok, ako zo snímok z Landsat-8, alebo Sentinel-2 (Obrázok 11)

Niektor by mohol tvrdiť, že nie je vhodné porovnávať dáta z LhpoMap z roku 2016 a dáta z google maps z roku 2017, pretože v čase medzi získaním dat do oboch aplikácií, mohol časť lesa niekto vyrúbať. Toto riziko je ošetrené tým, že z miesta, kde bol vyrúbaný les sa za jeden rok nestane lúka. Pretože les, ktorý rastie na danom území napr. 50 rokov, neobsahuje semennú banku lúčnych druhov rastlín.



Obrázok 11: Úprava hraníc lokality Přešednice podľa leteckých snímok z google maps



**Legenda:**

- Upravené územie lokality
- Pôvodné územie lokality

A – 4. kanál snímky z Landsat-8 z 24.6.2015

B – 4. kanál snímky z Snetinel-2 z 4.7.2015

C – Vrstva druhového zloženia lesov z LphoMap z roku 2016

D – Google maps, letecká snímky, 2017

*Autor: vlastné*

#### 4.5 Dáta z družíc

Na aplikáciu danej metódy sa využijú dáta s multispektrálnych družíc Landsat-8 a Sentinel-2. Ako vhodné územie pre výskum boli vybrané lesné porasty v Usteckom, Karlovarskom, Plzeňskom a Stredočeskom kraji v Českej republike.

Snímky z Landsat-8 je možné získať napr. od U.S. Geological Survey [a]. Snímky Sentinel-2 je možné zadarmo stiahnuť z webovej stránky Európskej vesmírnej agentúry z portálu Sentinel-2 Pre-Operations Data Hub [b].

Podklady pre kontrolu správnosti klasifikácie sú získané z mapového portálu LhpoMap od UHUL. Aktuálnosť údajov sa vzťahuje k roku 2016.

#### 4.5.1 Landsat-8

Družica Landsat-8 je súčasťou vesmírneho programu Landsat, na ktorom spolupracujú NASA (National Aeronautics and Space Administration) a USGS (United States Geological Survey). Družica začala naplno pracovať na orbite Zeme 30.5.2013. Plán poskytovania snímok družicou Landsat-8 sa predpokladá na 5 až 10 rokov.

Landsat-8 sníma dané územie raz za 16 dní a sníma s 8 dňovým posunom oproti družici Landsat-7. Družica disponuje dvomi „push-broom“ senzormi OLI (Operational Land Imager) a TIRS (Thermal Infrared Sensor). V tejto práci boli použité dáta zo senzoru OLI. [48]

Vlnové dĺžky v jednotlivých kanáloch boli v senzore nastavené tak, aby senzor snímал vegetáciu v dôležitých úsekoch vegetačnej krivky. Dôležitých v zmysle, aby sa z hodnoty odrazenia slnečného žiarenia dalo o vegetácii pokrývajúci daný pixel zistiť čo najviac. [49]

B kanál je vhodný pre batimetrické mapovanie. Teda je vhodný pre rozlišovanie pôdy od vegetácie a tiež pre rozlišovanie listnatej od ihličnatej vegetácie. G kanál je najvhodnejšie využiť, keď je vegetácia plne rozvinutá, inak povedané, keď dosahuje svoj vrchol. Tento kanál je vhodný pre posúdenie sily rastlín. R kanál je vhodný pre rozlišovanie vegetácie na svahoch a hodnotenie stability svahu. NIR kanál je vhodný pre skúmanie množstva biomasy. SWIR1 a SWIR2 pásma sú vhodné pre rozlišovanie vlhkosti pôdy a vegetácie. Všeobecný prehľad vlastností kanálov družice Landsat-8 je v *Prílohe 1.3 a 1.4*)

#### Úroveň predspracovania dát

Zo snímok ktoré boli urobené z družice Landsat-8 sa za mesiac júl v rokoch 2015 až 2017 nedala použiť žiadna snímka kvôli vysokej oblačnosti. Preto boli hľadané vhodné snímky z mesiacov jún a august v rokoch 2015 až 2017, kedy bola tiež vegetácia pomerne dobre rozvinutá. Boli nájdené tri vhodné snímky a to z dní: 24.6.2016, 27.8.2016 a 30.8.2017.

Snímky boli stiahnuté z EarthExplorer od USGS v dvoch verziách predspracovania v Level-1 a v Level-2.

Dáta predspracované na úrovni Level-1 (L1TP) obsahujú rádiometrickú korekciu. Pri rádiometrickej kalibrácii aj ortorektifikácii boli použité pozemné kontrolné body a digitálny výškový model (DEM). Dáta L1TP sú dáta s najvyššou kvalitou predspracovania dát poskytovateľom na úrovni Level-1 pre pixelovú analýzu.

Dáta predspracované na úrovni Level-2 obsahujú oproti dátam predspracovaným na úrovni Level-1 (L1TP) aj atmosférickú korekciu.

#### 4.5.2 Sentinel-2

Sentinel-2 je misia pozorovania Zeme, ktorá je súčasťou vesmírneho programu Copernicus, za ktorým je ESA (European Space Agency). Misia Sentinel-2 pozostáva z dvoch rovnakých družíc a to Sentinel-2A a Sentinel-2B. Sentinel-2A začala pracovať 23.6.2015 a Sentinel-2B 7.3.2017. V prípade všetkých použitých snímok z misie Sentinel-2 išlo o dáta z družice Sentinel-2A. Životnosť jednej družice sa odhaduje na 7 rokov. Sentinel-2 sníma Zemský povrch na jednom mieste s intervalom 5 dní.

Pre rozšírenie možnosti skúmania charakteristík vegetácie družica Sentinel-2 okrem tzv. klasických pásiem (B, G, R a NIR) sníma Zemský povrch aj pásmach Vegetation Red Edge 1 (VRE1), Vegetation Red Edge 2 (VRE2), Vegetation Red Edge 3 (VRE3) a Narrow NIR (viď. *Prílohy 1.3 a 1.4*). [50]

##### Úroveň predspracovania dát

Podobne ako v prípade dát z Landsat-8 aj v prípade dát zo Sentinel-2, nebolo možné pre vysokú oblačnosť použiť žiadnu snímku z mesiaca júl z rokov 2015 až 2017. Preto boli v tejto práci využité snímky z mesiacov jún a august z rokov 2015 až 2017. Boli nájdené tri vhodné snímky z dátumov: 4.7.2015, 13.8.2015 a 27.8.2016.

Snímky použité v tejto práci boli stiahnuté z portálu ESA. Boli stiahnuté vo verzii predspracovania Level-1C. Úroveň predspracovania dát Level-1C obsahuje rádiometrickú a geometrickú korekciu vrátane ortorektifikácie a priestorovej registrácie na globálnom referenčnom systéme so sub-pixelovou presnosťou. Údaje sú v DN hodnotách. DN hodnoty sú hodnoty odrazivosti namerané nad atmosférou (TOA Reflectance, Top of Atmosphere Reflectance) vynásobené číslom 10 000. [51]

Tvorba dát na úrovni Level-1C pozostáva z niekoľkých krokov:

- Združenie dlaždíc: výber vopred pred definovaných dlaždíc pretínajúcich stopu požadovaného obrázka
- Výpočet prevzorkovania mriežky: umožňuje prepojenie obrazu natívnej geometrie s obrazom celkovej geometrie (ortorektifikovaný obraz)
- Prevzorkovanie každého spektrálneho pásma v geometrii orto obrazu pomocou mriežky prevzorkovania a interpolačného filtra. V tomto kroku sa tiež nachádza výpočet TOA Reflectance (odrazivosť meraná na vrchu atmosféry)
- Tvorba masiek pre oblaky a vodu/zem
- Kompresia výsledných snímok pre Level-1C pomocou algoritmu JPEG2000 a záhlavie kódované geografickými obrázkami GML

### 4.5.3 Použité scény

V tabuľke nižšie je možné vidieť základné informácie o 6 scénach použitých v tejto práci. V tabuľke sa nachádza 9 scén, pretože scény z družice Landsat-8 sú v nej spomenuté 2-krát, pre každý level spracovania dát 1-krát.

Tabuľka 9: Základný prehľad informácií o scénach

Číslo	Družica	Level predspracovania dát	Dátum	Oblačnosť [%]	Označenie v práci
1	Landsat-8	Level-1TP	24.6.2016	0,4	L8_L1_160624
2	Landsat-8	Level-1TP	27.8.2016	0,02	L8_L1_160827
3	Landsat-8	Level-1TP	30.8.2017	0,04	L8_L1_170830
4	Landsat-8	Level-2	24.6.2016	0,4	L8_L2_160624
5	Landsat-8	Level-2	27.8.2016	0,02	L8_L2_160827
6	Landsat-8	Level-2	30.8.2017	0,04	L8_L2_170830
7	Sentinel-2	Level-1C	4.7.2015	0,47	S2_L1_150704
8	Sentinel-2	Level-1C	13.8.2015	1,93	S2_L1_150813
9	Sentinel-2	Level-1C	27.8.2016	0	S2_L1_160827

*Autor: vlastný*

### Očakávané problémy so snímkami z družíc Landsat-8 a Sentinel-2

Pixely u snímok z družíc Landsat-8 a Sentinel-2 sú dosť veľké 15 a 30 m a 10 až 20 m. Na týchto snímkach nie je možné identifikovať menšie diery (5 x 5 m) v lesnom poraste, vďaka ktorým senzor zachytáva podrast pod stromovou etážou. To spôsobuje spektrálnu nepresnosť pri analýze snímok a je jednou z príčin zníženej celkovej presnosti klasifikácie.

## 5 POSTUP

V tejto kapitole bol popísaný postup práce, ktorý bol popísaný v 3 podkapitolách. V prvej podkapitole bol opísaný proces predspracovania dát. V druhej podkapitole bol opísaný klasifikačný proces. V tretej podkapitole bol popísaný dopad výberu hodnôt pre parametre klasifikátora na celkovú presnosť klasifikácie.

Celá praktická časť práce bola robená v softvéroch ArcMap (verzia 10.5.1, jún 2017) a ArcGIS Pro (verzia 2.1.2, február 2018) od ESRI.

### 5.1 Predspracovanie dát

Predtým, ako mohlo dôjsť k samotnej klasifikácii museli byť data predspracované a upravené do vhodného formátu. V tejto podkapitole je opísaný proces predspracovania dát.

#### 5.1.1 Georeferencovanie

Na georeferencovanie bol použitý softvér ArcMap 10.5.1.

Pre presnejšie určenie hraníc lokalít lesných porastových jednotiek boli georeferencované prinskríny lesných máp z UHULu a prinskríny z Google maps na základe snímok z družice Landsat-8.

Dubové, bukové, smrekové a borovicové lesy majú rôzne vlastnosti a preto aj georeferencovanie bolo prispôbené týmto vlastnostiam.

V prípade polovice lokalít dubu nebolo potrebné georeferencovať žiadne mapy, pretože homogénne dubové lesy sa častokrát vyskytujú v rovinnom teréne v údolí a tvoria samostatné lesné celky ohraničené najčastejšie lúkami. Preto v polovici prípadov na presnú lokalizáciu stačila WMS vrstva Základnej mapy v mierke 1:10 000. Hranica polygónov lokalít nie je totožná so skutočnou hranicou lesa, ale je posunutá o približne 20 a viac metrov do vnútra lesa, aby krajné pixely v polygónoch neobsahovali hodnoty odrazivosti z lúky. V prípade hraničenia dubového lesa so zmiešaným lesom je hranica polygónu dubového lesa posunutá o 60 a viac metrov do vnútra dubového lesa. V prípade druhej polovice lokalít bolo potrebné georeferencovanie prinskrínov lesníckych máp z UHUL, pretože tieto lokality z veľkej časti hraničili so zmiešanými lesmi a nebolo možné určiť presnú hranicu dubového lesu z WMS vrstvy Základnej mapy 1:10 000.

V prípade bukových lesov boli georeferencované prinskríny lesných máp z UHUL, pretože bukové lesy sa vo väčšine prípadov nachádzajú na svahoch pohorí. Kde s nimi väčšinou susedí zmiešaný les.

V prípade smrekových lesov bolo potrebné na lokalizáciu lesných jednotiek georeferencovať prinskríny z Google maps, pretože v lesníckych mapách UHULu sa nachádzajú na pohoriach veľké smrekové plochy. Ale v skutočnosti v mnohých

prípadoch je tam veľmi mladý les, kde by na družicových snímkach veľmi výrazne presvitala pôda. Mladý les bolo potrebné pre účel tejto práce eliminovať.

V prípade borovicových lesov bolo potrebné georeferencovať prinskrióny vrstvy druhového zloženia lesov z aplikácie LhpoMap, pretože borovicové porasty sa nachádzali v susedstve s inými lesnými porastmi. Borovicové porasty netvorila rozsiahle plochy. Sú roztrúsené po celom území ČR.

### **5.1.2 Mozaikovanie scén**

Družica Sentinel-2 zaznamenáva územie o veľkosti 290 x 274 km. Každá snímka je rozdelená na 15 dlaždíc. Každá dlaždica má rozlohu 100 x 100 km. Celé výskumné územie sa rozprestiera na viacerých dlaždiciach a preto je potrebné urobiť mozaikovanie.

Jednotlivé dlaždice scény sa navzájom z časti prekrývajú. Na území, kde je prekryv aspoň 2 scén vidieť, že pixely majú rovnaké DN hodnoty u oboch dlaždíc. Preto pri spájaní scén nie je potrebné robiť prevzorkovanie. Pre účely tejto práce bolo potrebné mozaikovať 4 dlaždice v prípade každého kanála a u každej scény.

### **5.1.3 Formátovanie údajov**

Aby na údaje zo snímkov mohol byť aplikovaný klasifikačný proces, tak dáta boli upravené. Nad každou lokalitou bola urobená polygónová sieť. Nad snímkami z Landsat-8 každý polygónový záznam mal rozmer 30 x 30 m a v prípade Snetinel-2 mal 10 x 10 m.

Z pôvodných snímkov boli urobené výrezy lokalít a tieto výrezy boli pretransformované z rastrovej do polygónovej vrstvy. Keďže niektoré susedné pixely mali vo všetkých pásmach rovnaké DN hodnoty, tak pri transformácii rastrového obrazu na vektorový obraz dochádzalo k tomu, že dve až tri pixely tvorili jeden polygón. Preto bolo potrebné urobiť sieť (angl. Fishnet) spomenutú vyššie. V následnom kroku sieť pomohla rozdeliť spojené pixely do jedného polygónu na sieť, v ktorej každý polygón nadobudol rozmery pixelov snímky príslušnej družice.

Ďalším krokom bolo vypočítať rozlohu jednotlivých pixelov, ak polygón nedosahoval veľkosť rozlohy pixela t.j. u pixela Landsat-8 900 m<sup>2</sup> a v prípade pixela Sentinel-2 100 m<sup>2</sup>, tak následne tie polygóny boli odstránené. Išlo totiž o polygóny, ktoré ležali na hraniciach lokalít a tie pixely mohli obsahovať aj hodnoty zo susednej základnej porastovej jednoty t.j. napr. lúky, alebo vegetácie s iným druhom zloženia drevín.

Výsledná polygónová vrstva obsahovala v riadkoch záznamy za jednotlivé pixely a v stĺpcoch atribúty, ktoré sú popísané v nasledujúcej časti práce 5.2.4. Údaje z každej scény boli pretransformované do samostatnej polygónovej vrstvy.

#### 5.1.4 Príprava údajov pre použitie funkcie *RandomForestClassifier*

Na základe príkladov [23] [25] a [24] boli identifikované dve podmienky na použitie funkcie „RandomForestClassifier“ a to:

1. Doinštalovanie potrebných knižníc (scikit-learn a seaborn) do aplikácie ArcGIS Pro, ako je uvedené v zdroji [24]
2. Príprava údajov pre výpočet. Údaje majú mať formu atribútovej tabuľky, ktorá sa v novo vytvorenom projekte vloží ako vrstva.

Atribútová tabuľka musí mať tieto potrebné atribúty:

- Atribút vyjadrujúci polohu typu „Point“. (Ak „Shape“ je typu „Polygon“ je nutné ho prekonvertovať na typ „Point“ pre použitie funkcie „FeatureClassToNumPyArray“.)
- Atribúty vyjadrujúce jednotlivé kanály družice ('blue', 'red', 'NIR', 'green'), údaje sú v DN hodnotách
- Atribút vyjadrujúci druh stromu. V práci boli použité čísla 1 až 4 ('Dub', 'Buk', 'Smrek', 'Borovica').

Pre ďalšiu prácu s údajmi v aplikácii ArcGIS Pro boli použité ďalšie atribúty:

- Názov lokality
- Číslo lokality v tejto práci (1 až 24)

## 5.2 Klasifikačný proces

V poslednom období je značný nárast využívania a vývoja metód „učiacich sa strojov“ a „umelej inteligencie“. To sa odráža ako určitá nekonzistentnosť vo výsledkoch rešeršovaných prác, ako aj v ich záveroch a odporúčaní.

Tiež paralelný vývoj programovacích knižníc od rôznych spoločností na jednej strane posúva ich vývoj dopredu, ale na druhej strane neumožňuje generalizovať odporúčania z rešeršovaných prác. Napríklad parametre funkcie RandomForestClassifier, ktorá je použitá v tejto práci a je obsiahnutá v knižnici „scikit-learn“ verzia 0.19.1 sú odlišné od tej istej funkcie obsiahnutej vo verzii knižnici 0.15. (neobsahuje parametre *min\_weight\_fraction\_leaf = 0.0*, *min\_impurity\_decrease = 0.0*, *min\_impurity\_split = None*).

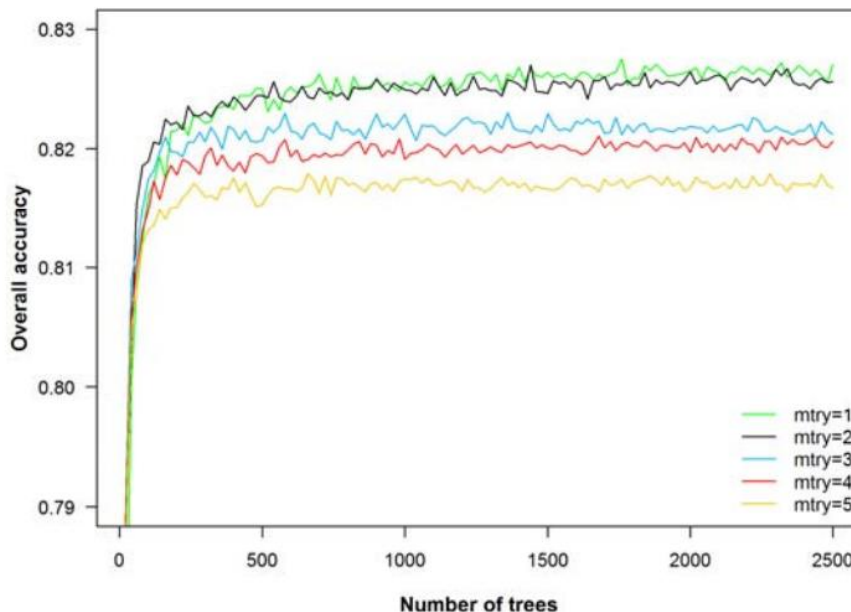
### 5.2.1 Nastavenie parametrov funkcie *RandomForestClassifier*

Pri použití funkcie *RandomForestClassifier* je možné nastaviť 17 parametrov tak, ako je uvedené v opise funkcie 3.2.4.3.1. *sklearn.ensemble.RandomForestClassifier* [21]. V tomto opise funkcie je uvedené, že je možné použiť všetky parametre s predvolenými hodnotami, ktoré maximálne využijú vlastnosti funkcie. Tiež je uvedené, že takto nastavené parametre vyžadujú vysoký výpočtový výkon. Toto posledné tvrdenie pri použití počítača sa neprejavilo (procesor i7-7500 2,7Ghz a pamäť 8Gb). Pri ďalšom zdroji [52] je uvedené, že zmena nárokov na výpočtový výkon a operačnú pamäť bola výrazne optimalizovaná zmenou verzie jadra z 3.13.0-57

na 4.4.0-28. Na základe tejto informácie boli porovnávané verzie dokumentácie k použitiu RF. V repozitári [21] je najstaršia dokumentácia z roku 2014 t.j. verzia 0.15. **Vývoj nástrojov pre „učiace sa stroje“ môže čiastočne vysvetliť nekonzistentnosť výsledkov v uvedenej literatúre, ako aj používanie rôznych knižníc pre programovanie v jazykoch Python a jazyka „R“.**

Immitzer M. a kol. (2012) [4] opisuje nastavenie parametra *mtry*, ktorý má vplyv na dosiahnutú celkovú presnosť klasifikácie (Obrázok 12: Immitzer M. a kol. (2012) Figure 5 ) na základe počtu stromov a počtu náhodného rozdelenia premenných v každom bode (*mtry*) pri použití RF klasifikácie 10 druhov stromov v ôsmich pásmach.

Obrázok 12: Immitzer M. a kol. (2012) Figure 5 [1]



Takýto parameter nie je uvedený vo funkcii „RandomForestClassifier“, ale miesto neho sú parametre, ktoré určujú minimálnu a maximálnu hodnotu [21]. Ale pri ďalšom skúmaní bolo zistené, že daný parameter odpovedá použitiu funkcie „RandomForest“ v programovacom jazyku „R“ [18] a [19] (viď. *Tabuľka 4: Porovnanie názvov kľúčových parametrov RF pre túto prácu medzi jazykom "R" a Python v podkapitole 2.9*)

Nastaveniu parametrov sa podrobne venuje článok „Tuning a Random Forest Classifier“ [53] kde je podrobne vysvetlené, ako vplyva nastavenie parametrov na výslednú dosiahnutú presnosť.

V literatúre sa viackrát objavuje upozornenie na efekt „overfitting“, ktorý sa objavuje pri prílišnej snahe o dosiahnutie vyššej presnosti.

*Poznámka: „Overfitting je štatistický jav, ktorý vzniká pri tvorení analýzy, ktorá chce byť príliš detailná, alebo presná vzhľadom k určitému súboru údajov a preto zlyháva pri analýze ďalších údajov a ich spoľahlivej predpovedi.“* [54]

Nastavenie parametrov na základe článku [53] a [4] , ktoré boli uvažované v tejto práci:

- `n_estimators`



- criterion
- bootstrap
- oob\_score

Nasledujúce parametre neslúžia ani tak na získanie lepšieho výsledku, ako na zabránenie javu „overfitting“. V tejto práci aj tieto parametre boli menené za účelom overenia či nebude dosiahnutý lepší výsledok.

- max\_features
- max\_depth
- min\_samples\_split
- min\_samples\_leaf

Ostatné parametre slúžia na nastavenie výkonnosti n\_jobs (počet použitých procesorov) na ladenie procesu RF random\_state, verbose, warm\_start či class\_weight.

## 5.2.2 Nastavenie parametrov RF podľa literatúry

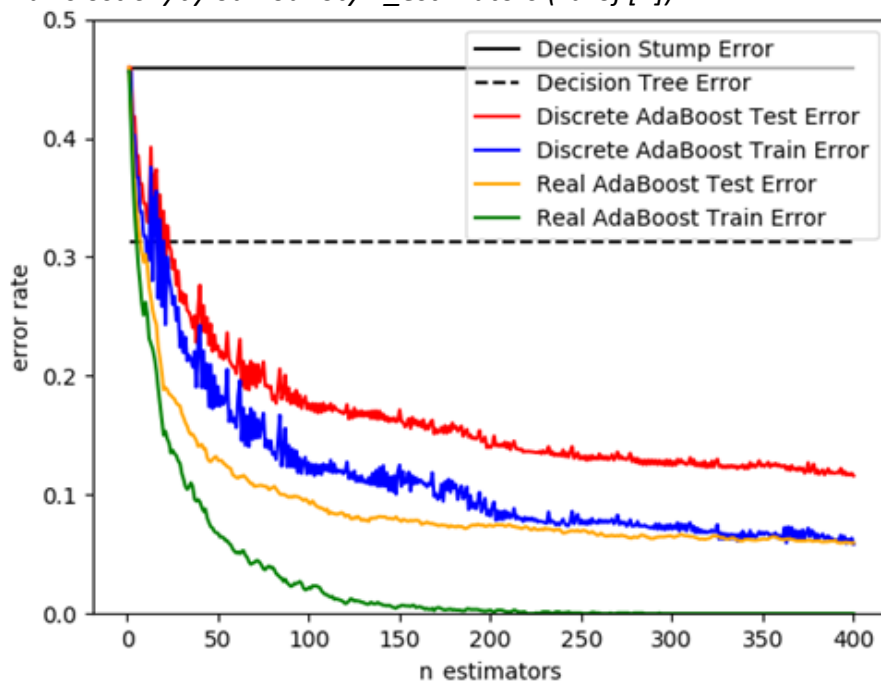
### 5.2.2.1 *n\_estimators*

(*n\_estimators* : integer, optional (default=10))

Podľa literatúry čím je hodnota *n\_estimators* väčšia, tým vyššiu presnosť klasifikácie je dosiahnutá, ale je potrebné dať pozor na „Overfitting“. V literárnej rešerši bola najčastejšie použitá hodnota 500.

V článku „1.11.3. AdaBoost“ [21] je uvedený spôsob testovania a výsledný graf z ktorého vyplýva, že po prekročení hodnoty 200 parametra *n\_estimators* dosiahnutá presnosť začína saturovať.

Graf č. 1: Závislosť chyby od hodnoty *n\_estimators* (zdroj [1])



V tejto práci bola použitá hodnota **500**.

### 5.2.2.2 *criterion*

(*criterion-(string)-Default = "gini"*) podľa [53] hodnota *gini* je vhodná pre väčšinu typov údajov. Zo zdroja HOW THE RANDOM FOREST ALGORITHM WORKS IN MACHINE LEARNING [4] vyplýva, že pre údaje, ktoré sú spojité je lepšie použiť hodnotu *entropy* (t.j. Information Gain – pokus predpovedať ďalší údaj na základe predchádzajúcej hodnoty) a pre diskkrétne údaje je lepšie použiť hodnotu *gini* (*gini* je miera štatistickej disperzie).

V práci bola použitá hodnota **entropy**.

### 5.2.2.3 *bootstrap*

(*bootstrap-(boolean)-Default=True*) V článku T. Plapinger, „Tuning a Random Forest Classifier,“ (12. 8. 2017) [53] je uvedené, že „bootstrap“ automaticky (samostatne) vytvára podmnožiny pixelov z celej trénovacej množiny pixelov pre jednotlivé stromy v lese (angl. trees in forest). V literatúre je uvedené že „je to niečo, čo by si nikdy nemal meniť z prednastavenej hodnoty, pretože model bude dosahovať lepšie výsledky, ak je zapnutý *bootstrap*. Ak by bol *bootstrap* vypnutý, tak posledné rozhodnutie daný strom robí z inej podmnožiny pixelov.

V práci bola použitá prednastavená hodnota **True**.

*Poznámka: Podmnožinou pixelov je v tejto práci označovaná konkrétna podmnožina pre daný strom (angl. tree) trénovacej množiny pixelov.*

### 5.2.2.4 *oob\_score*

(*oob\_score-(boolean)-Default=False*) tento parameter slúži na zapnutie, alebo vypnutie krížového overovania (angl. cross validation).

Podľa zdroja [53], ako aj ostatných zdrojov, kde boli príklady použitia RF je lepšie ak hodnota parametra je nastavená na *True*. Zjednodušenie povedané, hodnota slúži na validáciu výsledku tým, že model je ladený na  $n-1$  trénovacích údajoch a ten posledný je nechaný na kontrolu výsledku.

Vysvetlenie vhodnosti použitia parametra je v kapitole 3.4.5

V práci bola použitá hodnota **True**.

### 5.2.2.5 *max\_features*

(*max\_features-(integer, float, string, or None)-Default="auto"*) podľa [53] je toto najkomplikovanejší parameter. Je veľa možností, ako nastaviť jeho hodnoty, ale z literatúry vyplýva že okrem výslednej presnosti má efekt na rýchlosť výpočtu a „overfitting“. V prípade hodnoty *auto* počíta z hodnotou *sqrt*, čo znamená, že nastaví hodnotu  $max\_features = (n\_estimators)^2$ . Čím je hodnota parametra väčšia, tým je

viacej rozhodovacích procesov v strome a tým vyššia presnosť, ale pri veľkej hodnote  $n\_estimators$  sa môže prejaviť overfitting.

V práci bola použitá prednastavená hodnota **auto**.

#### 5.2.2.6 *max\_depth*

(*max\_depth*-(integer or none)- Default=None) Podľa zdrojov [21] a [53] tento parameter určuje maximálne na koľko rozhodovacích vetiev sa člení rozhodovací strom. Parameter má za úlohu zabrániť efektu „overfitting“. Hodnota *None* určuje že algoritmus použije toľko vetvenia, koľko uzná za potrebné.

V práci bola použitá prednastavená hodnota **None**.

#### 5.2.2.7 *min\_samples\_split*

(*min\_samples\_split*-(integer, float)-Default=2) Podľa zdrojov [21] a [53] hodnota určuje minimálny počet vzoriek (v tejto práci pixelov), ktoré musia byť z trénovacích podmnožine údajov prítomné, aby došlo ešte k rozdeleniu (vetveniu stromu).

V práci bola použitá prednastavená hodnota **2**.

#### 5.2.2.8 *min\_samples\_leaf*

(*min\_samples\_leaf*-(integer, float)-Default=1) Podľa [21] a [53] hodnota určuje minimálne vetvenie rozhodovacieho stromu. Tento parameter je opačný parameter k parametru *max\_depth*. Pri vetvení rozhodovacieho stromu sa koncový rozhodovací uzol (node) tiež nazýva list (leaf).

V práci bola použitá prednastavená hodnota **1**.

### 5.2.3 Overenie nastavenie parametrov RF

Na overenie nastavenia niektorých parametrov podľa vyššie uvedeného opisu bol použitý program, ktorý v cykle menil hodnoty, ktoré dával do výstupu. Na základe týchto hodnôt boli zostavené tabuľky a grafy, ktoré slúžili na nastavenie parametrov pre výslednú klasifikáciu druhov stromov.

Overenie prebiehalo na údajoch získaných z družicovej snímky L8\_L1\_160624 (označenie jednotlivých snímok je v *Tabuľka 9: Základný prehľad informácií o scénach*). Meranie prebehlo viackrát, vždy s inou trénovacou a kontrolou množinou pixelov, ktorá bola generovaná, ako náhodné body z celkovej množiny pixelov. Ďalšie nastavenie parametrov bolo nasledovné:

- $n\_estimators = 500$
- $oob\_score = „True“$
- Ostatné hodnoty parametrov boli prednastavené (tieto hodnoty sú uvedené v kapitole 6.1)

Zdrojové kódy použitého programu sú v *Prílohe 3*. Ako program na získanie údajov pri zmene nastavenia *min\_samples\_leaf* a *min\_samples\_split* bol upravený program na zmenu  $n\_estimators$ .

(názov súboru v elektronickej prílohe – „výpočet vplyvu  $n\_estimators$  na presnost.py“)

### 5.2.3.1 Parameter: $n\_estimators$

Pri skúmaní vplyvu hodnoty  $n\_estimators$  na výpočet presnosti bolo urobené porovnanie zo všetkých skúmaných skupín údajov (podľa družice a dátumu snímky). Údaje sú zhrnuté v tabuľke a nasledovných grafov. Hodnoty parametrov z literárnej rešerše, v ktorej bolo skúmané druhové zloženie lesov aj hodnoty parametrov z rešerše, v ktorej bolo skúmané použité RF v jazyku Python v prostredí ArcGIS Pro sú iné, ako hodnoty parametrov uvedené v práci [21], v ktorej sa opisuje, ako ladiť parametre.

Pri dátach z Lansat-8 sa prejavuje „overfitting“ omnoho viac, ako v prípade dát Sentinel-2. Pravdepodobne má na to vplyv veľkosť trénovacej množiny vzhľadom k hodnote  $n\_estimators$  (Lansat-8 = 2560 pixelov, Sentinel-2 = 27478 pixelov).

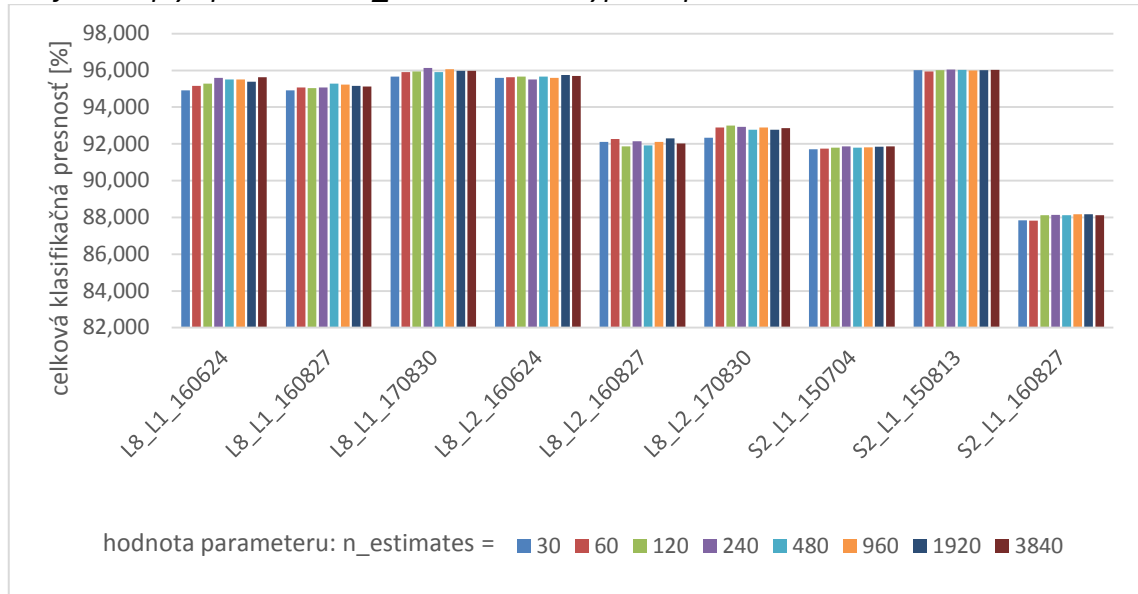
Na základe dosiahnutých výsledkov bola v tejto práci hodnota parametra  $n\_estimators$  vybratá 500.

Výsledky celkovej klasifikačnej presnosti z klasifikácii, v ktorých sa menili hodnoty parametra  $n\_estimators$  sú v tabuľke (Tabuľka 10). V nasledujúcom grafe (Graf č. 2) je možné vidieť tento vplyv za jednotlivé snímky.

Tabuľka 10: Vplyv parametra  $n\_estimators$  na výpočet presnosti [%]

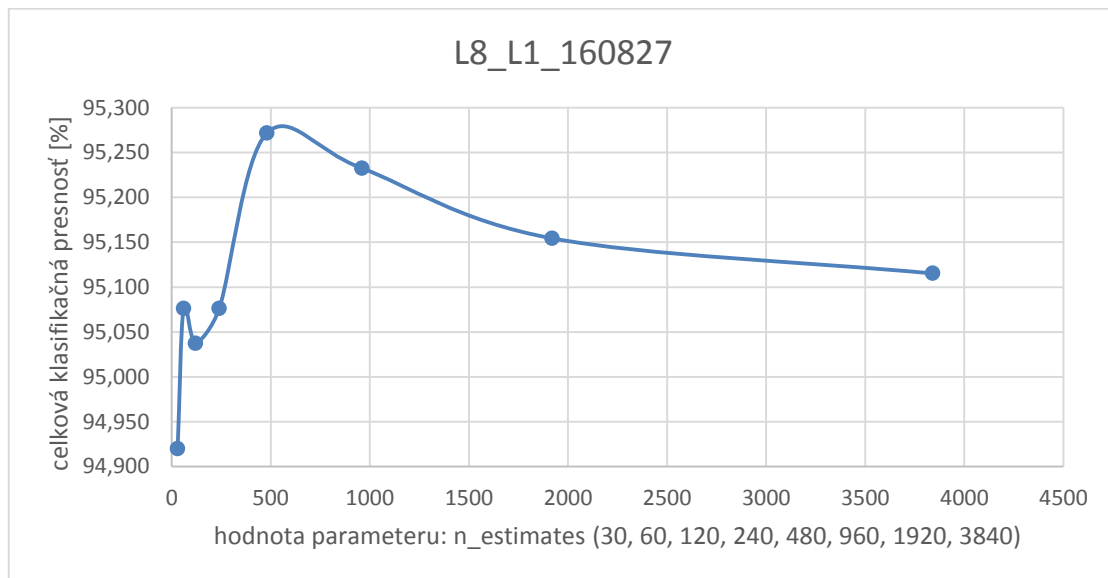
Použité snímky z:	hodnota parametra: $n\_estimators$							
	30	60	120	240	480	960	1920	3840
L8_L1_160624	94,920	95,154	95,272	95,584	95,506	95,506	95,389	95,623
L8_L1_160827	94,920	95,076	95,037	95,076	95,272	95,233	95,154	95,115
L8_L1_170830	95,661	95,897	95,937	96,134	95,897	96,055	95,976	95,976
L8_L2_160624	95,584	95,623	95,662	95,506	95,662	95,584	95,741	95,701
L8_L2_160827	92,106	92,263	91,872	92,145	91,911	92,106	92,302	92,028
L8_L2_170830	92,341	92,888	93,005	92,927	92,771	92,888	92,771	92,849
S2_L1_150704	91,703	91,746	91,790	91,870	91,801	91,819	91,845	91,870
S2_L1_150813	96,004	95,939	96,008	96,048	96,019	95,997	96,011	96,033
S2_L1_160827	87,834	87,820	88,125	88,140	88,118	88,166	88,169	88,115

Graf č. 2: Vplyv parametra  $n\_estimators$  na výpočet presnosti



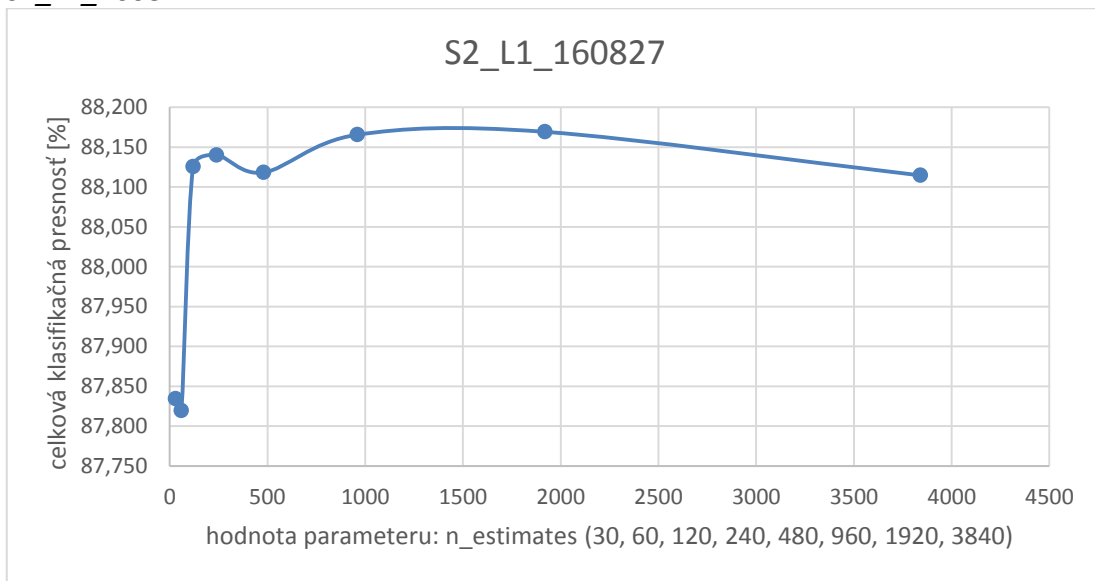
Nasledujúce dva grafy (Graf č. 3 a Graf č. 4) zobrazujú detailne zmenu presnosti na základe zmeny hodnôt  $n\_estimators$  z dvoch vybraných snímkov. Tabuľka 10 slúži, ako zdroj údajov pre uvedené grafy.

Graf č. 3: Vplyv parametra  $n\_estimators$  na výpočet presnosti údajov zo scény L8\_L1\_160827



Pri nízkej hodnote parametra  $n\_estimators$  (hodnoty medzi 30 a 120) a pri tej istej trénovacej množine výsledná celková presnosť klasifikácie kolíše (v rozmedzí 0,2%), čo môže byť dôsledok vytvorenia nedostatočného modelu (rozhodovacích stromov), ktorý by správne klasifikoval druhy drevín.

Graf č. 4: Vplyv parametra  $n\_estimators$  na výpočet presnosti údajov zo scény S2\_L1\_160827



### 5.2.3.2 Parameter: criterion

V tejto práci boli uvedené základné informácie o parametre *criterion* v podkapitole 5.2.2.2.

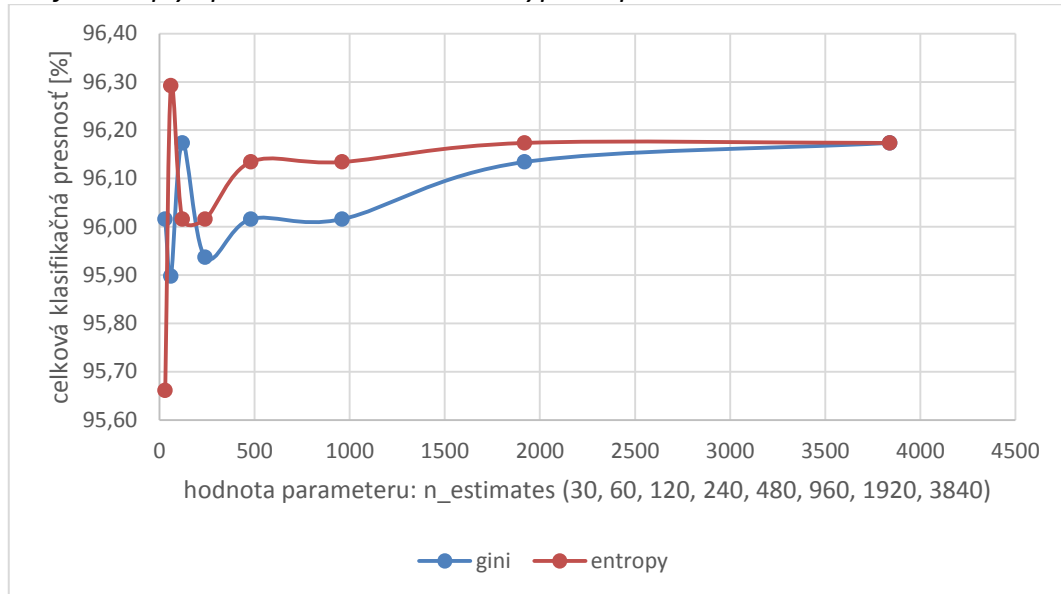
Pre dosiahnutie, čo najvyššej klasifikačnej presnosti bolo v tejto práci skúmané aj nastavenie parametra *criterion*. Nasledujúca tabuľka a graf ukazujú dosiahnutú rôznu celkovú klasifikačnú presnosť v závislosti od hodnoty  $n\_estimators$  pri odlišných hodnotách parametra *criterion* (*gini*, *entropy*).

Na základe dosiahnutých výsledkov testovania vzťahu  $n\_estimators$  a *criterion*, bolo pre klasifikáciu snímok vybratá hodnota **entropy**.

Tabuľka 11: Vplyv parametra *creation* na výpočet presnosti [%]

criterion	hodnota parametra: $n\_estimators$							
	30	60	120	240	480	960	1920	3840
<b>gini</b>	96,02	95,90	96,17	95,94	96,02	96,02	96,13	96,17
<b>entropy</b>	95,66	96,29	96,02	96,02	96,13	96,13	96,17	96,17

Graf č. 5: Vplyv parametra creation na výpočet presnosti



### 5.2.3.3 Parameter: min\_samples\_leaf

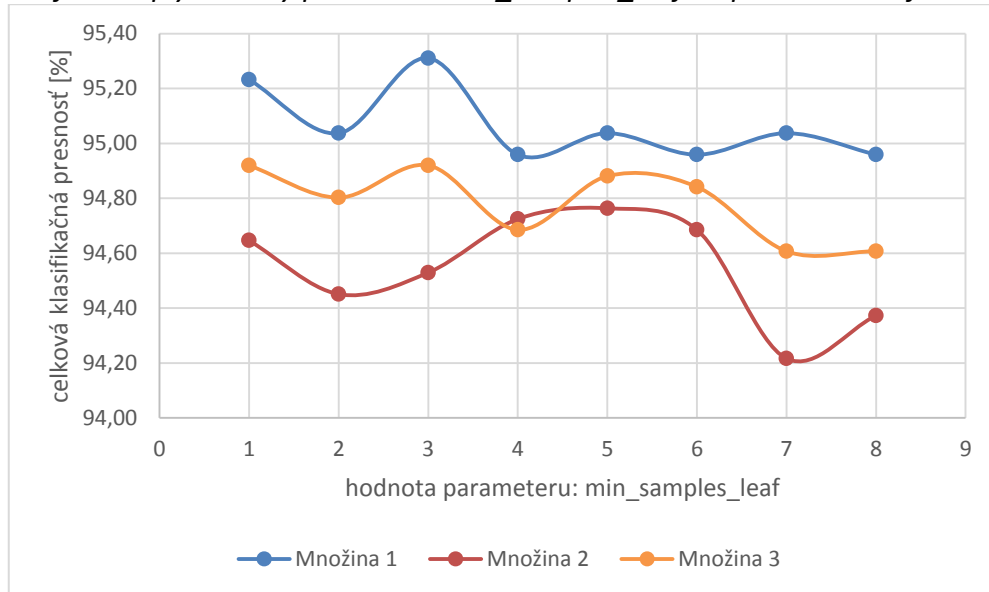
Na základe informácií uvedených v kapitole 5.2.2.8 v tejto práci bolo skúmané aj nastavenie parametra *min\_samples\_leaf* v dosiahnutí presnosti celkovej klasifikácie. Nasledujúca tabuľka a graf ukazujú zmenu dosiahnutej celkovej presnosti klasifikácie v závislosti od výberu odlišných hodnôt parametra *min\_samples\_leaf* (1 až 8). Pre klasifikáciu bolo použitých 500 rozhodovacích stromov (*n\_estimators*) a kritérium (*criterion*) *entropy*. Výsledok závisí aj od výberu trénovacej množiny pixelov, ktoré v tejto práci sú vyberané náhodne z celej množiny pixelov. Z toho dôvodu boli vykonané tri merania.

Na základe dosiahnutých výsledkov bolo v tejto práci ponechané základné nastavenie t.j. hodnota **1**.

Tabuľka 12: Vplyv zmeny parametra *min\_samples\_leaf* na presnosť klasifikácie [%]

	hodnota parametra: min_samples_leaf							
	1	2	3	4	5	6	7	8
<b>Množina 1</b>	95,23	95,04	95,31	94,96	95,04	94,96	95,04	94,96
<b>Množina 2</b>	94,65	94,45	94,53	94,72	94,76	94,69	94,22	94,37
<b>Množina 3</b>	94,92	94,80	94,92	94,69	94,88	94,84	94,61	94,61

Graf č. 6: Vplyv zmeny parametra *min\_samples\_leaf* na presnosť klasifikácie



#### 5.2.3.4 Parameter: *min\_samples\_split*

Na základe informácií uvedených v kapitole 5.2.2.7 v tejto práci bolo skúmané aj nastavenie parametra *min\_samples\_split* v dosiahnutí presnosti celkovej klasifikácie.

V tejto podkapitole je spísaná správa z testovania parametru *min\_samples\_split*. Urobila sa klasifikácia s nastaveniami hodnôt parametra *min\_samples\_split* 1 až 9. Pre klasifikáciu bolo použitých 500 rozhodovacích stromov ( $n_{estimators}$ ) a kritérium (*criterion*) *entropy*. Klasifikácia sa zopakovala pre tri rôzne množiny tréningových dát.

Na základe dosiahnutých výsledkov testovania parametra *min\_samples\_split* bolo pre klasifikáciu v kapitole 6 ponechané základné nastavenie t.j. hodnota 2, aj keď v niektorých prípadoch by bola lepšia hodnota 5, alebo 6. Pravdepodobne vplyv tohto parametra na klasifikáciu je dosť zložitý a miera jeho vplyvu závisí od výberu hodnôt iných parametrov. Rozdiel v dosiahnutej presnosti je vo všetkých prípadoch 0,2%.

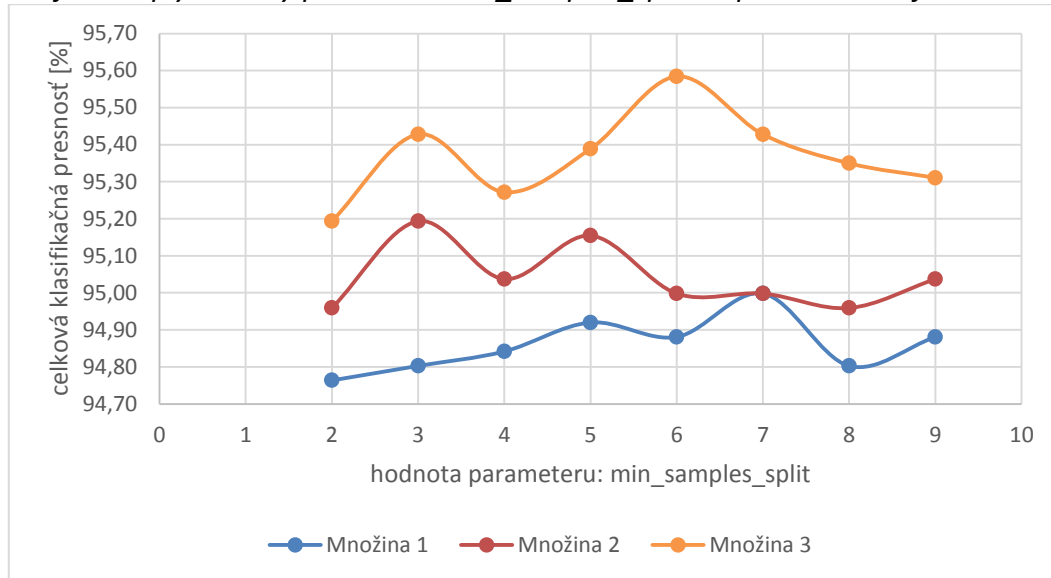
Z grafu je viditeľné, že tri rôzne tréningové množiny dávajú odlišné výsledky a nie je možné jednoznačne určiť závislosť dosiahnutej celkovej presnosti klasifikácie od nastavenej hodnoty parametra.

Tabuľka 13: Vplyv zmeny parametra *min\_samples\_split* na presnosť klasifikácie[%]

	hodnota parametra: min_samples_split							
	2	3	4	5	6	7	8	9
<b>Množina 1</b>	94,76	94,80	94,84	94,92	94,88	95,00	94,80	94,88
<b>Množina 2</b>	94,96	95,19	95,04	95,15	95,00	95,00	94,96	95,04
<b>Množina 3</b>	95,19	95,43	95,27	95,39	95,58	95,43	95,35	95,31



Graf č. 7: Vplyv zmeny parametra *min\_samples\_split* na presnosť klasifikácie



### 5.2.3.5 Overenie veľkosti trénovacej množiny na výslednú presnosť

Počas ladenia programu bola zisťovaná závislosť výslednej presnosti na veľkosti trénovacej množiny. Na základe získaných údajov je zrejmé, že čím je trénovacia množina reprezentatívnejšia t.j. čím má bližšie ku kvalifikovanému výberu, tým je dosiahnutá celková presnosť klasifikácie vyššia.

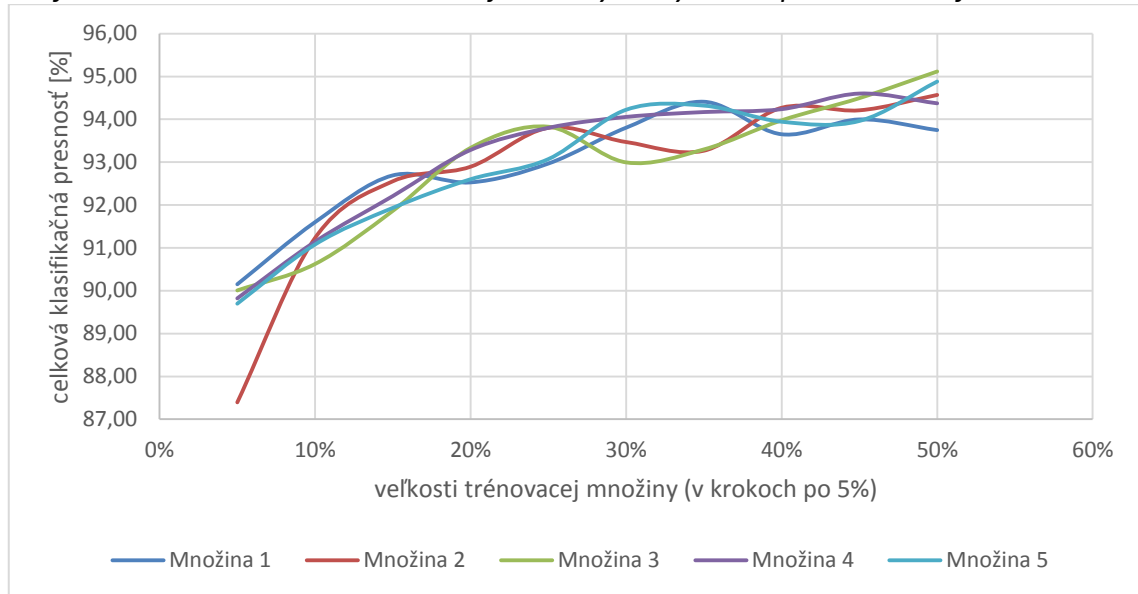
Zo získaných údajov vyplýva, že územie 138,24 [ha], ktoré predstavuje 30% z množiny všetkých skúmaných údajov je dostatočné a vyššia presnosť už nie je dosiahnutá. Samozrejme to závisí aj od výberu trénovacích pixelov, ktoré v tejto práci sú vyberané náhodne. Z toho dôvodu bolo vykonaných päť meraní.

Pri ladení veľkosti trénovacej množiny boli všetky parametre nastavené defaultne okrem *n\_estimators* (500) a *criterion* (*entropy*). V kapitole 6 bolo v klasifikácii pre trénovaciu množinu použitých 50 % pixelov celej množiny.

Tabuľka 14: Závislosť veľkosti trénovacej množiny na výslednú presnosť klasifikácie

Počet bodov	256	512	768	1024	1280	1536	1792	2048	2304	2560
%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
<b>Množina 1</b>	90,15	91,60	92,69	92,53	92,97	93,80	94,41	93,65	94,00	93,75
<b>Množina 2</b>	87,39	91,23	92,55	92,89	93,80	93,47	93,27	94,27	94,21	94,57
<b>Množina 3</b>	90,01	90,62	91,86	93,33	93,83	92,99	93,30	93,98	94,49	95,12
<b>Množina 4</b>	89,82	91,14	92,21	93,28	93,80	94,06	94,17	94,24	94,60	94,37
<b>Množina 5</b>	89,70	91,08	91,93	92,60	93,07	94,22	94,32	93,94	93,96	94,88

Graf č. 8: Závislosť veľkosti trénovacej množiny na výslednú presnosť klasifikácie



## 5.2.4 Opis algoritmu - programu použitý na klasifikáciu

(názov súboru v elektronickej prílohe – „výpočet vplyvu  $n\_estimators$  na presnosť.py“)

### 5.2.4.1 Príprava

Program na to, aby mohol používať funkcie musí najprv naimportovať knižnice pre RF, matematické, ArcGIS, práce s polom, štatistické a s operačným systémom.

```
from sklearn.ensemble import RandomForestClassifier
import numpy as NUM
import arcpy as ARCPY
import arcpy.da as DA
import pandas as PD
import seaborn as SEA
import matplotlib.pyplot as PLOT
import os as OS
```

### 5.2.4.2 Načítanie údajov

V ďalšom kroku je potrebné zdefinovať premenné a načítať údaje. Načítané údaje majú tvar matice (údaje sú vybrané z atribútovej tabuľky vektorovej vrstvy obsahujúcej údaje zo snímok) a prekonvertované do dvojrozmerného poľa.

```
# definovanie počiatočných premenných
inputFC = r'Vrstva_body'
#predictVars = ['gridcode', 'gridcode_1', 'gridcode_2', 'gridcode_3']
predictVars = [ 'blue', 'red', 'NIR', 'green', ]
classVar = ['Trieda'] # [trieda_stromu]
allVars = predictVars + classVar #mená všetkých atribútov

#pole trénovacích bodov zo všetkými atribútmi
trainFC = DA.FeatureClassToNumPyArray(inputFC, ["SHAPE@XY"] + allVars)
spatRef = ARCPY.Describe(inputFC).spatialReference

#namerané údaje
data = PD.DataFrame(trainFC, columns = allVars)
```

```
#zobrazenie prvych 3 a poslednych 3 udajov v matici „trainFC“, stĺpce odpovedajú
hodnotám ["SHAPE@XY"] + allVars
trainFC
```

**Výstup:**

```
array([[[-840100.7468, -994459.1536], 8243, 6780, 13404, 7672, 3),
       ([-859300.7468, -1067629.1536], 8629, 6917, 14655, 7897, 4),
       ([-821380.7468, -1062739.1536], 8659, 6816, 14444, 7808, 4),
       ...,
       ([-814690.7468, -983869.1536], 8459, 6608, 28359, 7816, 2),
       ([-858640.7468, -1067719.1536], 8598, 6876, 14876, 7940, 4),
       ([-869860.7468, -994009.1536], 8293, 6616, 14210, 7635, 3)],
      dtype=[('SHAPE@XY', '<f8', (2,)), ('blue', '<i4'), ('red', '<i4'), ('NIR',
'<i4'), ('green', '<i4'), ('Trieda', '<i4')])
```

```
#zobrazenie prvych x riadkov v tabuľke
data.head()
```

**Výstup:**

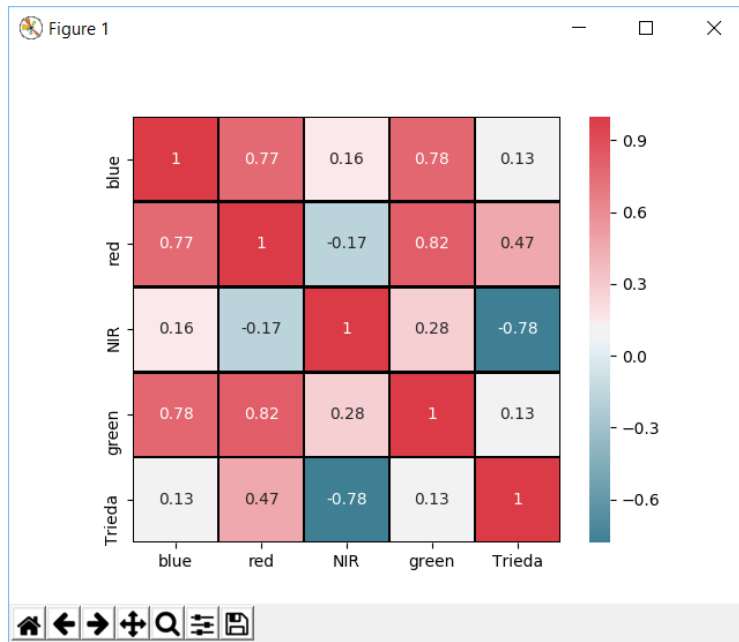
	blue	red	NIR	green	Trieda
0	8243	6780	13404	7672	3
1	8629	6917	14655	7897	4
2	8659	6816	14444	7808	4
3	8497	6524	21732	7655	1
4	8486	6567	20820	7620	1

### 5.2.4.3 Vykreslenie korelačnej tabuľky

Vykreslenie korelačnej tabuľky nie je potrebné pre samotný beh programu, ale slúži na overenie vhodnosti výberu údajov (v našom prípade kanálov). Na základe korelačnej tabuľky sa dá tvrdiť, že korelácia medzi kanálmi B, G, R je relatívne vysoká, ale ich korelácia s kanálom NIR je nízka.

```
#vykreslenie korelačnej tabuľky
corr = data.astype('float64').corr()
ax = SEA.heatmap(corr, cmap=SEA.diverging_palette(220, 10, as_cmap=True),
square=True, annot = True, linecolor = 'k', linewidths = 1)
PLOT.show()
```

Obrázok 13: Test korelácie kanálov



#### 5.2.4.4 Vytvorenie tréningovej a testovacej množiny

V tejto práci bola použitá veľkosť tréningovej množiny polovica celej množiny údajov. Vplyv veľkosti tejto množiny na presnosť klasifikácie je priamo úmerný kvalite reprezentatívnosti tréningovej množiny pixelov. Reprezentatívnosťou tréningovej množiny pixelov je myslené, či tréningová množina dát zachytáva rôzne podmienky, v ktorých rastie konkrétny lesný typ (t.j. rôzny sklon svahu, alebo orientáciu svahu voči svetovým stranám).

```
#vytvorenie tréningovej a testovacej množiny (tréningová je 50% z celkovej množiny pixelov)
fracNum = 0.5
train_set = data.sample(frac = fracNum) # 50% náhodných z data
test_set = data.drop(train_set.index) # zvyšných 50%
print('Veľkosť množiny: tréningovej = {0}; testovacej = {1}'.format(train_set.shape[0], test_set.shape[0]))
```

#### Výstup:

Veľkosť množiny: tréningovej = 2560; kontrolnej = 2559

#### 5.2.4.5 Použitie funkcie *RandomForestClassifier*

Pred použitím funkcie je potrebné najprv klasifikovať t.j. očíslovať druh stromu. (V prípade tejto práce to môže byť metúce nakoľko vyjadrenie druhu stromu je označené diskretným číslom (1, 2, 3, 4), ale mohlo to byť označené textovým výrazom (Buk, Dub, ...). V každom prípade je potrebné aby, si funkcia vytvorila vlastné číslovanie druhov stromov.

```
# Prekódovanie vstupných hodnôt ako vymenovaný typ alebo kategorizovaná premenná indicator, labels = PD.factorize(train_set[classVar[0]]) # classVar[0] je vlastne 'Trieda'
```

Vytvoríme objekt RF a zavoláme funkciu *fit* ktorá sa „naučí“ rozoznávať druhy stromov.

```
# vytvoríme objekt RandomForest
rfco = RandomForestClassifier(n_estimators = 500, oob_score = True)
# priradenie trénovacej množiny
# fit(X, y[, sample_weight])      Build a forest of trees from the training set
(X, y).
rfco.fit(train_set[predictVars], indicator)
```

Funkcia *predict* urobí predpoklad druhov stromov na validačnej množine.

```
#a spolu s validačnou množinou sa urobí predpoklad
treesPred = rfco.predict(test_set[predictVars])
```

Urobí sa spätný preklad druhov stromov, ktorý bol vytvorený funkciou *factorize*.

```
x = []
for kod in treesPred:
    x.append(labels[kod])
treesPredLables = NUM.array(x, dtype=NUM.float64)
```

#### 5.2.4.6 Výsledky procesu RF

Teraz sú pripravené všetky premenné potrebné na zobrazenie výsledkov.

```
# Vytvorenie confusion matrix
PD.crosstab(test_set[classVar[0]], treesPredLables, rownames=['Druhy skutočné'],
colnames=['Druhy predpovedané'])
```

##### Výstup:

Predicted Species	1.0	2.0	3.0	4.0
Actual Species				
1	392	50	0	0
2	15	763	0	3
3	0	1	587	36
4	0	5	21	686

```
# Vytvorenie confusion matrix v %
PD.crosstab(test_set[classVar[0]], treesPredLables, rownames=['Druhy skutočné
[%]'], colnames=['Druhy predpovedané']).apply(lambda r: (r/r.sum())*100, axis=1)
```

##### Výstup:

Druhy predpovedané	1.0	2.0	3.0	4.0
Druhy skutočné [%]				
1	88.687783	11.312217	0.000000	0.000000
2	1.920615	97.695262	0.000000	0.384123
3	0.000000	0.160256	94.070513	5.769231
4	0.000000	0.702247	2.949438	96.348315

```
# Vytvorenie poľa z matice validačných údajov
test_trees = test_set[classVar].as_matrix()
test_trees = test_trees.flatten()
```

```
# Vymenovanie počtu skutočných a predpovedaných druhov stromov
druhStromov = ['nil', 'Dub', 'Buk', 'Smrek', 'Borovica']
for i in range(1,5):
    print('{0} Výskyt = {1} Predikcia = {2}'.format(druhStromov[i],
len(NUM.where(test_trees==i)[0]), len(NUM.where(treesPredLables==i)[0])))
```

##### Výstup:

Dub Výskyt = 442 Predikcia = 407  
 Buk Výskyt = 781 Predikcia = 819

Smrek Výskyt = 624 Predikcia = 608  
Borovica Výskyt = 712 Predikcia = 725

```
# výpočet celkovej presnosti klasifikácie
treesPrd = 0
for i in range(len(treesPredLables)):
    if (test_trees[i] == treesPredLables[i]): treesPrd += 1
error = (NUM.abs(len(treesPredLables) - treesPrd)/len(treesPredLables))*100
print('trénovacia množina = {0} % (počet bodov = {1}) Presnosť = {2} %
'.format(fracNum*100, train_set.shape[0], (100 - error)))
```

**Výstup:**

trénovacia množina = 50.0 % (počet bodov = 2560) Presnosť = 94.88081281750684 %

```
# Zobrazenie zoznamu atribútov a ich dôležitosť na vypočítanie odhadu
list(zip(train_set[predictVars], rfco.feature_importances_))
```

**Výstup:**

[('blue', 0.33766716116292078), ('red', 0.21669033197695942), ('NIR',  
0.36347437074045075), ('green', 0.082168136119668972)]

Funkcia *predict\_proba* zobrazí hodnoty na základe ktorých sa robí konečné rozhodnutie o aký druh stromu sa jedná. Z výstupu je zrejmé, že rozhodnutia sú dosť jednoznačné.

```
# Zobrazí predpokladané pravdepodobnosti prvých 10 bodov (pixelov)
rfco.predict_proba(test_set[predictVars])[0:10]
```

**Výstup:**

```
array([[ 0.01 ,  0.99 ,  0.    ,  0.    ],
       [ 0.    ,  0.    ,  0.138,  0.862],
       [ 0.    ,  1.    ,  0.    ,  0.    ],
       [ 1.    ,  0.    ,  0.    ,  0.    ],
       [ 1.    ,  0.    ,  0.    ,  0.    ],
       [ 0.    ,  0.    ,  0.994,  0.006],
       [ 0.    ,  0.    ,  0.014,  0.986],
       [ 0.002,  0.    ,  0.22 ,  0.778],
       [ 1.    ,  0.    ,  0.    ,  0.    ],
       [ 0.    ,  0.    ,  0.086,  0.914]])
```

### 5.2.4.7 Vloženie výsledku predikcie do geofile databázy

Nasledujúce kroky slúžia na vloženie výsledku predikcie do jednotlivých vrstiev podľa druhu stromu.

*Poznámka: Tu je potrebné dať správnu cestu a názov projektu / geofile databázy.*

```
# zostavenie cesty a názvu projektu
projectName = 'L8_L1_160624'
outputDir = r'D:\gis\{0}\{0}.gdb'.format(projectName)
print(outputDir)
```

**Výstup:**

D:\gis\L8\_L1\_160624\L8\_L1\_160624.gdb

```
# vytvorenie vrstiev z predikcie podľa druhu stromov
for i in range(1,5):
    nameFC = 'Predpoved{0}_{1}'.format(i, druhStromov[i])
```

```
print(nameFC, ' - ', outputDir)
stromy = trainFC[["SHAPE@XY"]][test_set.index[NUM.where(treesPredLables==i)]]
ARCPY.da.NumPyArrayToFeatureClass(stromy, OS.path.join(outputDir, nameFC),
['SHAPE@XY'], spatRef)
```

**Výstup:**

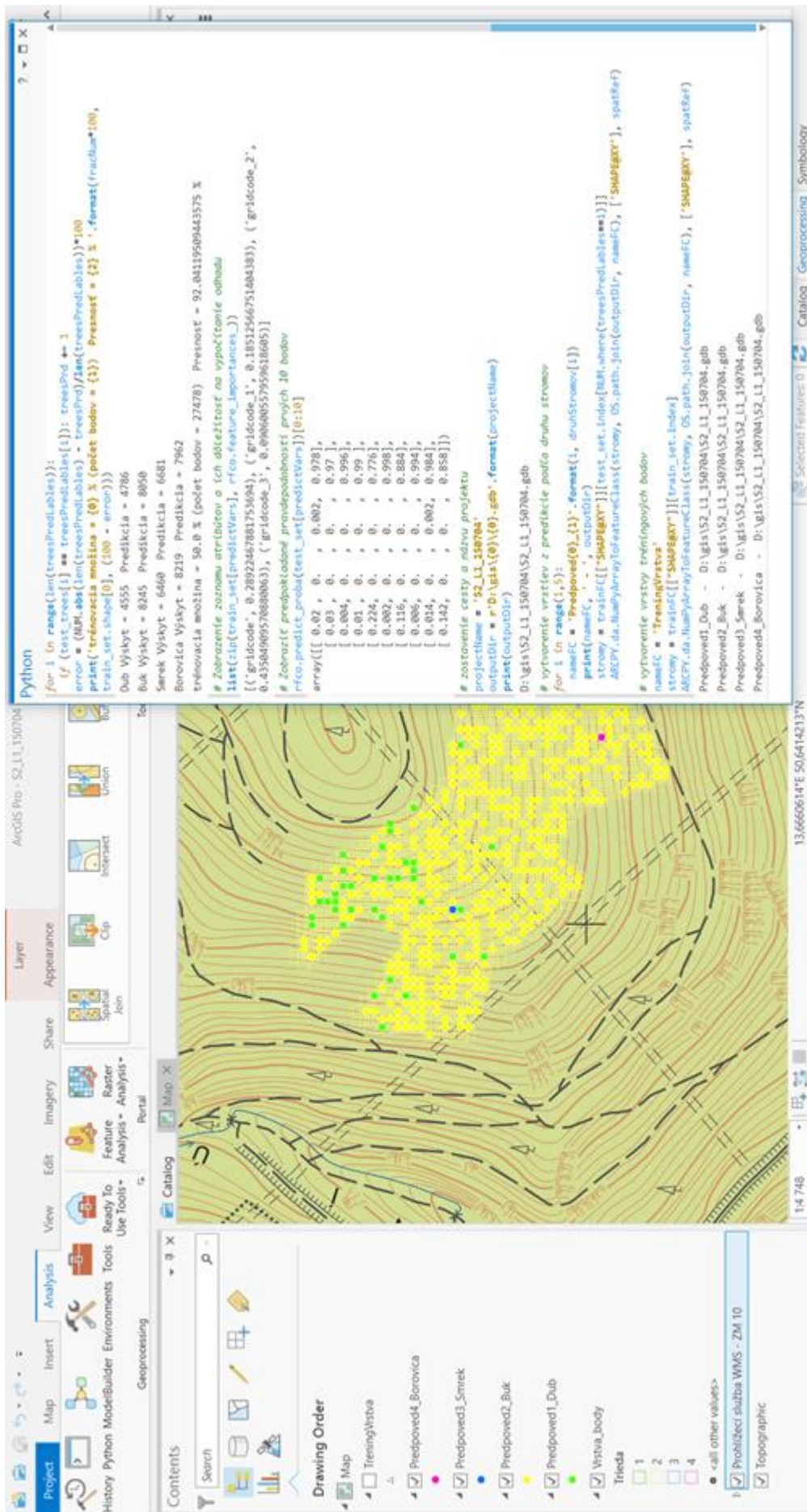
```
Predpoved1_Dub - D:\gis\L8_L1_160624\L8_L1_160624.gdb
Predpoved2_Buk - D:\gis\L8_L1_160624\L8_L1_160624.gdb
Predpoved3_Smrek - D:\gis\L8_L1_160624\L8_L1_160624.gdb
Predpoved4_Borovica - D:\gis\L8_L1_160624\L8_L1_160624.gdb
```

```
# vytvorenie vrstvy tréningových bodov
nameFC = 'TreningVrstva'
stromy = trainFC[["SHAPE@XY"]][train_set.index]
ARCPY.da.NumPyArrayToFeatureClass(stromy, OS.path.join(outputDir, nameFC),
['SHAPE@XY'], spatRef)
```

Po aplikovaní výpočtu predikcie je potrebné pre ďalšie použitie v aplikácii ArcGIS Pro túto vrstvu z geofile databázy zobrazíť, ako vektorovú vrstvu.



Obrázok 14: Použitie Random Forest v aplikácii "ArcGIS Pro"





## 6 VÝSLEDKY

V tejto práci boli použité 3 kategórie snímok. Prvou kategóriou boli snímky z Landsat-8 s úrovňou predspracovania dát Level-1TP, druhou boli snímky tiež z Landsat-8 s úrovňou predspracovania dát Level-2 a treťou kategóriou boli snímky zo Sentinel-2 s úrovňou predspracovania dát Level-1C z troch odlišných dátumov (*Tabuľka 9*). Účelom použitia 3 kategórii snímok bolo vyhodnotiť, pomocou ktorej kategórie snímok je vhodné dosiahnuť najvyššiu klasifikačnú presnosť.

Pre účely vyhodnotenia druhové zloženia lesov bola použitá metóda RF. Postup vyhodnotenia bol opísaný v kapitole č. 5.

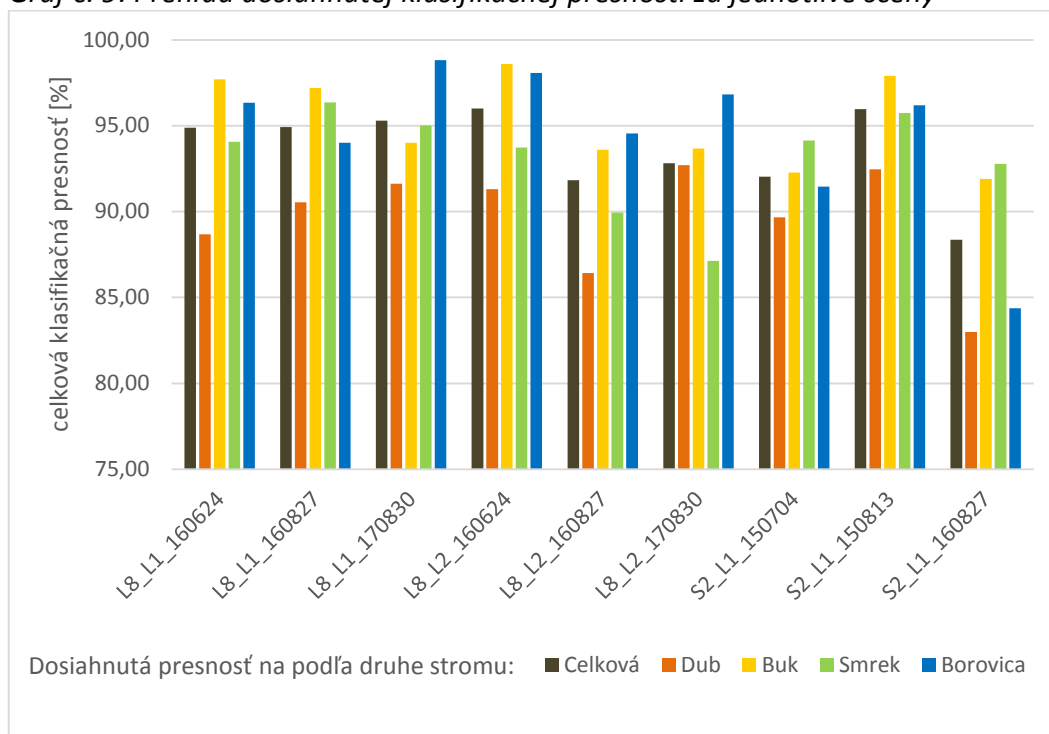
### 6.1 Vyhodnotenie výsledkov

Celkový prehľad dosiahnutých presností klasifikácie za jednotlivé scény je možné získať z nasledujúcej tabuľky a grafu. Pre detailnejší rozbor sa použili údaje z chybových matíc, ktoré sú uvedené v *Prílohe 2*.

*Tabuľka 15: Celková a užívateľská presnosť klasifikácii jednotlivých scén*

Scény	Klasifikačná presnosť [%]				
	Celková	Dub	Buk	Smrek	Borovica
L8_L1_160624	94,88	88,69	97,70	94,07	96,35
L8_L1_160827	94,92	90,54	97,19	96,35	94,01
L8_L1_170830	95,31	91,63	94,00	95,02	<b>98,81</b>
L8_L2_160624	<b>96,01</b>	91,31	98,59	93,74	98,07
L8_L2_160827	91,83	86,42	93,61	89,95	94,56
L8_L2_170830	92,81	92,71	93,67	87,14	96,82
S2_L1_150704	92,04	89,66	92,29	94,15	91,46
S2_L1_150813	95,98	92,47	97,91	95,75	96,19
S2_L1_160827	<b>88,37</b>	<b>83,00</b>	91,91	92,79	84,37

Graf č. 9: Prehľad dosiahnutej klasifikačnej presnosti za jednotlivé scény



**Zoznam nastavených hodnôt parametrov RF pre všetky výstupy v tejto kapitole:**

- **n\_estimators = 500**
- **criterion = 'entropy'**
- oob\_score = „True“
- max\_depth = None
- min\_samples\_split = 2
- min\_samples\_leaf = 1
- min\_weight\_fraction\_leaf = 0.0
- max\_features = 'auto'
- max\_leaf\_nodes = None
- min\_impurity\_decrease = 0.0
- min\_impurity\_split = None
- bootstrap = True
- **oob\_score = True**
- n\_jobs = 1
- random\_state = None
- verbose = 0
- warm\_start = False
- class\_weight = None

*Poznámka: Hrubo vyznačené parametre sú iné, ako prednastavené pre funkciu `sklearn.ensemble.RandomForestClassifier`*

V tejto podkapitole je rozbor získaných výsledkov na základe nasledujúceho členenia:

1. Dosiahnutá klasifikačná presnosť jednotlivých scén
2. Dosiahnutá presnosť klasifikácie na základe veľkosti rastra
3. Dosiahnutá presnosť klasifikácie na základe druhu stromov
4. Vplyv kanálov na vyhodnotenie
5. Chybové odchýlky v identifikácii druhov stromov

## 6.2 Dosiahnutá klasifikačná presnosť jednotlivých scén

Ako zobrazuje *Graf č. 9* najvyššiu celkovú presnosť klasifikácie dosiahla scéna z Landsat-8 s dátami predspracovanými na úrovni Level-2 s dátumom 24. 6. 2016 a druhú najvyššiu klasifikačnú presnosť scéna zo Sentinel-2 s dátumom 15. 8. 2015. Najnižšiu celkovú presnosť dosiahla klasifikácia scény zo Sentinel-2 s dátumom 27. 8. 2016.

Výsledky poukazujú na to, že v prípade Landsat-8 a Sentinel-2 s použitím rôznej úrovne predspracovania dát vyšli výsledky podobne. Tiež vidieť, že v prípade použitia klasifikátora RF pre klasifikáciu snímok nemalo vplyv, z ktorého letného mesiaca boli dáta.

*Tabuľka 16: Vyhodnotenie celkovej klasifikačnej presnosti podľa použitých scén*

Scény	Klasifikačná presnosť
L8_L2_160624	96,01%
S2_L1_150813	95,98%
L8_L1_170830	95,31%
L8_L1_160827	94,92%
L8_L1_160624	94,88%
L8_L2_170830	92,81%
S2_L1_150704	92,04%
L8_L2_160827	91,83%
S2_L1_160827	88,37%

## 6.3 Dosiahnutá presnosť na základe veľkosti pixela

V niektorých prácach spomenutých v rešerši sa autori vyjadrili, že pomocou družíc, ktoré produkujú snímky s vyšším priestorovým rozlíšením je možné dosiahnuť vyššiu klasifikačnú presnosť obrazu. Dokonca *Immitzer M.* tvrdil, že na základe výsledkov v jeho práci, dospel k záveru, že snímky z družice Landsat-8 nie sú vhodné pre klasifikáciu druhového zloženia lesných porastov, pretože priestorové rozlíšenie je príliš veľké.

Výsledky v tejto práci obidva tvrdenia vyvracajú, pretože klasifikácia snímok z obidvoch družíc dosiahla približne podobnú klasifikačnú presnosť.

#### 6.4 Dosiahnutá klasifikačná presnosť na základe druhu stromov

Na základe získaný údajov bola dosiahnutá najvyššia klasifikačná presnosť pri smreku a najnižšia pri dube. Toto bolo pravdepodobne spôsobené, tým že najhomogénnejšie porasty sú smrekové a najmenej homogénne sú dubové porasty. Pri hľadaní vhodných lokalít pre túto prácu bol najväčší problém nájsť homogénne porasty dubu.

Tabuľka 17: Dosiahnutá užívateľská presnosť na základe druhu stromov

Druh stromu	Užívateľská presnosť	Scéna
<b>Borovica</b>	98,814%	L8_L1_170830
<b>Buk</b>	98,588%	L8_L2_160624
<b>Smrek</b>	96,352%	L8_L1_160827
<b>Dub</b>	92,711%	L8_L2_170830

Tabuľka 18: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu „dub“

Scény	Dub
L8_L2_170830	92,711%
S2_L1_150813	92,465%
L8_L1_170830	91,626%
L8_L2_160624	91,314%
L8_L1_160827	90,544%
S2_L1_150704	89,660%
L8_L1_160624	88,688%
L8_L2_160827	86,417%
S2_L1_160827	83,005%

Tabuľka 19: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu „buk“

Scény	Buk
L8_L2_160624	98,588%
S2_L1_150813	97,912%
L8_L1_160624	97,695%
L8_L1_160827	97,193%
L8_L1_170830	94,003%
L8_L2_170830	93,669%
L8_L2_160827	93,606%
S2_L1_150704	92,286%
S2_L1_160827	91,913%

Tabuľka 20: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu „smrek“

Scény	Smrek
L8_L1_160827	96,352%
S2_L1_150813	95,750%
L8_L1_170830	95,025%
S2_L1_150704	94,149%
L8_L1_160624	94,071%
L8_L2_160624	93,740%
S2_L1_160827	92,787%
L8_L2_160827	89,950%
L8_L2_170830	87,138%

Tabuľka 21: Vyhodnotenie dosiahnutej užívateľskej presnosti pri druhu stromu "borovica"

Scény	Borovica
L8_L1_170830	98,814%
L8_L2_160624	98,066%
L8_L2_170830	96,823%
L8_L1_160624	96,348%
S2_L1_150813	96,195%
L8_L2_160827	94,555%
L8_L1_160827	94,013%
S2_L1_150704	91,459%
S2_L1_160827	84,369%

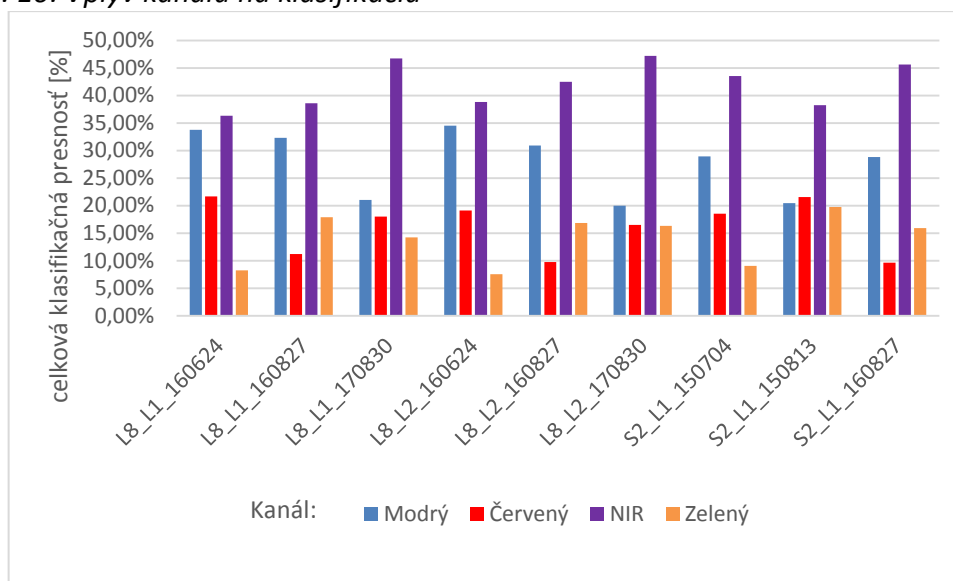
## 6.5 Vplyv kanálov na klasifikáciu

Na základe získaných údajov najväčší vplyv na rozpoznanie druhu drevín mal kanál NIR a B, čo zodpovedá záverom získaných z literárnej rešerše.

Tabuľka 22: Vplyv kanálu na klasifikáciu

Družica / dátum	Kanál			
	Modrý	Červený	NIR	Zelený
L8_L1_160624	33,77%	21,67%	36,35%	8,22%
L8_L1_160827	32,29%	11,19%	38,61%	17,90%
L8_L1_170830	21,03%	18,00%	46,75%	14,21%
L8_L2_160624	34,54%	19,11%	38,81%	7,54%
L8_L2_160827	30,93%	9,77%	42,45%	16,84%
L8_L2_170830	19,98%	16,50%	47,21%	16,31%
S2_L1_150704	28,92%	18,51%	43,50%	9,06%
S2_L1_150813	20,47%	21,54%	38,22%	19,76%
S2_L1_160827	28,85%	9,63%	45,60%	15,92%

Graf č. 10: Vplyv kanálu na klasifikáciu



## 6.6 Chybové odchýlky v identifikácii druhov stromov

Na základe výsledkov klasifikácie, ktoré sú uvedené v *Prílohe 2* v tabuľkách chybových matíc bolo možné zostaviť nasledujúcu tabuľku, ktorá uvádza ktoré druhy drevín boli najviac zamieňané.

Tabuľka 23: Chybové odchýlky v klasifikácii druhov drevín

Zámena druhu stromu	Najnižšia hodnota zámeny	Najvyššia hodnota zámeny
buk za dub	3,94%	14,13%
dub za buk	1,28%	7,22%
borovicu za smrek	3,91%	12,86%
smrek za borovicu	0,66%	14,25%*

\* - táto hodnota vychádza z nameraných údajov, lebo v ostatných výsledkoch zámena je nízka a pohybuje sa do hodnoty 7,08%

## 6.7 Mapové výstupy

V *Prílohe 3* sú mapové výstupy klasifikácie zo snímok L8\_L1\_150624 a S2\_L1\_150704 (označenie v *Tabuľka 9: Základný prehľad informácií o scénach*) za jednotlivé lokality.

V prípade mapových výstupov vytvorených na základe dát z Landsat-8 štvorčky predstavujú jednotlivé pixely celkovej množiny dát (t.j. trénovacie a validačné dáta). Krúžky znamenajú jednotlivé pixely z validačnej množiny dát.

Farba štvorčeka predstavuje, aká drevina rastie na danom území (30 x 30 m, respektíve 10 x 10 m, záleží na družici) podľa referenčných dát.

Farba krúžku znamená do akej klasifikačnej triedy bol klasifikovaný pixel ležiaci pod krúžkom.

Trénovacie množina pixelov je na mapových výstupoch zobrazená, ako štvorčeky bez krúžkov.

Mapové výstupy, ktoré zobrazujú dáta z klasifikácie zo snímok zo Sentinel-2 sa od tých z Landsat-8 trochu líšia. Na mapové výstupy neboli pridané štvorčeky (celková množina pixelov), pretože by na mapových výstupoch splývali štvorčeky s krúžkami, keďže snímky zo Sentinel-2 majú vyššie priestorové rozlíšenie, ako snímky z Landsat-8.

## 7 DISKUSIA

Kapitola Diskusia bola rozdelená do troch častí. Prvá časť bola venovaná výskumným otázkam a odpovediam na tieto otázky. Druhá časť bola venovaná porovnaniu dosiahnutých výsledkov v tejto práci s výsledkami z iných prác. Tretia časť bola venovaná možnostiam dosiahnutia lepších výsledkov.

### 7.1 Odpovede na výskumné otázky

1. Dáta z ktorej multispektrálnej družice, vzhľadom na vlastností pasiem, je vhodnejšie použiť pri skúmaní druhového zloženia lesov na území ČR?

- Na základe literárnej rešerše je pre klasifikáciu druhového zloženia lesa vhodnejšie použiť snímky s vysokým priestorovým rozlíšením, ako snímky z družíc so stredným priestorovým rozlíšením.
- V práci boli použité snímky z družíc Landsat-8 a Sentinel-2, pretože snímky z iných družíc neboli voľne dostupné. Na základe výsledkov v práci je možné tvrdiť, že aj napriek tomu, že obidve družice majú rôzne priestorové rozlíšenie, tak dáta z oboch družíc sú približne rovnako vhodné pre skúmanie druhového zloženia lesov. Takže je možné tvrdiť, že v prípade použitia klasifikátora RF nemá priestorové rozlíšenie snímok družice vplyv na presnosť klasifikácie.

2. Použitie, ktorého klasifikátora a nastavením akých hodnôt jeho parametrov je najvhodnejšie pre klasifikáciu drevín na území ČR?

- Na základe rešerše je najvhodnejšie použiť metódu Random Forest s použitím parametrov  $n\_estimators = 500$ ,  $criterion = 'Gini'$ ,  $mtry = 2$ .
- Na základe výsledkov v práci je možné tvrdiť, že metóda Random Forest je najvhodnejšia pre skúmanie druhového zloženia lesov s použitím parametrov  $n\_estimators = 500$ ,  $criterion = 'entropy'$ ,  $oob\_score=True$ .

3. Je možné na základe súčasných poznatkov dosiahnuť celkovú klasifikačnú presnosť druhového zloženia lesov nad 80 %?

- Na základe literárnej rešerše, je možné tvrdiť, že celková presnosť v prípade pixelovej klasifikácie dosiahla až 86 % a v prípade objektovej klasifikácie ešte vyššiu presnosť. Z rešerši je vidieť, že každý z klasifikátorov (RF, SVM, MLC, LDA), okrem PDA, dosiahol aspoň v jednej štúdiu vyššiu celkovú presnosť, ako 80 %. V prípade klasifikátora RF iba v jednej štúdiu klasifikácia snímok dosiahla celkovú presnosť pod 80 % dôvod bol, že územie bolo príliš druho



heterogénne vzhľadom na priestorové rozlíšenie družíc, ktoré boli pre prácu použité.

- Celková presnosť klasifikácie všetkých snímok v tejto práci bola vyššia ako 88 %. Dôvodom bol výber vhodného klasifikátora pre špecifická zadania práce a vhodné zvolenie hodnôt do parametrov klasifikátora.
- Takže odpoveď na výskumnú otázku je, že je možné dosiahnuť vyššiu klasifikačnú presnosť, ako 80 %.

4. Pri klasifikácii sa ktoré druhy drevín (dub, buk, smrek, borovica) najviac zamieňajú?

- V literárnej rešerši sa v prípade práce *Immitzera M. a kol.* [4] sa najčastejšie zamieňal borovica za smrek a v druhom najčastejšom prípade smrek za borovicu. V prípade druhej štúdie *Immitzera M. a kol.* [9] sa najčastejšie zamieňal buk za dub a v druhom najčastejšom prípade sa zamieňal smrek za borovicu.
- Na základe výsledkov tejto práce je možné tvrdiť, že najčastejšie sa zamieňa smrek za borovicu. V poradí druhá najčastejšia zámena je buk za dub.
- Listnaté a ihličnaté druhy drevín sa navzájom takmer vôbec nezamieňajú ani v rešerši ani v tejto práci.
- Keďže lesné porasty rôznych druhov drevín majú rôznu rozlohu v rôznych prácach, tak zamieňanie druhov drevín bolo porovnávané v relatívnych hodnotách.

## 7.2 Porovnanie dosiahnutých výsledkov s výsledkami v iných prácach

V tejto práci bola dosiahnutá celková presnosť klasifikácie jednotlivých scén od 88,37 % (v prípade Sentinel-2) do 96,01 % (v prípade Landsat-8).

Pre porovnanie *Immitzer M.* [4] vo svojej práci v prípade klasifikácie 4 rovnakých druhov drevín a s použitím rovnakých kanálov dosiahol celkovú klasifikačnú presnosť len 86,8 %. V tej istej práci autor dosiahol najvyššiu celkovú presnosť klasifikácie 95,9 %, ale v tom prípade použil objektovú klasifikáciu a 8 na miesto 4 kanálov, čo je v hodné v prípade klasifikácie snímok z WorldView-2.

V inej práci *Immitzer M.* [9] použitím metódy RF dosiahol v prípade klasifikácie snímky z WorldView-2 celkovú presnosť 74,4 %, v prípade Sentinel-2 68,1 % a v prípade Landsat-8 len 49,4 %. V závere práce dodáva, že snímky z družíc landsat-8 a Sentinel-2 nie sú kvôli svojmu priestorovému rozlíšeniu vhodné pre klasifikáciu druhového zloženia lesov. Toto jeho tvrdenie však vyvracajú výsledky dosiahnuté v tejto práci.

V inej štúdií *Sothe C. a kol.* [12] použitím metódy RF dosiahli v prípade klasifikácie pomocou snímok z družíc Landsat-8 a Sentinel-2 celkovú presnosť klasifikácii jednotlivých scén nad 90 %. V tejto štúdií sa neklasifikovali jednotlivé druhy drevín, ale 3 skupiny drevín, ktorá každá z nich obsahovala iné zloženie druhov drevín.

Vo všetkých troch prácach išlo o použité rovnakých hodnôt v parametroch klasifikátora RF t.j. Giniho kritérium (*criterion*), 500 rozhodovacích stromov (*nmtree*), 2 rozdelenia v každom uzle stromu (*mtry*). Ostatné hodnoty parametrov boli ponechané v prácach v rešerši, ako default. V prácach v rešerši autori vyskúšali nastavenie rôznych parametrov, ale vo všetkých troch prácach sa autori zhodli v optimalizácii ladenia hodnôt parametrov, ktoré použili vo výslednej klasifikácii.

Na základe literárnej rešerše je vidieť že metóda RF sa začala pre klasifikáciu druhového zloženia lesov používať len pred pár rokmi. Z literatúry je vidieť, že autori použili jazyk „R“. Nebola nájdená žiadna práca, v ktorej by išlo o klasifikáciu družicových snímok so zameraním na druhové zloženie lesov v jazyku „Python“.

### 7.3 Možnosti dosiahnutia vyššej klasifikačnej presnosti

Možnosti dosiahnutia vyššej klasifikačnej presnosti je možné v oblasti zlepšenia pracovného postupu vo viacerých krokoch tejto práce. Zlepšenie sa dá dosiahnuť napr. v ladení parametrov, alebo v starostlivejšom výbere tréningových a validačných množín pixelov.

Čo sa týka ladenia hodnôt vstupných parametrov, treba brať do úvahy, že metóda RF sa neustále vyvíja a podľa údajov z podkapitoly 5.2.3 nie je možné v súčasnosti dosiahnuť lepšie výsledky, ako boli dosiahnuté v tejto práci o viac ako 0,5 %, tým že by sa ladili hodnoty parametrov. Ako bolo overené v tejto práci výsledný efekt ladenia parametrov nezodpovedá jednoduchému súčtu zlepšení výslednej presnosti pri zmene hodnôt jednotlivých parametrov.

V niektorých prácach spomenutých v rešerši sa odporúča, aby pri použití metódy RF sa ladili hodnoty parametrov a hľadalo sa najvyššie dosiahnutie klasifikačnej presnosti. V tejto práci boli použité rovnaké parametre pre všetky snímky, aby bolo možné jednoducho urobiť porovnanie výsledkov.

Najväčší priestor pre dosiahnutie výsledkov s vyššou klasifikačnou presnosťou je v zlepšení algoritmu pre výber tréningovej, ale aj validačnej množiny pixelov. V tejto práci bolo dokázané, že údaje z mapového portálu LhpoMap nie vždy presne ukazovali skutočnú hranicu lesa. Tiež detekcia „dier“ v súvislom lesnom korunovom závoji nie je úplne vhodná s pomocou služby google maps, pretože aplikácia používa aj družicové snímky zo zimného ročného obdobia a v prípade listnatých lesov sa ťažko detekujú diery v súvislom lesnom korunovom závoji počas zimného obdobia, keď stromy nemajú olistené koruny.

Pre dosiahnutie čo najlepších výsledkov tréningová množina by mala čo najviac reprezentovať všetky podmienky rastu stromov t.j. klimatické, terénne, zloženie pôdy, vlhkosť pôdy a aj sklon a smer svahov.

Pri vysokej dosahovanej presnosti tiež vzniká otázka, že ak v tréningovej množine je identifikovaný iný druh stromu, či je to chyba metódy, alebo v skutočnosti je tam iný druh stromu. Toto by mohlo byť skúmané (okrem zistenia na mieste terénnym

šetrením) porovnávaním výsledkov klasifikácie z iných snímok, alebo pri tej istej snímke opakované vyhodnotenie (požitím funkcie RF), ktoré by bolo založené na inej množine tréningových údajov.

## 8 ZÁVER

Práca bola zameraná na skúmanie druhového zloženia lesných porastov z optických družíc. Išlo o skúmanie dubových, bukových, smrekových a borovicových lesných porastov v 24 lokalitách na území 4 krajov Českej republiky.

Bolo klasifikovaných 9 snímok, ktoré boli urobené v letných mesiacoch jún, júl a august. Išlo o 3 snímky družice z Landsat-8 s predspracovaním dát na úrovni Level-1TP, ďalej 3 snímok z Landsat-8 s dátami na úrovni predspracovania Level-2 a 3 snímky zo Sentinel-2 s predspracovaním dát na úrovni Level-1C. V prípade dát z obidvoch družíc boli použité kanály B, G, R a NIR. Ako referenčné dáta boli použité údaje o druhovej skladbe lesa za základné porastné jednoty z mapového portálu LhpoMap, ktorý prevádzkuje UHUL.

Ako klasifikátor bola použitá metóda Random Forest. Úprava a klasifikácia dát prebehla v softvéroch ArcMap a ArcGIS Pro od spoločnosti ESRI. V prípade klasifikačného procesu bol použitý programovací jazyk Python.

V tejto práci bola dosiahnutá celková presnosť klasifikácie jednotlivých scén od 88,37 % (v prípade Sentinel-2) do 96,01 % (v prípade Landsat-8). Všetky triedy druhov drevín dosiahli pri všetkých scénach vysokú užívateľskú aj procesnú klasifikačnú presnosť t.j. vyššiu, ako 80 %.

V literatúre dosiahla celková presnosť pixelovej klasifikácie vybraných druhov drevín na základe použitia snímok z rovnakých družíc, ako v tejto práci od 49,4 % do 86,8 %. Z týchto výsledkov je možné tvrdiť, že klasifikácia v tejto práci dosiahla vyššiu celkovú presnosť, ako klasifikácie v literatúre. Dôvodmi lepších výsledkov je jednak vhodnejšie nastavenie hodnôt parametrov, ale tiež aj novšia verzia klasifikátora Random Forest použitá v tejto práci.

Ciel práce bol splnený, klasifikácia lesných porastov z dát optických družíc dosiahla u všetkých snímok celkovú presnosť vyššiu ako 88 %.

Vývoj metód RF a nových GIS nástrojov uľahčuje použitie tejto metódy pre praktické používanie, čo dokazuje aj táto práca a ukazuje pomerne jednoduchú možnosť integrácie najnovších knižníc „učiacich sa strojov“ a novo pripravovaného GIS nástroja (ArcGis Pro od spoločnosti ESRI).

## Zoznam použitých dát:

[a] Portál LandsatLook Viewer (U.S. Geological Survey)

<http://landsatlook.usgs.gov/>

[b] Portál Sentinel-2 Pre-Operations Data Hub (ESA)

<https://scihub.copernicus.eu/s2/#/home>

[c] Portál LhpoMap (UHUL)

<http://geoportal.uhul.cz/LHPOMap/>

[d] Google maps (Google)

<https://www.google.com/maps>

## 9 BIBLIOGRAFIA

- [1] Feranec J. a kol., Meniace sa Slovensko očami satelitov, Bratislava: VEDA, Vydavateľstvo SAV, 2012.
- [2] Bucha T. a kol., Satelity v službách lesa, Zvolen: Národné lesnícke centrum, 2014.
- [3] Ministerstvo zemného hospodárstva, „Zpráva o stavu lesa a lesního hospodářství 2016, Praha,“ 6 12 2017. [Online]. Available: [http://eagri.cz/public/web/file/567452/Zprava\\_o\\_stavu\\_lesa\\_2016.pdf](http://eagri.cz/public/web/file/567452/Zprava_o_stavu_lesa_2016.pdf). [Cit. 22 12 2018].
- [4] Immitzer M. a kol., „Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data,“ *Remote Sensing*, zv. 4, %1. vyd.10.3390/rs4092661, pp. 2661-2693, 2012.
- [5] Verlič A. a kol., „Tree species classification using WorldView-2 satellite images and laser scanning data in a natural urban forest,“ *Šumarski list*, %1. vyd.9-10, pp. 477-488, 2014.
- [6] Lin C. a kol. , „Classification of Tree Species in Overstorey Canopy of Subtropical Forest Using QuickBird Images,“ *Plos One*, %1. vyd.10, 2015.
- [7] Carleer A. a Wolff E., „Exploitation of Very High Resolution Satellite Data for Tree Species Identification,“ *Photogrammetric Engineering and Remote Sensing*, %1. vyd.1, pp. 135 - 140, 2004.
- [8] Krahwinkler P. a Rossmann J., „Tree Species Classification Based on the Analysis of Hyperspectral Remote Sensing Data,“ *EARSeL, Rainer Reuter (Editor)*, %1. vyd.1, pp. 321 - 327, 2010.

- [9] Immitzer M. a kol., „[www.researchgate.net](http://www.researchgate.net),“ Dreiländertagung der DGPF, der OVG und der SGPF in Bern, Schweiz, 10 Jun 2016. [Online]. Available: [https://www.researchgate.net/publication/303882915\\_Verwendung\\_von\\_multispektralen\\_Sentinel-2\\_Daten\\_fur\\_die\\_Baumartenklassifikation\\_und\\_Vergleich\\_mit\\_anderen\\_Satellitensensoren](https://www.researchgate.net/publication/303882915_Verwendung_von_multispektralen_Sentinel-2_Daten_fur_die_Baumartenklassifikation_und_Vergleich_mit_anderen_Satellitensensoren). [Cit. 5 2 2018].
- [10] Immitzer M. a kol., „Suitability of Sentinel-2 Data for Tree Species Classification in Central Europe,“ rev. *LivingPlanet16*, Praha, 2016.
- [11] Bucha T., „Klasifikácia drevinového zloženia lesov Slovenska zo satelitných snímok ako súčasť monitorovania biodiverzity lesných ekosystémov,“ rev. *Lesnícky výskumný ústav*, Zvolen, 2007.
- [12] Sothe C. a kol., „Evaluating Sentinel-2 and Landsat-8 Data to Map Sucessional Forest Stages in a Subtropical Forest in Southern Brazil,“ *Remote Sensing*, %1. vyd.9, pp. 1 - 22, 2017.
- [13] Le Louarn M. a kol. , „"Kill Two Birds with One Stone ": Urban Tree Species Classification Using Bi-Temporal Pléiades Images to Study Nesting Preferences of an Invasive Bird,“ 2017. [Online]. Available: <http://www.mdpi.com/2072-4292/9/9/916>. [Cit. 8 1 2018].
- [14] Škvaraninová J., „Priebeh vegetatívnych fenologických fáz autochtónnych populácií smreka obyčajného (*Picea abies* L.) zo Slovenska,“ *Lesnícky časopis – Forestry Journal*, zv. 55, %1. vyd.1, p. 13–27, 2009.
- [15] Snopková Z. , „Sieť fenologických staníc: Monitorovacia sieť, merané veličiny a metódy monitorov,“ SHMÚ, 2017. [Online]. Available: <http://www.shmu.sk/sk/?page=311>. [Cit. 30 3 2018].
- [16] Larcher W. , *Physiological Plant Ecology*, Springer, 2003.
- [17] The R Foundation, „What is R?,“ 15 3 2018. [Online]. Available: <https://www.r-project.org/about.html>. [Cit. 29 3 2018].
- [18] A. Liaw, „randomForest v4.6-12 - Classification And Regression With Random Forest,“ 6 10 2015. [Online]. Available: <https://www.rdocumentation.org/packages/randomForest/versions/4.6-12/topics/randomForest>. [Cit. 20 3 2018].
- [19] A. S. Walia, „Random Forests in R,“ 24 7 2017. [Online]. Available: <https://www.r-bloggers.com/random-forests-in-r/>.
- [20] Python Software Foundation, „Python Software Foundation - Download the latest version for Windows,“ 5 2 2018. [Online]. Available: <https://www.python.org/downloads/>. [Cit. 30 3 2018].

- [21] Funding provided by INRIA and others, „scikit-learn Machine Learning in Python,“ 2007 - 2017. [Online]. Available: <http://scikit-learn.org/stable/>. [Cit. 23 3 2018].
- [22] KDnuggets, „R GUIs you use frequently,“ KDnuggets, April 2011. [Online]. Available: <https://www.kdnuggets.com/polls/2011/r-gui-used.html>. [Cit. 14 4 2018].
- [23] Geographic Information Systems, „memoryerror in Supervised Random Forest Classification in Python sklearn,“ 7 júl 2017. [Online]. Available: <https://gis.stackexchange.com/questions/247901/memoryerror-in-supervised-random-forest-classification-in-python-sklearn>. [Cit. 20 3 2018].
- [24] Learn ArcGIS (© 2018 Esri), „Predict Seagrass Habitats with Machine Learning,“ 2018. [Online]. Available: <https://learn.arcgis.com/en/projects/predict-seagrass-habitats-with-machine-learning/lessons/create-a-training-dataset.htm>. [Cit. 20 3 2018].
- [25] C. Albon, „Random Forest Classifier Example,“ 20 december 2017. [Online]. Available: [https://chrisalbon.com/machine\\_learning/trees\\_and\\_forests/random\\_forest\\_classifier\\_example/](https://chrisalbon.com/machine_learning/trees_and_forests/random_forest_classifier_example/). [Cit. 20 3 2018].
- [26] scikit-learn, „scikit-learn user guide, Release 0.19.1,“ 21 november 2017. [Online]. Available: [http://scikit-learn.org/stable/\\_downloads/scikit-learn-docs.pdf](http://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf). [Cit. 20 3 2018].
- [27] Ho T. K., „Random Decision Forests,“ 1995. [Online]. Available: <http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf>. [Cit. 28 11 2017].
- [28] Ho T. K., „The Random Subspace Method for Constructing Decision Forests,“ 1998. [Online]. Available: <http://ect.bell-labs.com/who/tkh/publications/papers/df.pdf>. [Cit. 28 11 2011].
- [29] Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning*, Stanford, California: Springer, 2008, pp. 352, 587 - 588, 592.
- [30] Zararsız G. a kol. , „A comprehensive simulation study on classification of RNA-Seq data,“ 23 August 2017. [Online]. Available: <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0182507&type=printable>. [Cit. 20 12 2017].
- [31] Kleinberg E. M., „AN OVERTRAINING-RESISTANT STOCHASTIC MODELING METHOD FOR PATTERN RECOGNITION,“ *The Annals of Statistics*, zv. 24, %1. vyd.6, pp. 2319 - 2349, State University of New York, Buffalo 1996.
- [32] Breiman L., „RANDOM FORESTS,“ 2001. [Online]. Available: <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>. [Cit. 24 11 2017].
- [33] Liaw A. , „Breiman and Cutler's Random Forests for Classification and Regression,“ 7 October 2015. [Online]. Available: <https://cran.r->

- project.org/web/packages/randomForest/randomForest.pdf. [Cit. 28 11 2017].
- [34] A. D. Resnick, „Random Forests“. USA Patent 3185828, 19 12 2006.
- [35] Breiman L. a Cutler A. , „Random Forests Leo Breiman and Adele Cutler,“ 2004. [Online]. Available: [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm). [Cit. 22 11 2017].
- [36] Amit Y. a Geman D., „Shape Quantization and Recognition with Randomized Trees,“ August 1996. [Online]. Available: [http://www.wisdom.weizmann.ac.il/~vision/courses/2003\\_2/shape.pdf](http://www.wisdom.weizmann.ac.il/~vision/courses/2003_2/shape.pdf). [Cit. 24 11 2017].
- [37] Kleinberg E., „AN OVERTRAINING-RESISTANT STOCHASTIC MODELING METHOD FOR PATTERN RECOGNITION,“ *The Annals of Statistics*, zv. 24, %1. vyd.6, pp. 2319 - 2349, 1996.
- [38] Diettrich T. G., „An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization,“ *Machine Learning*, zv. 40, %1. vyd.2, pp. 139 - 157, 2000.
- [39] S. Polamuri, „How the random forest algorithm works in machine learning,“ 22 máj 2017. [Online]. Available: <https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning/>. [Cit. 13 04 2018].
- [40] M. Gehlot, „What is the difference between bagging and bootstrapping in ML/statistics?,“ 16 nov 2015. [Online]. Available: <https://www.quora.com/What-is-the-difference-between-bagging-and-bootstrapping-in-ML-statistics>. [Cit. 14 4 2018].
- [41] Geurts P., Witten D., Hastie T., Tibshirani R., An Introduction, University of Southern California, University of Washington, Stanford University: Springer, 2013, pp. 316 - 319.
- [42] Svobodová H., Věžník A., Hofmann E., KGeog, Pedagogická fakulta, Masarykova univerzita, „Základní geografické charakteristiky ČR,“ 2013. [Online]. Available: <https://is.muni.cz/do/rect/el/estud/pedf/js13/geograf/web/pages/01-geograficke-charakteristiky.html>. [Cit. 21 3 2018].
- [43] Ministerstvo Zemědělství ČR, „Lesnictví,“ 2018. [Online]. Available: <http://eagri.cz/public/web/mze/lesy/lesnictvi/>. [Cit. 22 3 2018].
- [44] Chytrý M., „Vegetace České republiky, Ústav botaniky a zoologie PŘF MU,“ 18 2 2018. [Online]. Available: <http://www.sci.muni.cz/botany/chytry/veg-cr/Veg-CR02-Lesy.pdf>. [Cit. 22 3 2018].
- [45] Ministerstvo zemědělství, „Zpráva o stavu lesa a lesního hospodářství České republiky v roce 2008, Praha,“ 2009. [Online]. Available: <http://web.archive.org/web/20120306092907/http://www.uhul.cz/zelenazprava/2008/>



- zz2008.pdf. [Cit. 22 3 2018].
- [46] Divíšek J. a kol. , „Vegetační stupně střední Evropy,“ Geografický ústav, Přírodovědecká fakulta, Masarykova univerzita, 2010. [Online]. Available: [https://is.muni.cz/el/1431/jaro2010/Z0005/18118868/index\\_VS.html](https://is.muni.cz/el/1431/jaro2010/Z0005/18118868/index_VS.html). [Cit. 30 3 2018].
- [47] UHUL, „Druhová skladba,“ 5 1 2017. [Online]. Available: <http://www.uhul.cz/mapy-a-data/469-druhova-skladba>. [Cit. 14 11 2017].
- [48] USGS, „Landsat 8,“ U.S. Department of the Interior | U.S. Geological Survey, 2018. [Online]. Available: <https://landsat.usgs.gov/landsat-8>. [Cit. 31 3 2018].
- [49] Barsi J. A. a kol., „The Spectral Response of the Landsat-8 Operational Land Imager,“ Remote Sensing, Volume 6, Issue 10, 2014. [Online]. Available: <http://www.mdpi.com/2072-4292/6/10/10232/htm>. [Cit. 31 3 2018].
- [50] Gatti A. a Naud C., „Sentinel-2 Products Specification Document,“ ESA, 27 9 2017. [Online]. Available: <https://sentinel.esa.int/documents/247904/685211/Sentinel-2+Products+Specification+Document+%28PSD%29/0f7bedeb-9fbb-4b60-91aa-809162de456c>. [Cit. 31 3 2018].
- [51] ESA, „Level-1,“ 2000 - 2018. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/processing-levels/level-1>. [Cit. 5 3 2018].
- [52] Stack Overflow, „RandomForestClassifier.fit uses different amount of RAM on different machines,“ Founded in 2008, Stack Overflow, 7 11 2016. [Online]. Available: <https://stackoverflow.com/questions/40293169/randomforestclassifier-fit-uses-different-amount-of-ram-on-different-machines>. [Cit. 20 3 2018].
- [53] T. Plapinger, „Tuning a Random Forest Classifier,“ 12 8 2017. [Online]. Available: <https://medium.com/@taplapinger/tuning-a-random-forest-classifier-1b252d1dde92>. [Cit. 25 3 2018].
- [54] Copyright © 2016-2018 EliteDataScience.com, „Overfitting in Machine Learning: What It Is and How to Prevent It,“ 7 9 2017. [Online]. Available: <https://elitedatascience.com/overfitting-in-machine-learning>. [Cit. 28 3 2018].

# PRÍLOHA 1: REŠERŠ A GEOREFERENCOVANIE MAPOVÝCH PODKLADOV

## 1.1. Tabuľkový prehľad rešerše 1

Číslo	Autor	Miesto	Počet druhov drevín	Dreviny	Zdroj dát	Podobné pásma s L-8 a S-2	Použité klasifikátory	Dátum snímky	Ročné obdobie získania snímky	Predspracovanie vykonané poskytovateľom dát	Predspracovanie vykonané autormi	Validačné dáta
1.	Immitzer	Wien	10	dub, buk, borovica, smrek	WorldView 2	B, G, R, NIR	RF a LDA	10.7.2010	LETO	Neuvedené	Prevod DN -> Žiara, Atmosferická korekcia (FLASH (ENVI)), (TOA Odrazivost), Georeferencovanie pomocou DMT	1465 referenčných polygónov
2.	Verlič	Lužana	5	dub, buk, borovica, smrek	WorldView 2	B, G, R, NIR	SVM	24.3.2011(Air) a 1.8.2010(WV2)	JAR, LETO	Level 1B (základné radiometrické a senzorové korekcie)	Ortorektifikácia (RPC) na 1 m pomocou DMT	Podrobná virtuálna sieť (Grid of plots 100 x 100 m)
3.	Chinsu Lin	Taiwan a USA	40	x	Quick Bird	B, G, R, NIR	SVM, MLC, MD, SAM, SID, NN	24.8.2002 (QB)	LETO	Dáta s geometrickou korekciou.	Atmosferická korekcia, Dátová fúzia.	Dáta pozemného prieskumu
4.	E. Wolff	Brusel	7	dub, buk, borovica	Ikonos 2	B, G	MLC	8.6.2000, 7.10.2000	LETO, JESEŇ	Georeferencované na 50 m	Snímky boli ortorektifikované pomocou DMT a 1256 geodetickými kontrolnými bodmi. Mean Filter, Modal Filter	Sonlan Forest parcel maps (trénovacie aj validačné dáta)
5.	Krahwinkel a Rossmann	Arnsberg, Nemecko	6	dub, buk, smrek	SPOT	x	PDA (Rozhodovací strom)	Neuvedené	Neuvedené	Neuvedené	Bolo odstránených približne polovica pixlov, lebo ležali v tieňovej časti koruny.	Ako pozemné pravdivé dáta bolo vybratých 7 forest stands
6.	Immitzer a kol.	Mnichov, Nemecko	12	dub, buk, borovica, smrek	WorldView-2, Sentinel-2, Landsat-8	OK	RF	12.7.2011(WV2)16.8.2013(WV2) 13.8.2015 (S2), 9.8.2015	LETO	L8SR (L8)	Georeferencia a atmosferická korekcia (ACTOR (CATENA)) pre WV2, Sen2Cor pre S2	Neuvedené
7.	Bucha	Slovensko	15	dub, buk, borovica, smrek	Landsat-4-5 (TM)	B, G, R	MLC	23.7.1990, 31.7.1990, 30.7.1992, 6.8.1992, 8.7.1993, 31.7.1993	LETO	Neuvedené	Do prespracovania vstupuje model zatienenia reliéfu Slovenska nad DMT, Georeferencovanie, Topologická normalizácia	Lesnícke porastové mapy
8.	Sathe a kol.	Brazília	12	x	Sentinel-2 a Landsat-8	OK	SVM a RF	12.6.2016 (L8), 3.11.2016 (L8), 8.6.2016 (S2), 6.12.2016 (S2)	LETO, JESEŇ, ZIMA	Level L1T (L8) Level 1C (S2) (geokódované a s terénou korekciou)	Atmosferická korekcia DN -> BOA Reflectance (FLASH) (L8), Pansharped na 10 m pre S2 a 15 m pre L8.	Lesnícke mapy a terény prieskum
9.	Le Louarn	Marseille	4+2	x	Pleiades	x	SVM a RF	28.7.2015, 24.3.2016	JAR, LETO	Neuvedené	Atmosferická korekcia (Optical Calibration module, Orfeo ToolBox), DN -> TOA Reflectance,	Terény prieskum v lete 2016, 2017
10.	Kärgel a Janzen	Barnim a Uckermark, Nemecko	4+1 / 3	dub, buk, borovica, smrek	RapidEye	B, G	MLC	02.05.2009, 24.09.2010	JAR, JESEŇ	Level 3A	Mozaikovanie Neuvedené	Lesnícke mapy

Vysvetlivky:  
zelené pole - práca spĺňa dané kritérium  
červené pole - práca nespĺňa požadované kritérium  
OK - totožné vlastnosti pásiem s L-8 a S-2  
x - znamená to, že buď v práci nie je klasifikovaná žiadna zo 4 vybraných drevín, alebo, že v danom článku bola použitá družica, ktorej jednotlivé kanály majú iné vlastnosti ako družice L-8 a S-2

## 1.2. Tabuľkový prehľad rešerše 2

Číslo	Autor	Miesto	Naj. Klasifikátor	Pixel./ Objekt. klasifikácia	Parametre s naj. výsledkami	Pásma s naj. výsledkami	Naj. Výsledky	Chyba nahodnoteni a (Proces. A.)	Chyba podhodnotenia (User A.)
1.	Immitzer	Wien	RF	Objektová klasifikácia	Gini criterion, 500 trees, 2 variables at each node	C, B, G, Y, R, RE, NIR1, NIR2	OA 4c, 8b = 95,9%	dub 97,4%, buk 98,4%, borovica 93,4% smrek 94,9%	dub 97,4%, buk 98%, borovica 93,7% smrek 94,6%
2.	Verlič	Lubľana	SVM	Objektová klasifikácia	Pokus - omyl <b>Neuvedené</b>	C, B, G, Y, R, RE, NIR1, NIR2	OA 5c, 8b = 58%	dub 70%, buk 38%, borovica 50%, smrek 80%	dub 47%, buk 64%, borovica 57%, smrek 69%
3.	Chinsu Lin	Taiwan a USA	MLC	Pixelová klasifikácia	<b>Neuvedené</b>	B, G, R, NIR	OA 5c, 13b = 100%	x	x
4.	E. Wolff	Brusel	MLC	Pixelová klasifikácia	11 tried <b>Neuvedené</b>	B, G, R, NIR, NDVI, PCA1	OA 11c, 6b = 85,8%	dub 92,7%, ml. buk 97%, st. buk 44,2%, Scots pine 56,4%, Corsican pine 99,5 %	dub 87%, ml. buk 97%, st. buk 100%, Scots pine 63,9%, Corsican pine 43,5%
5.	Krahwinkler a Rossman	Arnsberg, Nemecko	PDA	Objektová klasifikácia	<b>Neuvedené</b>	IR-R, IR-G, IR-B	OA 5c, = 78%	dub 72,6%, buk 73,2%, smrek 71,9%	dub 74%, buk 79%, smrek 79,1%
6.	Immitzer a kol.	Mníchov, Nemecko	RF	Pixelová klasifikácia	Segmentácia a Mean shift, 500 trees,	C, B, G, Y, R, RE, NIR1, NIR2	A 12c, 74,4	dub 55,9%, buk 78,5%, borovica 84%, smrek 94,5, (pre WV2)	dub 58,9%, buk 66,4%, borovica 89,4%, smrek 86,4%, (pre WV2)
7.	Bucha	Slovensko	MLC	Pixelová klasifikácia	Bayassovým klasifikátorom	PCA1, PCA2, PCA3, DMT	15c, 4b, OA od scény 80 - 90%	dub 83,4%, buk 96,2%, borovica 90,1%, smrek 89,3%	dub 79%, buk 100%, borovica 82,1%, smrek 89,3%
8.	Sothe a kol.	Brazília	SVM	Pixelová klasifikácia	RBF (cost a gamma)	C, B, G, R, RE1, RE2, RE3, NIR1, NIR2 (S2)	5c, 98,4%	x	x
9.	Le Louarn	Marseille	RF	Objektová klasifikácia	CARTs, 500 trees	B, G, R, NIR	OA 98,5%	x	x
10.	Kärgel a Janzen	Barnim a Uckermark, Nemecko	MLC	Pixelová klasifikácia	<b>Neuvedené</b>	B, G, R, RE, NIR	OA 81,3%	dub, buk 89,4%, borovica, smrek	dub, buk 52,5%, borovica, smrek

### 1.3. Spektrálne rozlíšenie senzorov satelitov

Spektrálne rozlíšenie senzorov satelitov									
[nm]	satelit	Landsat-4 a 5	Landsat-8	Sentinel-2	WorldView-2	Quick-Bird	Ikonos 2	RapidEye	Pléiades
č. pásma	pásma/senzor	TM	OLI			BGIS 2000			
1	Ultra Blue (Violet/Coastal/Aerosol/Deep)		430 - 450	433 - 453	400 - 450				
2	Blue	450 - 520	450 - 510	457,5 - 522,5	450 - 510	450 - 520	450 - 530	440 - 510	430 - 550
3	Green	520 - 600	530 - 590	542,5 - 577,5	510 - 580	520 - 600	520 - 610	520 - 590	500 - 620
4	Yellow				585 - 625				
5	Red	630 - 690	640 - 670	650 - 680	630 - 690	630 - 690	640 - 720	630 - 685	590 - 710
6	(Vegetation) Red Edge1			697,5 - 712,5				690 - 730	
7	(Vegetation) Red Edge2			732,5 - 747,5	705 - 745				
8	(Vegetation) Red Edge3			773 - 793					
9	Near Infrared 1 (wide)	760 - 900		784,5 - 899,5	770 - 895	760 - 900	760 - 860	760 - 850	740 - 940
10	Near Infrared 2 (narrow)		850 - 880	855 - 875	860 - 1040				
11	Panchromatic		500 - 680		450 - 800	450 - 900	450 - 900		470 - 830
12	Water Vapor			935 - 955					
13	Cirrus clouds		1360 - 1380	1360 - 1390					
14	Shortwave Infrared 1	1550 - 1750	1570 - 1650	1565 - 1655					
15	Shortwave Infrared 2	2080 - 2350	2110 - 2290	2100 - 2280					
16	Thermal infrared 1	10400 - 12500	10620 - 11190						
17	Thermal infrared 2		11500 - 12510						

- hodnoty v tabuľke sú udávané v nanometroch [nm]

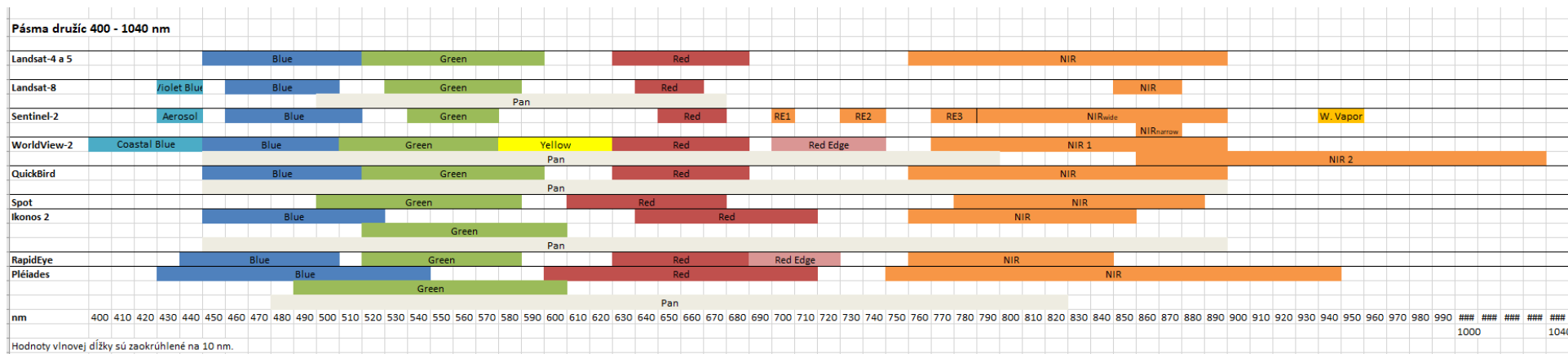
#### 1.4. Priestorové rozlíšenie senzorov satelitov

Priestorové rozlíšenie senzorov satelitov									
[m]	satelit	Landsat-4 a 5	Landsat-8	Sentinel-2	WorldView-2	Quick-Bird	Ikonos 2	RapidEye	Pléiades
č. pásma	pásma/senzor	TM	OLI			BGIS 2000			
1	Ultra Blue (Violet/Coastal/Aerosol/Deep)		30	60	1,84				
2	Blue	30	30	10	1,84	2,62	3,2	5	2
3	Green	30	30	10	1,84	2,62	3,2	5	2
4	Yellow				1,84				
5	Red	30	30	10	1,84	2,62	3,2	5	2
6	(Vegetation) Red Edge1			20				5	
7	(Vegetation) Red Edge2			20	1,84				
8	(Vegetation) Red Edge3			20					
9	Near Infrared 1 (wide)	30		10	1,84	2,62	3,2	5	2
10	Near Infrared 2 (narrow/RE4)		30	20	1,84				
11	Panchromatic		15		0,46	0,651	0,82		0,5
12	Water Vapor								
13	Cirrus clouds		30	60					
14	Shortwave Infrared 1	30	30	60					
15	Shortwave Infrared 2	30	30	20					
16	Thermal infrared 1	120* (30)	100* (30)	20					
17	Thermal infrared 2		100* (30)						

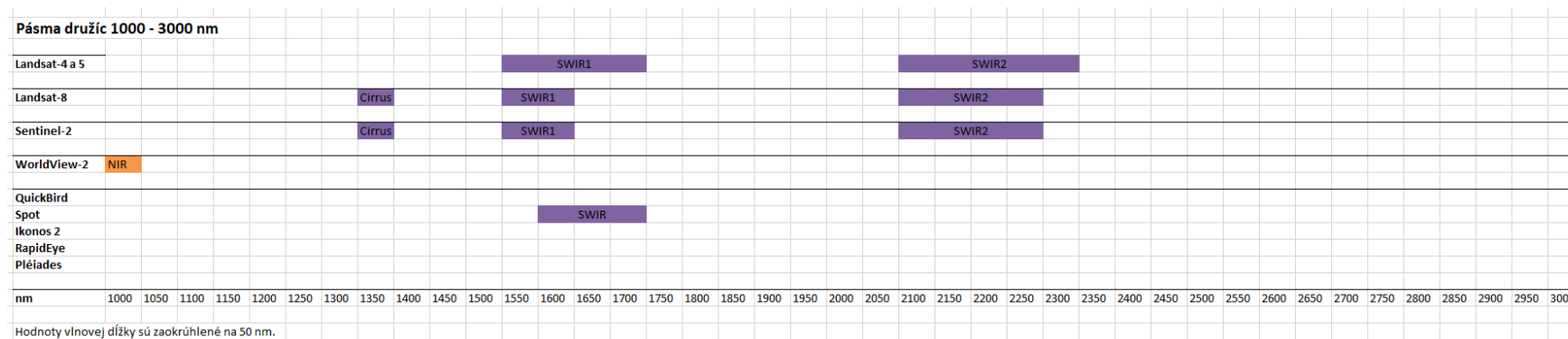
\*snímky z pásiem sú získané s rozlíšením 100/120 m, ale pre užívateľov sú poskytované v prevzorokovanej podobe 30 m

- hodnoty v tabuľke sú udávané v metroch [m]

## 1.5. Pásma družíc 400 - 1040 nm



## 1.6. Pásma družíc 1000 - 3000 nm



### 1.7. Vlastnosti lokalít a georeferencovania mapových podkladov

DUB	LOKALITY						GEOREFERENCOVANIE			
				Veľkosť plochy		Počet pixelov		Vlicovacie body		
č. lokality	Názov lokality	Katastrálne územie	Okres	[m <sup>2</sup> ]	[ha]	L-8	S-2	Urobené	Použité	presnosť [m]
1	Nechvalice	Velvěty	Teplice	415749	41,57	462	4157	x	x	x
2	Libčice nad Vltavou	Libčice nad. Vltavou	Praha - západ	105644	10,56	117	1056	x	x	x
3	Evaň 1	Horka u Libochovic	Litoměřice	129075	12,91	143	1291	3	3	2,3
4	Evaň 2	Levousy	Litoměřice	107803	10,78	120	1078	3	3	2,3
5	Bažantnice	Nehalozeves	Mělník	108007	10,80	120	1080	x	x	x
6	Masečín	Slapy nad Vltavou	Praha - západ	135752	13,58	151	1358	3	3	4

BUK	LOKALITY						GEOREFERENCOVANIE			
				Veľkosť plochy		Počet pixelov		Vlicovacie body		
č. lokality	Názov lokality	Katastrálne územie	Okres	[m <sup>2</sup> ]	[ha]	L-8	S-2	Urobené	Použité	presnosť [m]
7	Osek	Hrad Osek	Teplice	165085	16,51	183	1651	4	4	2,3
8	Jiřetín 1	Horní Jiřetín	Most	152485	15,25	169	1525	7	7	1,1
9	Jiřetín 2	Černice u Horno Jiřetína	Most	134501	13,45	149	1345	7	7	1,1
10	Jiřetín 3	Černice u Horno Jiřetína	Most	115695	11,57	129	1157	7	7	1,1
11	Bečov	Blatno	Chomútov	865929	86,59	962	8659	7	7	4,2
12	Kalivody	Kalivody	Rakovník	329927	32,99	367	3299	9	7	5,6

SMREK	LOKALITY							GEOREFERENCOVANIE		
				Veľkosť plochy		Počet pixelov		Vlicovacie body		
č. lokality	Názov lokality	Katastrálne územie	Okres	[m <sup>2</sup> ]	[ha]	L-8	S-2	Urobené	Použité	presnosť [m]
13	Kryštofovy Hamry	Přísečnice	Chomútov	77736	7,77	86	777	9	7	5,2
14	Klinovec	Jachymov	Jáchymov	415897	41,59	462	4159	7	4	3,7
15	Vršek	Jachymov	Jáchymov	177917	17,79	198	1779	6	4	3,8
16	Nová Ves u Krasnic	Krasnice	Sokolov	377570	37,76	420	3776	5	5	2,3
17	Stará Voda 1	Nové Mohelno	Cheb	120764	12,08	134	1208	6	5	1,8
18	Stará Voda 2	Slatina u Staré Vody	Cheb	231137	23,11	257	2311	5	4	1,9

BOROVICA	LOKALITY							GEOREFERENCOVANIE		
				Veľkosť plochy		Počet pixelov		Vlicovacie body		
č. lokality	Názov lokality	Katastrálne územie	Okres	[m <sup>2</sup> ]	[ha]	L-8	S-2	Urobené	Použité	presnosť [m]
19	Prameny	Prameny	Cheb	320315	32,03	356	3203	7	4	4,5
20	Studánka	Studánka u Tachova	Tachov	213013	21,30	237	2130	7	5	0,9
21	Benešovice	Holostřevy	Tachov	411408	41,14	457	4114	6	5	0,9
22	Plzeň 1	Zálužní u Třemošní	Plzeň - sever	284620	28,46	316	2846	7	7	1,7
23	Plzeň 2	Chotíkov	Plzeň - sever	294090	29,41	327	2941	8	7	1,7
24	Valov (Podbořany)	Vroutek	Louny	256290	25,63	285	2563	8	7	3,7



## 1.8. Prehľad scén

Číslo	Družica	Level predspracovania dát	Dátum	Čas	Oblačnosť	ID číslo scény
1	Landsat-8	Level-1TP	24.6.2016	9:56:41	0,4	LC81920252016176LGN01
2	Landsat-8	Level-1TP	27.8.2016	9:57:03	0,02	LC81920252016240LGN01
3	Landsat-8	Level-1TP	30.8.2017	9:27:00	0,04	LC81920252017242LGN00
4	Landsat-8	Level-2	24.6.2016	9:56:41	0,4	LC81920252016176LGN01
5	Landsat-8	Level-2	27.8.2016	9:57:03	0,02	LC81920252016240LGN01
6	Landsat-8	Level-2	30.8.2017	9:27:00	0,04	LC81920252017242LGN00
7	Sentinel-2	Level-1C	4.7.2015	10:13:37	0,47	S2A_OPER_MSI_L1C_DS_EPA__20160809T015434_S20150704T101337_N02.04
8	Sentinel-2	Level-1C	13.8.2015	10:16:17	1,93	S2A_OPER_MSI_L1C_DS_EPA__20160720T233845_S20150813T101617_N02.04
9	Sentinel-2	Level-1C	27.8.2016	10:10:22	0	S2A_OPER_MSI_L1C_DS_SGS__20160827T153533_S20160827T101025_N02.04

1. Oblačnosť je udávaná v [%]

## 1.9. Rešerš 33 prác

**Table 11.** Results from tree species classifications in temperate and boreal forests with spectral remote sensing data using different sensors and classification methods from the last 10 years (ordered by the number of tree species and overall accuracy).

Platform <sup>1</sup> Datatyp <sup>2</sup>	Sensor	Acquisition date	LIDAR <sup>3</sup>	Classification algorithm <sup>4</sup>	Approach <sup>5</sup>	Species no.	Overall acc. [%]	Kappa <sup>6</sup>	Tree species	Reference
A M	DMC	Jun	x	ML	OI	3	<b>96</b>	<b>0.93</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> & broadleaf trees	[69]
A M	DMC	Oct		ML	OI	3	<b>91</b>	<b>0.87</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> & broadleaf trees	[69]
A M	Wild RC 20	May & Jul		MD	OI	3	<b>91</b>	<b>0.86</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> & <i>Betula pendula</i>	[84]
A M	DMC	Oct		LDA	OI	3	<b>89</b>	<b>0.82</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> & broadleaf trees	[50]
A M	ADS40-SH52	Aug		SVM	OI	3	<b>88</b>	n.s.	<i>Pinus sylvestris</i> , <i>Picea abies</i> & <i>Betula sp.</i>	[85]
H M	DCRTRV20	Oct		ML	OS	3	<b>87</b>	n.s.	<i>Larix sp.</i> , <i>Cryptomeria japonica</i> , <i>Fagus sp.</i>	[75]
A M	DMC	Jun		ML	OI	3	<b>84</b>	<b>0.76</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> & broadleaf trees	[69]
A M	n.s.	n.s.	x	DT	OI	3	<b>84</b>	n.s.	<i>Fagus sylvatica</i> , <i>Quercus sp.</i> / <i>Carpinus betulus</i> & conifers	[73]
A M	Wild RC 20	Jun		LDA	OI	3	<b>78</b>	n.s.	<i>Picea abies</i> , <i>Pinus sylvestris</i> & broadleaf trees	[47]
A M	DMC	Sep		LDA	OI	3	<b>68</b>	n.s.	<i>Picea abies</i> , <i>Pinus sylvestris</i> & <i>Betula sp.</i>	[52]
A M	ADS40-SH52	Sep		ANN	OI	4	<b>84</b>	<b>0.73</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Larix decidua</i> & <i>Betula sp.</i>	[61]
A M	Wild RC30/4	Aug		DT	OI	4	<b>77</b>	n.s.	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Betula pubescens</i> & <i>Populus tremula</i>	[86]
A M	n.s.	Aug		LDA	OI	4	<b>67</b>	n.s.	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Betula pubescens</i> & <i>Populus tremula</i>	[49]
S M	ASTER	Apr & Jun		ACA	OS	5	<b>87</b>	<b>0.83</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Pseudotsuga menziesii</i> , <i>Fagus sylvatica</i> , & <i>Quercus sp.</i>	[87]
S M	ASTER	Apr & Jun		ML	OS	5	<b>82</b>	<b>0.77</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Pseudotsuga menziesii</i> , <i>Fagus sylvatica</i> , & <i>Quercus sp.</i>	[87]
A M	CASI	Jun		ML	OI	6	<b>93</b>	n.s.	<i>Pseudotsuga menziesii</i> , <i>Abies grandis</i> , <i>Abies amabilis</i> , <i>Thuja plicata</i> , <i>Tsuga heterophylla</i> & hardwood	[88]
A M	ATM	Oct		ML	OI	6	<b>84</b>	<b>0.79</b>	<i>Fraxinus excelsior</i> , <i>Quercus robur</i> , <i>Acer campestre</i> , <i>Betula pendula</i> , <i>Populus tremula</i> & <i>Ulmus minor</i>	[89]
A M	ATM	Mar, Jul & Oct		ML	OI	6	<b>71</b>	<b>0.63</b>	<i>Fraxinus excelsior</i> , <i>Quercus robur</i> , <i>Acer campestre</i> , <i>Betula pendula</i> , <i>Populus tremula</i> & <i>Ulmus minor</i>	[89]
A H	AISA	Jul	x	NN	OI	6	<b>57</b>	n.s.	<i>Pinus strobus</i> , <i>Picea glauca</i> , <i>Gleditsia triacanthos</i> , <i>Acer saccharum</i> , <i>Tilia Americana</i> & <i>Quercus palustris</i>	[63]
A H	AISA	Jul		NN	OI	6	<b>48</b>	n.s.	<i>Pinus strobus</i> , <i>Picea glauca</i> , <i>Gleditsia triacanthos</i> , <i>Acer saccharum</i> , <i>Tilia Americana</i> & <i>Quercus palustris</i>	[63]
A H	AISA	Oct		NN	OI	6	<b>45</b>	n.s.	<i>Pinus strobus</i> , <i>Picea glauca</i> , <i>Gleditsia triacanthos</i> , <i>Acer saccharum</i> , <i>Tilia Americana</i> & <i>Quercus palustris</i>	[63]
S M	IKONOS	Jun		ML	P	7	<b>86</b>	<b>0.84</b>	<i>Pinus sylvestris</i> , <i>Pinus nigra subsp. Laricio</i> , <i>Larix decidua</i> , <i>Pseudotsuga menziesii</i> , <i>Fagus sylvatica</i> , <i>Fagus sylvatica purpure</i> & <i>Quercus sp.</i>	[60]
A H	AISA	Jul	x	SVM	OI	7	<b>83</b>	<b>0.77</b>	<i>Picea abies</i> , <i>Abies alba</i> , <i>Pinus mugo</i> , <i>Pinus sylvestris</i> , <i>Larix decidua</i> , <i>Fagus sylvatica</i> , & other broadleaf (& non forest)	[27]
S M	IKONOS	May		ML	PA	7	<b>77</b>	<b>0.73</b>	<i>Pinus rigida</i> , <i>Pinus koraiensis</i> , <i>Larix leptolepis</i> , <i>Quercus mongolica</i> , <i>Quercus variabilis</i> , <i>Quercus acutissima</i> & <i>Castanea crenata</i>	[90]
A M	ADS40-SH40 & RC30	Sep		ANN	OI	7	<b>76</b>	<b>0.70</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Abies alba</i> , <i>Larix decidua</i> , <i>Fagus sylvatica</i> , <i>Fraxinus excelsior</i> , <i>Acer sp.</i> , <i>Alnus sp.</i> , & <i>Betula sp.</i>	[61]
A H	MIVIS	Jul		ML	OI	7	<b>75</b>	<b>0.67</b>	<i>Pinus sp.</i> , <i>Castanea sativa</i> , <i>Alnus sp.</i> , <i>Salix sp.</i> , <i>Populus sp.</i> , <i>Quercus sp.</i> & Alien species	[91]
A H	AISA	Jul		SVM	OI	7	<b>74</b>	<b>0.66</b>	<i>Picea abies</i> , <i>Abies alba</i> , <i>Pinus mugo</i> , <i>Pinus sylvestris</i> , <i>Larix decidua</i> , <i>Fagus sylvatica</i> , & other broadleaf (& non forest)	[27]
A M	DMC	Aug		LR	OI	8	<b>88</b>	<b>0.86</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Acer sp.</i> , <i>Fagus sp.</i> , <i>Fraxinus sp.</i> , <i>Populus sp.</i> , <i>Salix sp.</i> & <i>Quercus sp.</i>	[62]
A H	AISA	Jul		SVM	P	11	n.s.	<b>0.75</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Pinus mugo</i> , <i>Abies alba</i> , <i>Larix decidua</i> , <i>Fagus sylvatica</i> , <i>Fraxinus excelsior</i> , <i>Acer pseudoplatanus</i> , <i>Ostrya carpinifolia</i> , <i>Quercus petraea</i> , <i>Ulmus glabra</i> (& non forest)	[92]
A H	AISA	Jul	x	SVM	OI	11	<b>74</b>	<b>0.60</b>	<i>Abies grandis</i> , <i>Thuja plicata</i> , <i>Pseudotsuga menziesii</i> , <i>Tsuga heterophylla</i> , <i>Pinus contorta</i> , <i>Populus balsamifera</i> , <i>Populus tremuloides</i> , <i>Alnus rubra</i> , <i>Acer macrophyllum</i> , <i>Quercus garryana</i> & <i>Arbutus menziesii</i> ,	[64]
A H	AISA	Jul		SVM	OI	11	<b>72</b>	<b>0.60</b>	<i>Abies grandis</i> , <i>Thuja plicata</i> , <i>Pseudotsuga menziesii</i> , <i>Tsuga heterophylla</i> , <i>Pinus contorta</i> , <i>Populus balsamifera</i> , <i>Populus tremuloides</i> , <i>Alnus rubra</i> , <i>Acer macrophyllum</i> , <i>Quercus garryana</i> & <i>Arbutus menziesii</i> ,	[64]
S M	GeoEye-1	Sep		SVM	P	11	n.s.	<b>0.51</b>	<i>Picea abies</i> , <i>Pinus sylvestris</i> , <i>Pinus mugo</i> , <i>Abies alba</i> , <i>Larix decidua</i> , <i>Fagus sylvatica</i> , <i>Fraxinus excelsior</i> , <i>Acer pseudoplatanus</i> , <i>Ostrya carpinifolia</i> , <i>Quercus petraea</i> , <i>Ulmus glabra</i> (& non forest)	[92]
S M	IKONOS	Jul		ML	P	16	<b>58</b>	<b>0.50</b>	<i>Abies concolor</i> , <i>Abies magnifica</i> , <i>Juniperus occidentalis</i> , <i>Pinus albicaulis</i> , <i>Pinus contorta</i> , <i>Pinus jeffreyi</i> , <i>Pinus monticola</i> , <i>Tsuga mertensiana</i> , <i>Populus tremuloides</i> , <i>Alnus incana</i> , <i>Arctostaphylos patula</i> , <i>Artemisia tridentata</i> , <i>Ceanothus cordulatus</i> , <i>Ceanothus velutinus</i> , <i>Quercus vaccinifolia</i> , <i>Salix sp.</i> (& 3 grass species: results without class 'water')	[39]

<sup>1</sup> A: Airplane, H: Helicopter, S: Satellite; <sup>2</sup> H: Hyperspectral, M: Multispectral; <sup>3</sup> Additional information for the tree species classification; <sup>4</sup> ACA: Adaptive Classification Approach, DT: Decision Tree, LDA: Linear Discriminant Analysis, LR: Logistic Regression, MD: Minimum Distance, ML: Maximum Likelihood, NN: Nearest Neighbor, ANN: Artificial Neural Network, SVM: Support Vector Machine; <sup>5</sup> OI: object-based classification of individual trees or clusters of some trees, OS: object-based classification of stands, P: pixel-based classification, PA: pixel-based classification and aggregation of classified pixels to objects; <sup>6</sup> n.s.: not specified.

Autor: Immitzer M. a kol. [4], upravené

## PRÍLOHA 2: CHYBOVÉ MATICE KLASIFIKÁCIE SNÍMOK

Tabuľka 24: Chybová matica scény L8\_L1\_160624

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
Druhy skutočné	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	392	88,69	50	11,31	0	0,00	0	0,00	88,69
Buk	15	1,92	763	97,70	0	0,00	3	0,38	97,70
Smrek	0	0,00	1	0,16	587	94,07	36	5,77	94,07
Borovica	0	0,00	5	0,70	21	2,95	686	96,35	96,35
Procesná presnosť	407	96,31	819	93,16	608	96,55	725	94,62	

Tabuľka 25: Chybová matica scény L8\_L1\_160827

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
Druhy skutočné	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	383	90,54	35	8,27	0	0,00	5	1,18	90,54
Buk	17	2,27	727	97,19	4	0,53	0	0,00	97,19
Smrek	0	0,00	0	0,00	581	96,35	22	3,65	96,35
Borovica	8	1,02	0	0,00	39	4,97	738	94,01	94,01
Procesná presnosť	408	93,87	762	95,41	624	93,11	765	96,47	

Tabuľka 26: Chybová matica scény L8\_L1\_170830

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
Druhy skutočné	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	372	91,63	30	7,39	1	0,25	3	0,74	91,63
Buk	44	5,74	721	94,00	1	0,13	1	0,13	94,00
Smrek	5	0,83	0	0,00	573	95,02	25	4,15	95,02
Borovica	4	0,53	0	0,00	5	0,66	750	98,81	98,81
Procesná presnosť	425	87,53	751	96,01	580	98,79	779	96,28	

Tabuľka 27: Chybová matica scény L8\_L2\_160624

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
Druhy skutočné	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	410	91,31	39	8,69	0	0,00	0	0,00	91,31
Buk	10	1,28	768	98,59	1	0,13	0	0,00	98,59
Smrek	0	0,00	0	0,00	569	93,74	38	6,26	93,74
Borovica	6	0,83	0	0,00	8	1,10	710	98,07	98,07
Procesná presnosť	426	96,24	807	95,17	578	98,44	748	94,92	

Tabuľka 28: Chybová matica scény L8\_L2\_160827

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	369	86,42	56	13,11	0	0,00	2	0,47	86,42
Buk	49	6,27	732	93,61	0	0,00	1	0,13	93,61
Smrek	1	0,17	0	0,00	537	89,95	59	9,88	89,95
Borovica	3	0,40	5	0,66	33	4,38	712	94,56	94,56
Procesná presnosť	422	87,44	793	92,31	570	94,21	774	91,99	

Tabuľka 29: Chybová matica scény L8\_L2\_170830

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	407	92,71	26	5,92	3	0,68	3	0,68	92,71
Buk	48	6,20	725	93,67	0	0,00	1	0,13	93,67
Smrek	0	0,00	0	0,00	542	87,14	80	12,86	87,14
Borovica	4	0,55	0	0,00	19	2,62	701	96,82	96,82
Procesná presnosť	459	88,67	751	96,54	564	96,10	785	89,30	

Tabuľka 30: Chybová matica scény S2\_L1\_150704

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	4084	89,66	413	9,07	3	0,07	55	1,21	89,66
Buk	595	7,22	7609	92,29	14	0,17	27	0,33	92,29
Smrek	1	0,02	14	0,22	6082	94,15	363	5,62	94,15
Borovica	106	1,29	14	0,17	582	7,08	7517	91,46	91,46
Procesná presnosť	4786	85,33	8050	94,52	6681	91,03	7962	94,41	

Tabuľka 31: Chybová matica scény S2\_L1\_150813

Druhy predpovedané	Dub		Buk		Smrek		Borovica		Užívateľská presnosť
	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
Dub	4246	92,47	181	3,94	2	0,04	163	3,55	92,47
Buk	129	1,58	7971	97,91	19	0,23	22	0,27	97,91
Smrek	2	0,03	20	0,31	6218	95,75	254	3,91	95,75
Borovica	72	0,87	6	0,07	236	2,86	7938	96,19	96,19
Procesná presnosť	4449	95,44	8178	97,47	6475	96,03	8377	94,76	

Tabuľka 32: Chybová matica scény S2\_L1\_160827

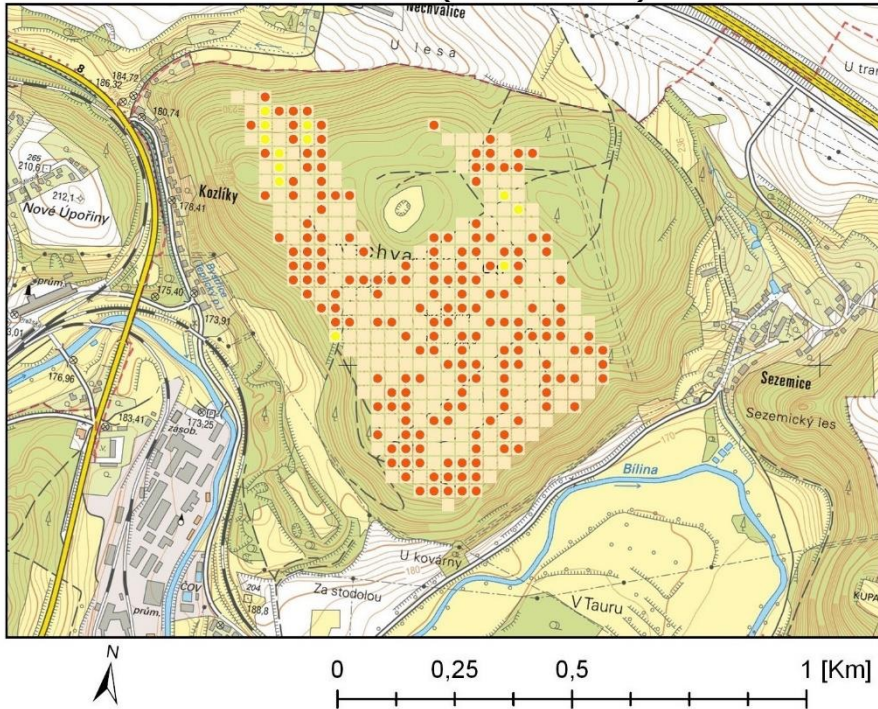
<b>Druhy predpovedané</b>	<b>Dub</b>		<b>Buk</b>		<b>Smrek</b>		<b>Borovica</b>		<b>Užívateľská presnosť</b>
<b>Druhy skutočné</b>	počet	[%]	počet	[%]	počet	[%]	počet	[%]	
<b>Dub</b>	3790	<b>83,00</b>	645	<b>14,13</b>	25	<b>0,55</b>	106	<b>2,32</b>	<b>83,00</b>
<b>Buk</b>	550	<b>6,68</b>	7569	<b>91,91</b>	81	<b>0,98</b>	35	<b>0,43</b>	<b>91,91</b>
<b>Smrek</b>	6	<b>0,09</b>	42	<b>0,65</b>	5956	<b>92,79</b>	415	<b>6,47</b>	<b>92,79</b>
<b>Borovica</b>	101	<b>1,22</b>	13	<b>0,16</b>	1177	<b>14,25</b>	6968	<b>84,37</b>	<b>84,37</b>
<b>Procesná presnosť</b>	4447	<b>85,23</b>	8269	<b>91,53</b>	7239	<b>82,28</b>	7524	<b>92,61</b>	

## **PRÍLOHA 3: MAPOVÉ VÝSTUPY KLASIFIKÁCIE LESNÝCH PORASTOV**



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 1 (Nechvalice)



### Legenda:

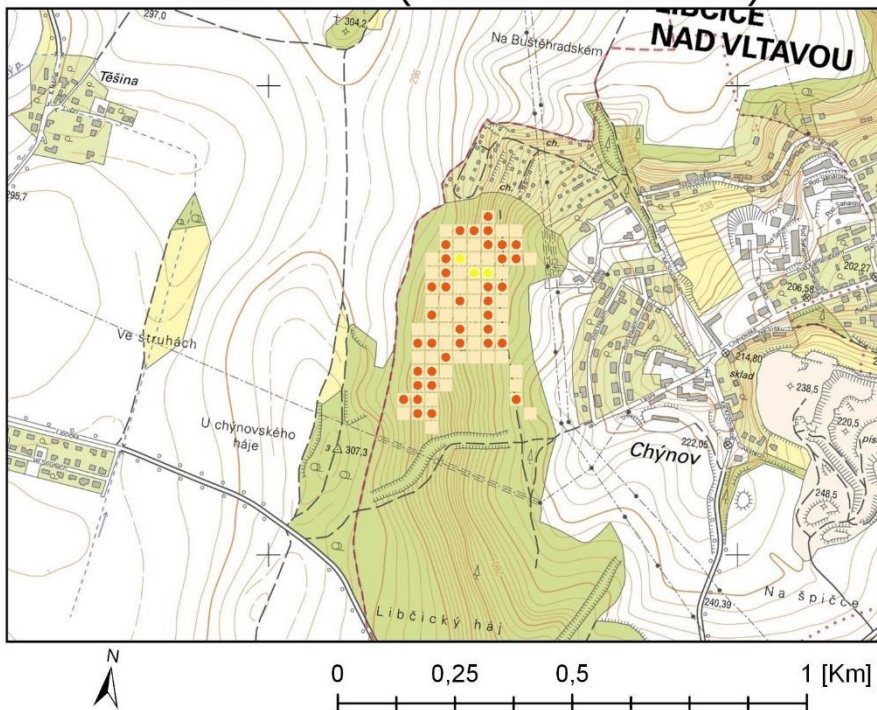
#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Dub
- Základní mapa 1 : 10 000

## Lokalita č. 2 (Libčice nad Vltavou)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

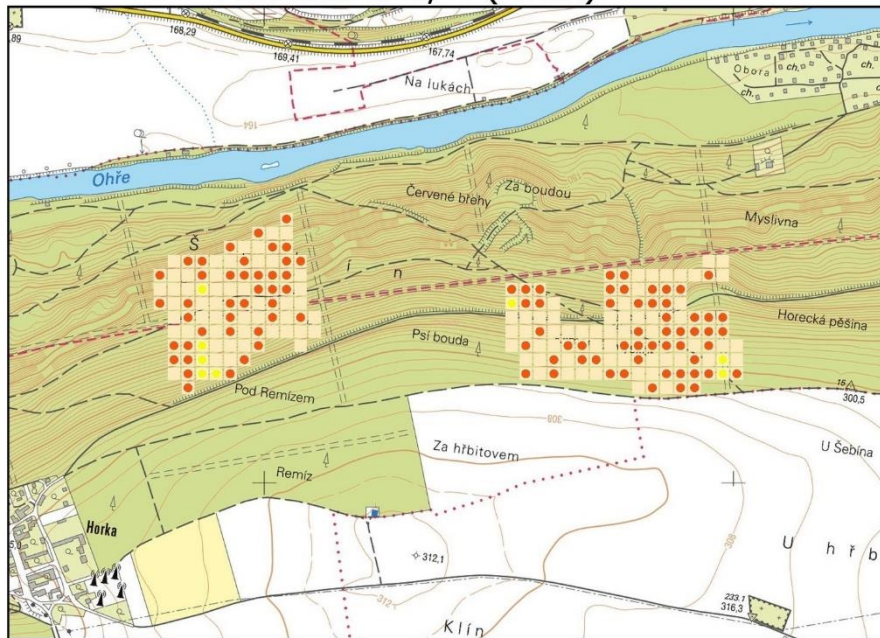
#### Druh lokality:

- Dub
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 3, 4 (Evaň)



### Legenda:

#### Výsledek klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Dub
- Základní mapa 1 : 10 000

## Lokalita č. 5 (Bažantnice)



### Legenda:

#### Výsledek klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

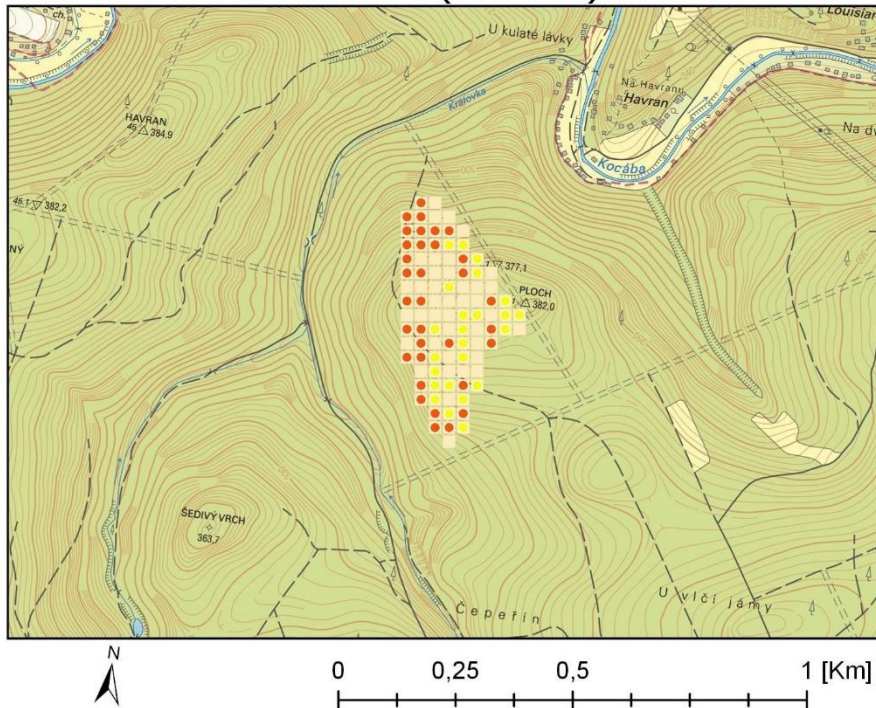
#### Druh lokality:

- Dub
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 6 (Mačetín)



### Legenda:

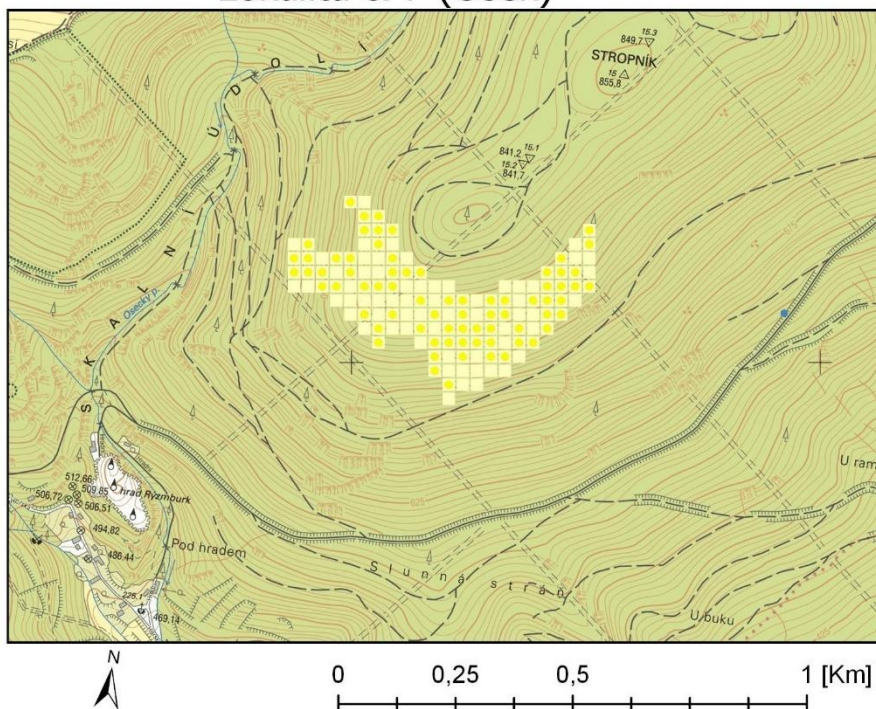
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Dub
- Základní mapa 1 : 10 000

## Lokalita č. 7 (Osek)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

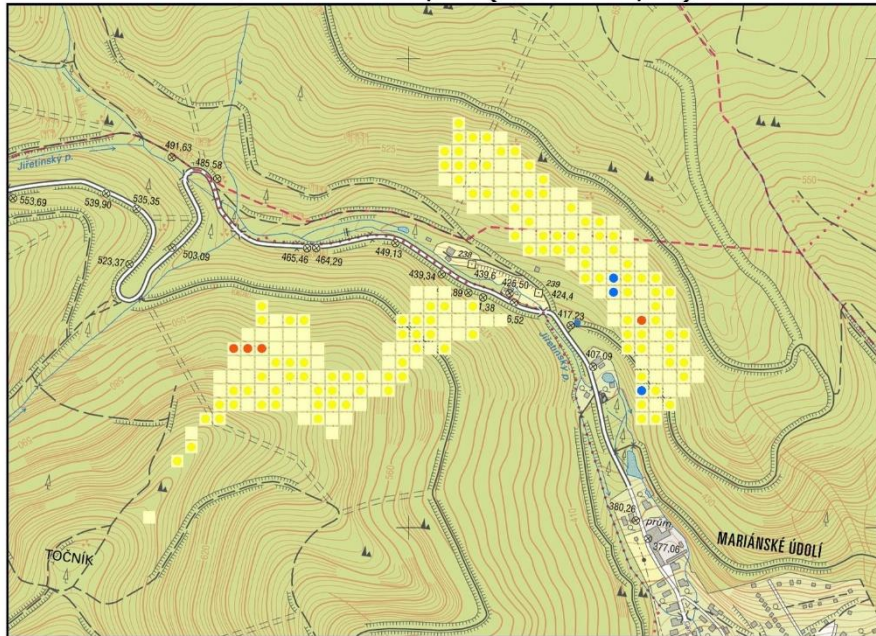
Druh lokality:

- Buk
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 8, 9 (Jiřetín 1, 2)



### Legenda:

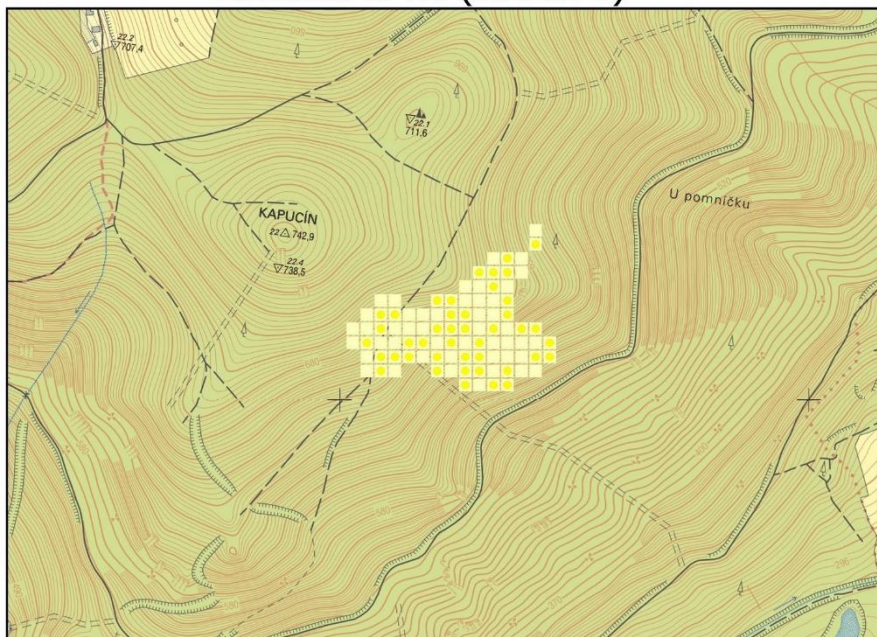
#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Buk
- Základní mapa 1 : 10 000

## Lokalita č. 10 (Jiřetín 3)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

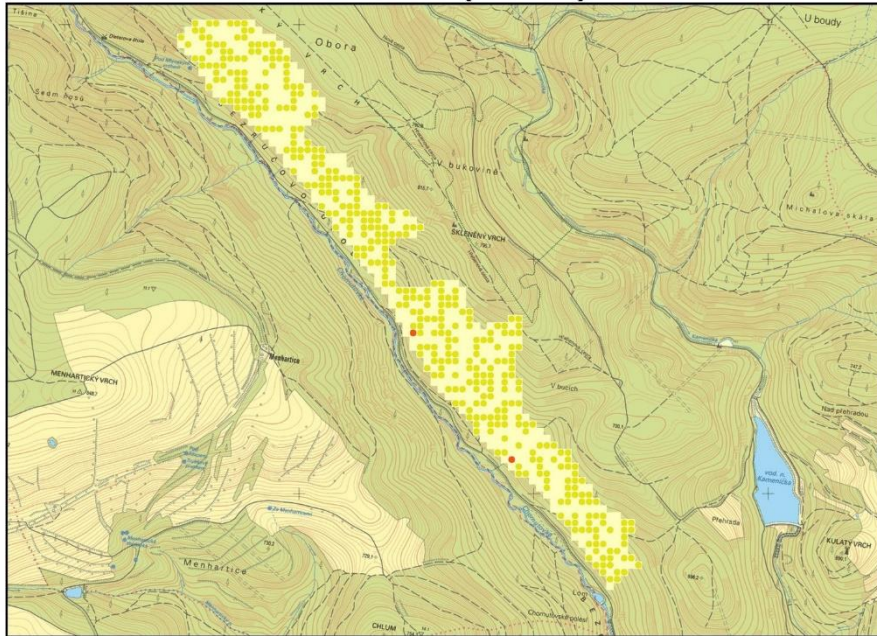
#### Druh lokality:

- Buk
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 11 (Bečov)



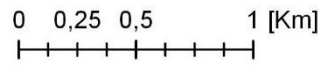
### Legenda:

Výsledok klasifikácie:

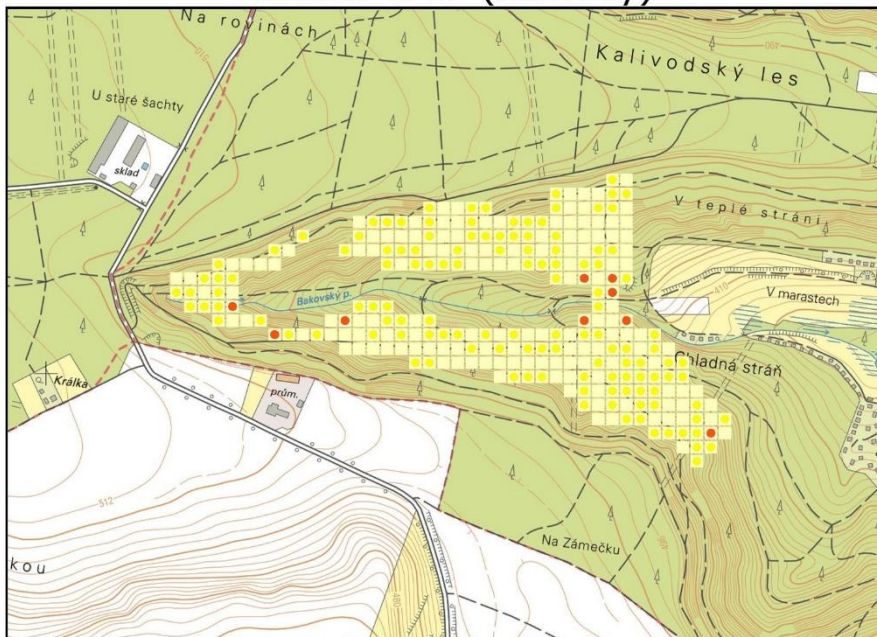
- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Buk
- Základní mapa 1 : 10 000



## Lokalita č. 12 (Kalivody)



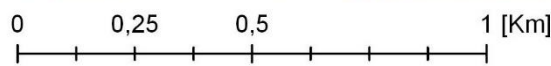
### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

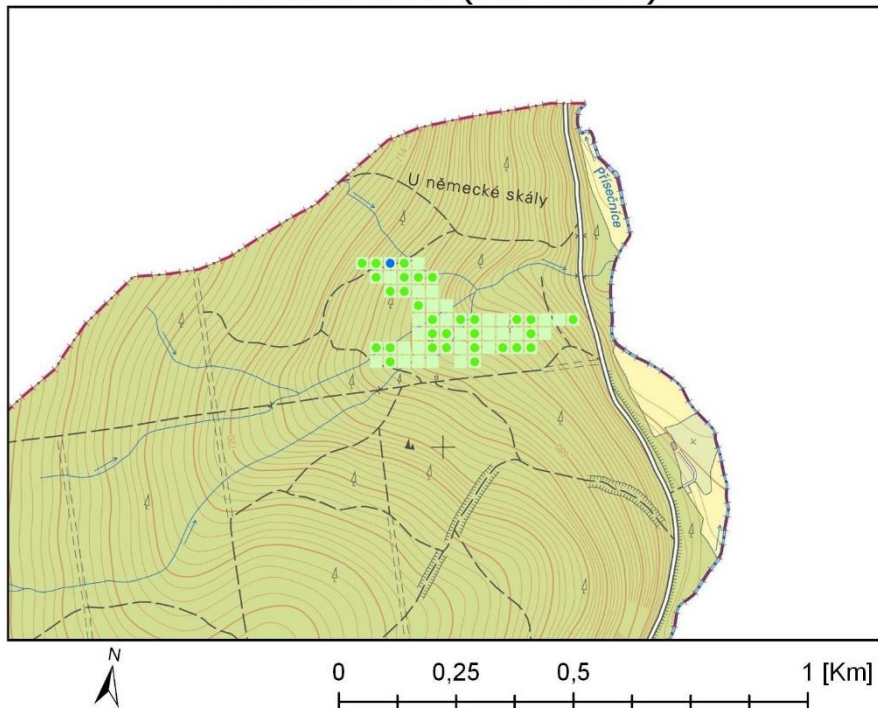
Druh lokality:

- Buk
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 13 (Přísečnice)



### Legenda:

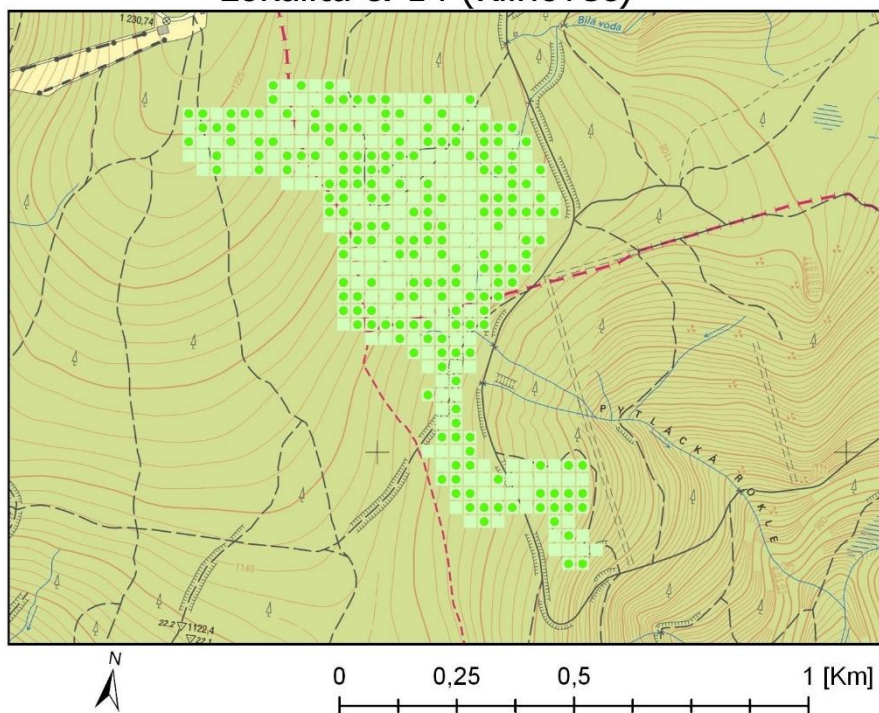
#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Smrek
- Základní mapa 1 : 10 000

## Lokalita č. 14 (Klinovec)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

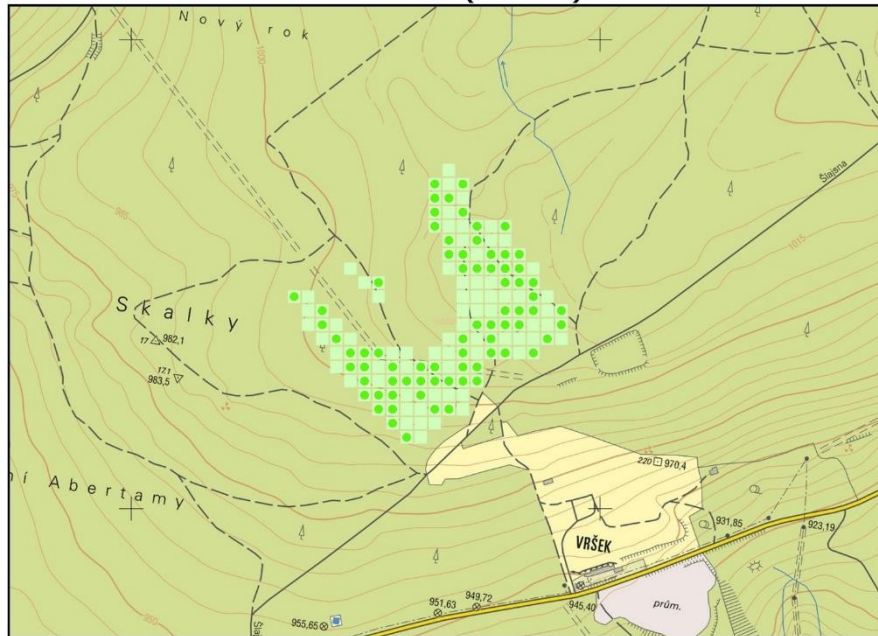
#### Druh lokality:

- Smrek
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 15 (Vršek)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

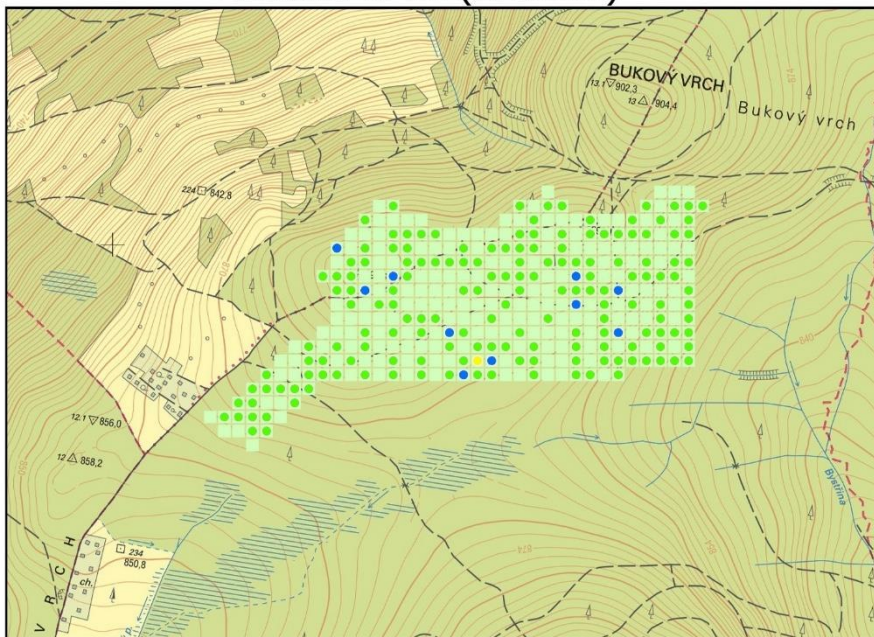
#### Druh lokality:

- Smrek
- Základní mapa 1 : 10 000



0 0,25 0,5 1 [Km]

## Lokalita č. 16 (Kraslice)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

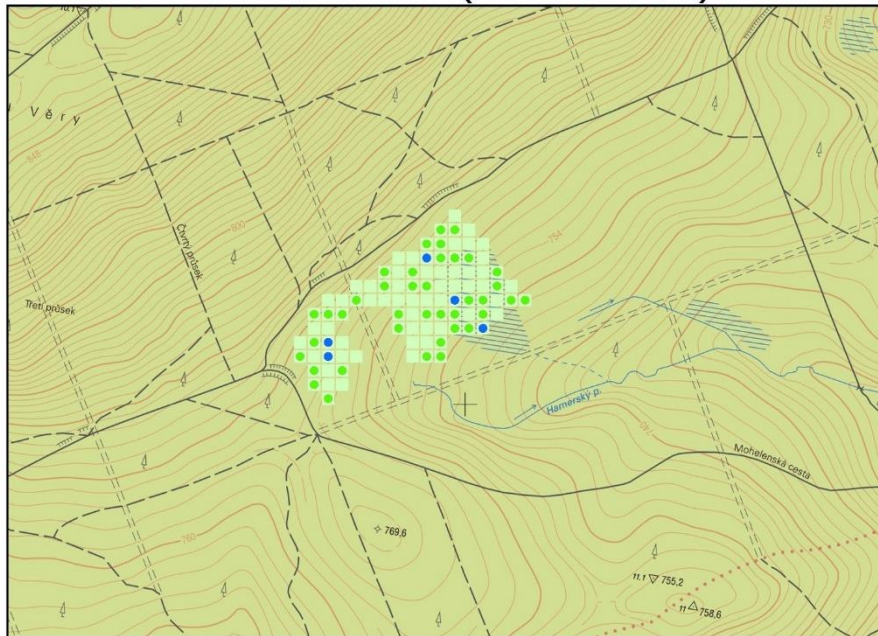
- Smrek
- Základní mapa 1 : 10 000



0 0,25 0,5 1 [Km]

# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 17 (Stará Voda 1)



### Legenda:

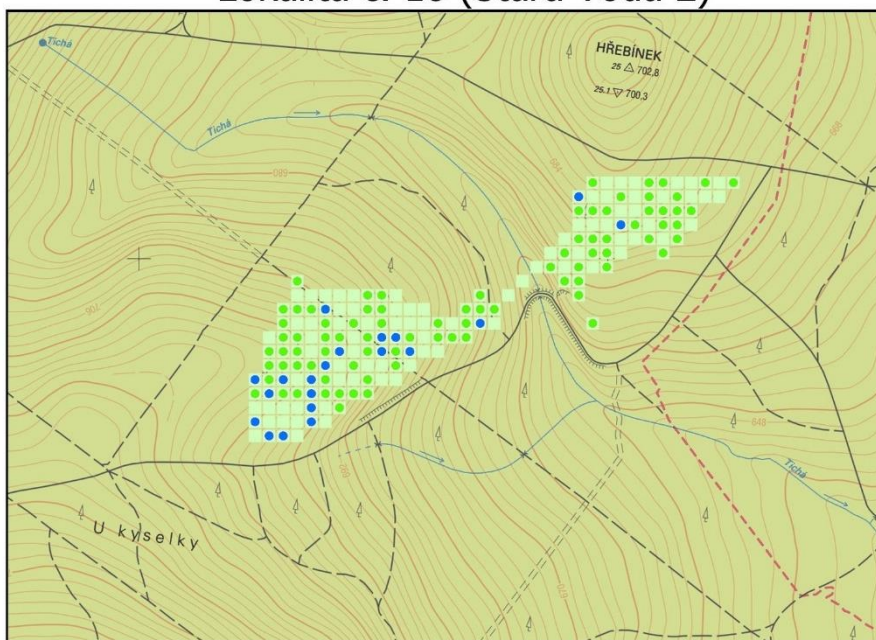
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Smrek
- Základní mapa 1 : 10 000

## Lokalita č. 18 (Stará Voda 2)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

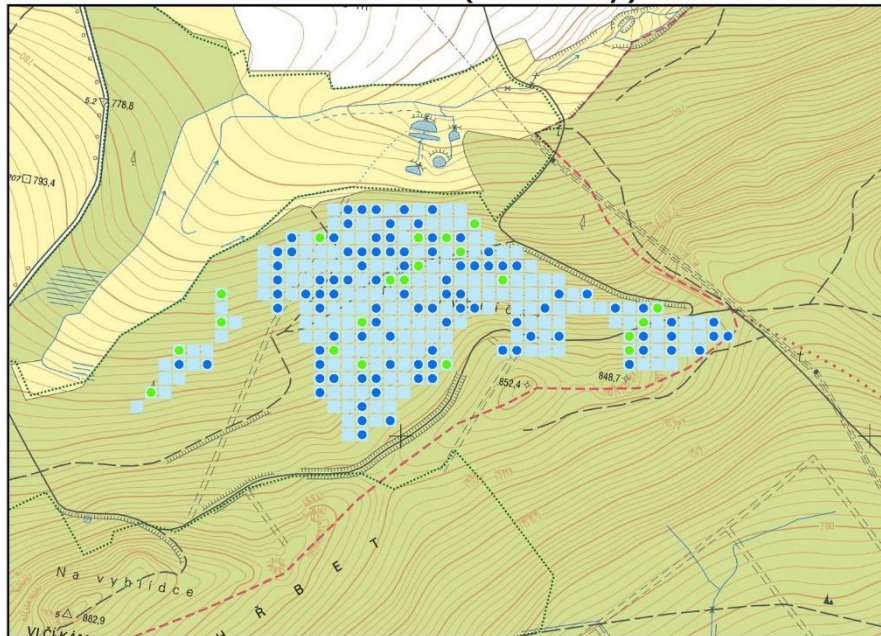
Druh lokality:

- Smrek
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 19 (Prameny)



### Legenda:

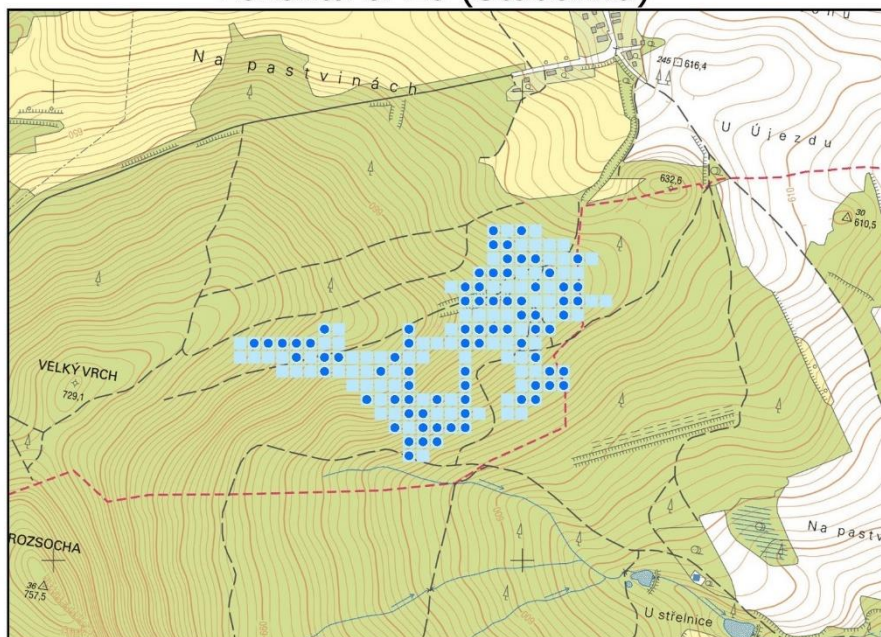
#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Borovica
- Základní mapa 1 : 10 000

## Lokalita č. 20 (Studánka)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

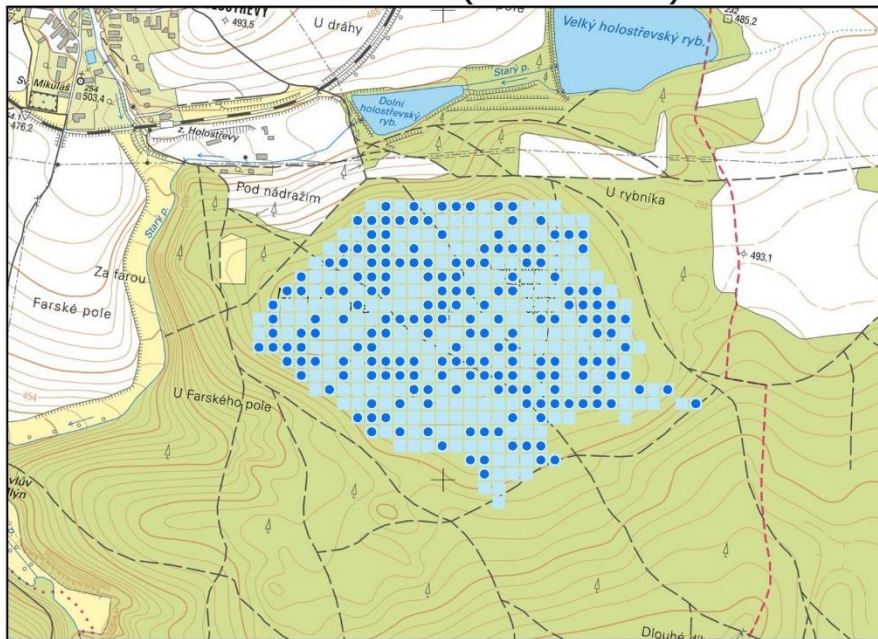
#### Druh lokality:

- Borovica
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 21 (Benešovice)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

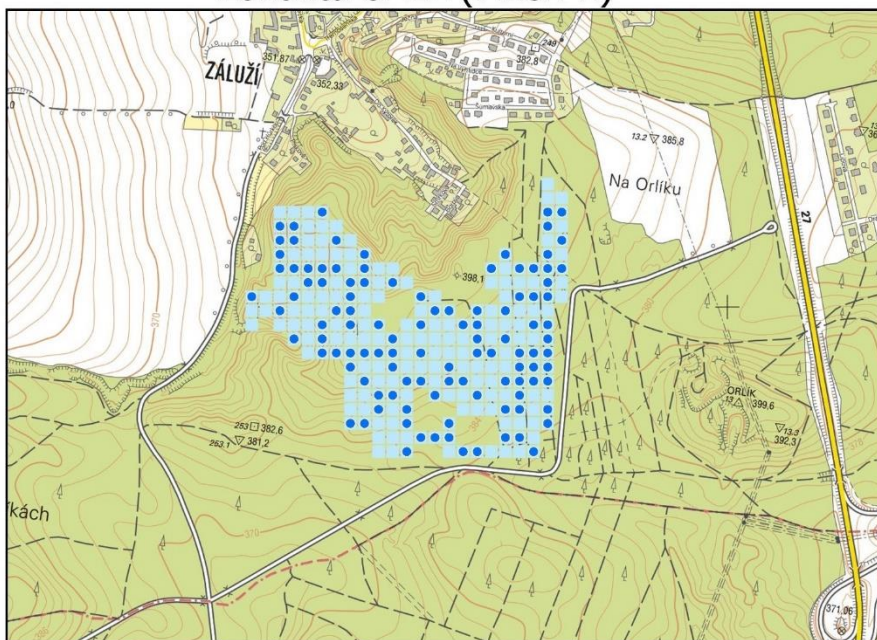
Druh lokality:

- Borovica
- Základní mapa 1 : 10 000



0 0,25 0,5 1 [Km]

## Lokalita č. 22 (Plzeň 1)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Borovica
- Základní mapa 1 : 10 000

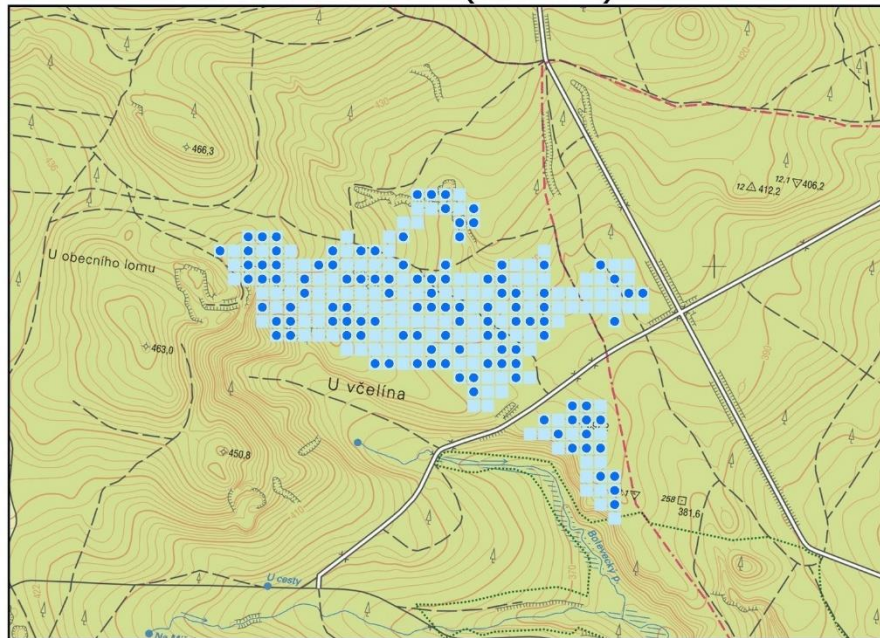


0 0,25 0,5 1 [Km]



# Klasifikácia lesných porastov z dát Landsat-8 L1 z 24.6.2016

## Lokalita č. 23 (Plzeň 2)



### Legenda:

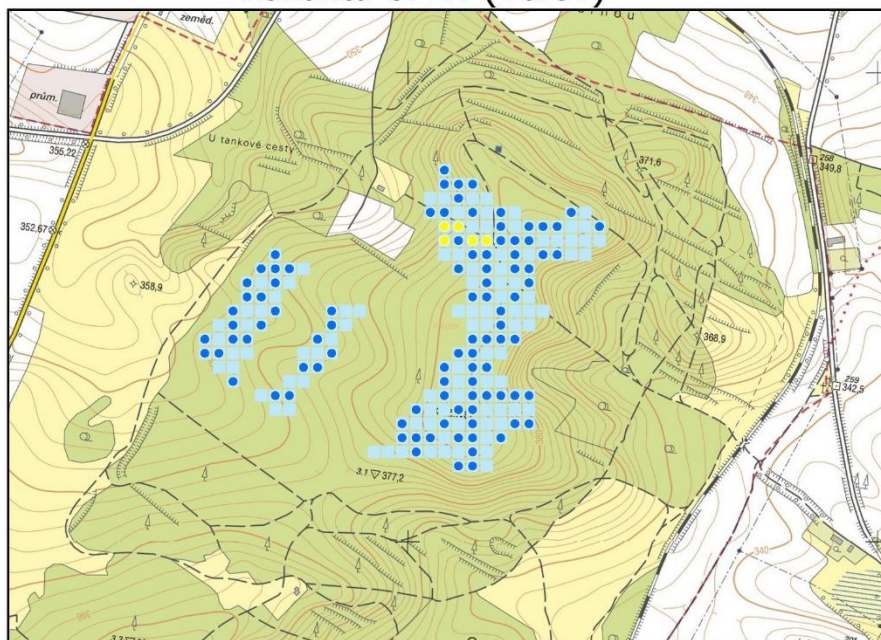
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Borovica
- Základní mapa 1 : 10 000

## Lokalita č. 24 (Valov)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

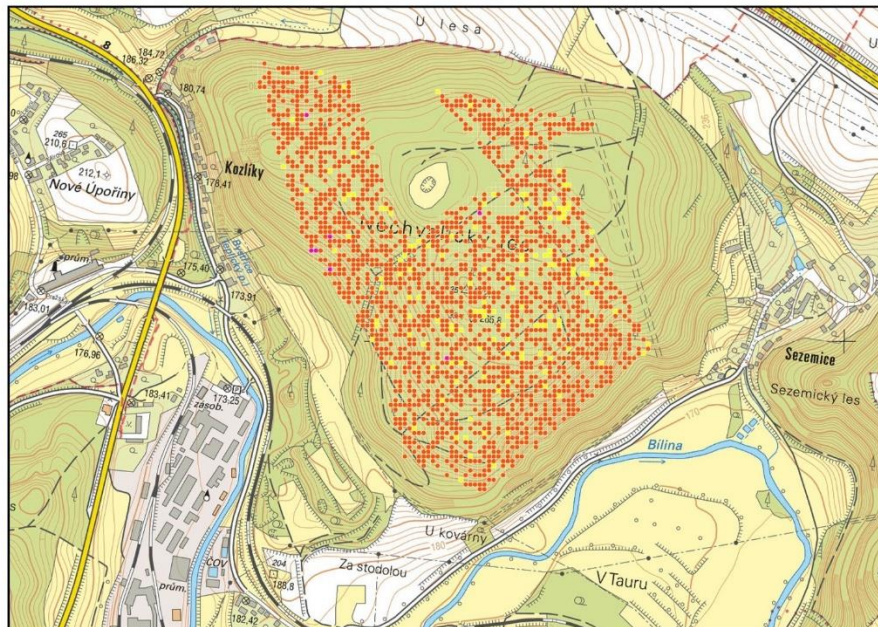
Druh lokality:

- Borovica
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 1 (Nechvalice)



### Legenda:

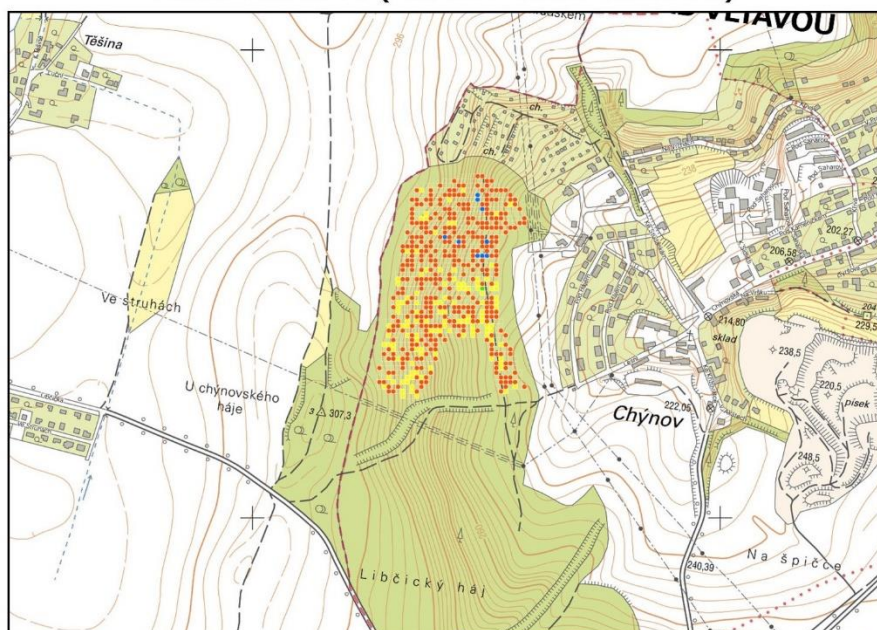
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Dub
- Základní mapa 1 : 10 000

## Lokalita č. 2 (Libčice nad Vltavou)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

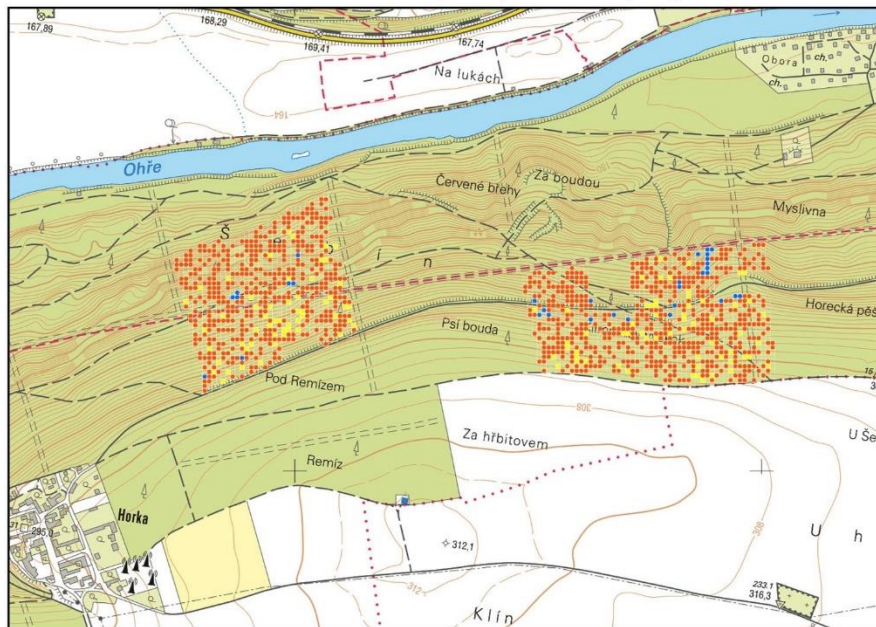
Druh lokality:

- Dub
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 3, 4 (Evaň 1, 2)



### Legenda:

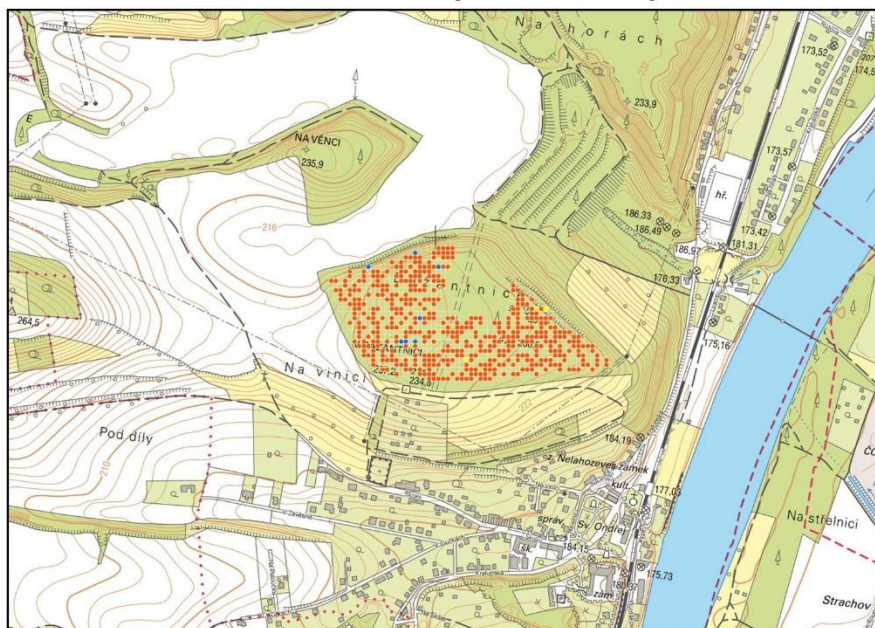
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Dub
- Základní mapa 1 : 10 000

## Lokalita č. 5 (Bažantnice)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

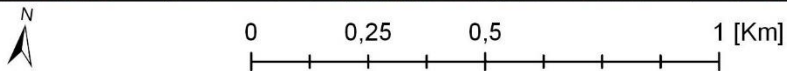
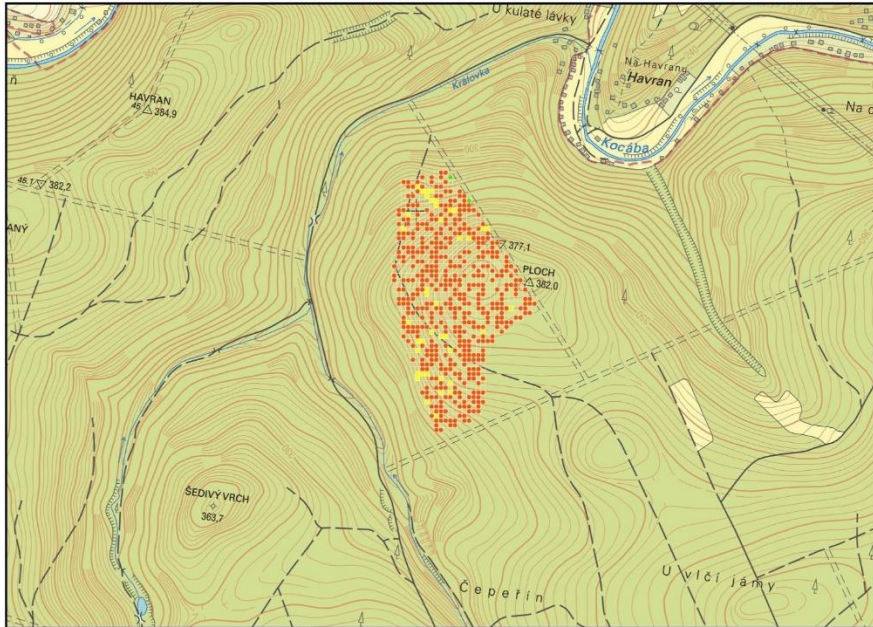
Druh lokality:

- Dub
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 6 (Mačetín)



### Legenda:

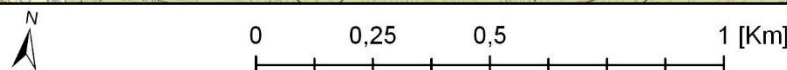
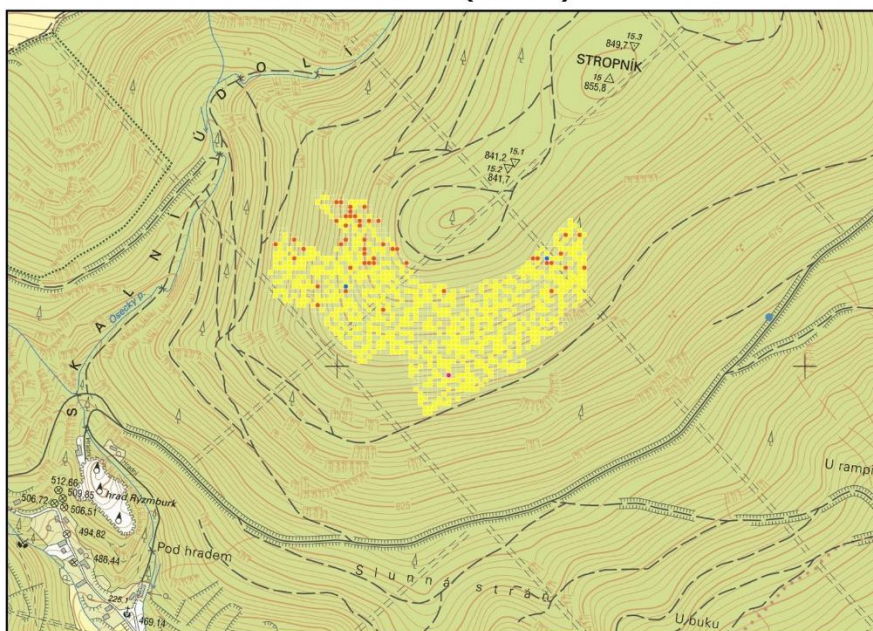
#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Dub
- Základní mapa 1 : 10 000

## Lokalita č. 7 (Osek)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

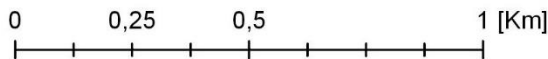
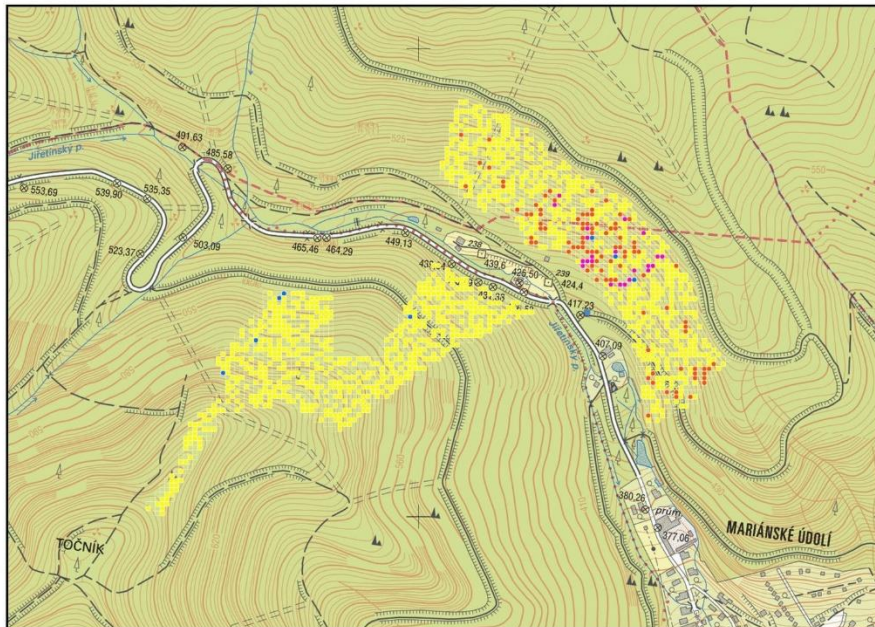
#### Druh lokality:

- Buk
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 8, 9 (Jiřetín 1, 2)



### Legenda:

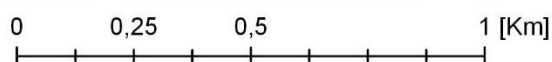
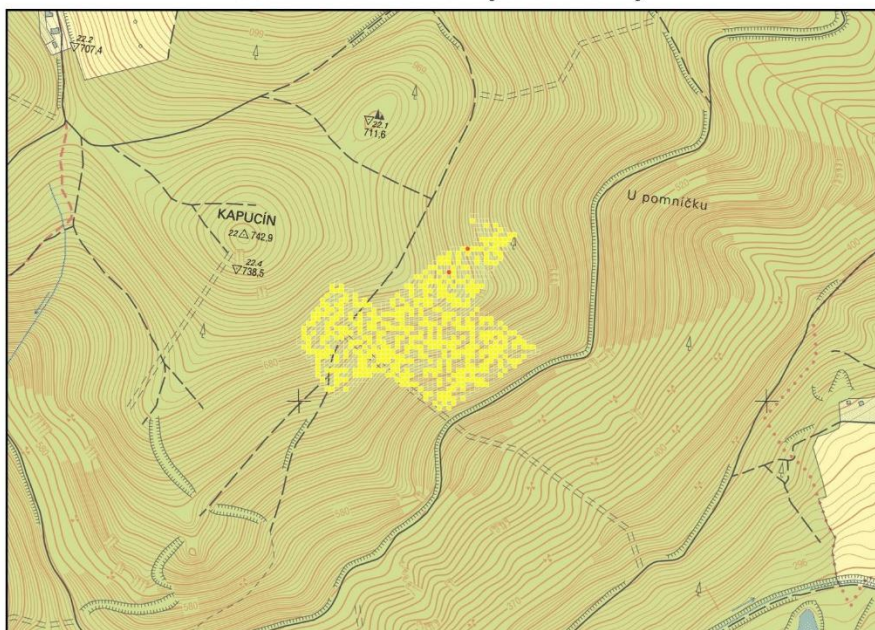
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Buk
- Základní mapa 1 : 10 000

## Lokalita č. 10 (Jiřetín 3)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

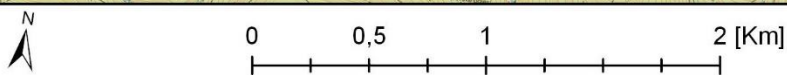
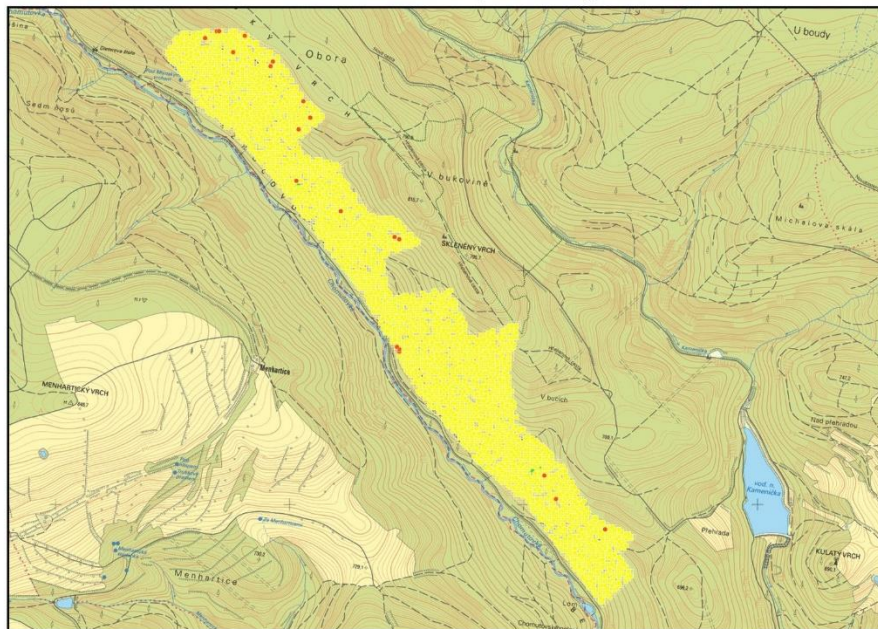
Druh lokality:

- Buk
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 11 (Bečov)



### Legenda:

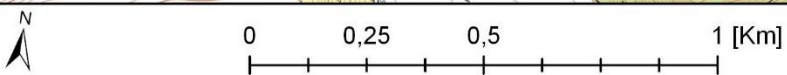
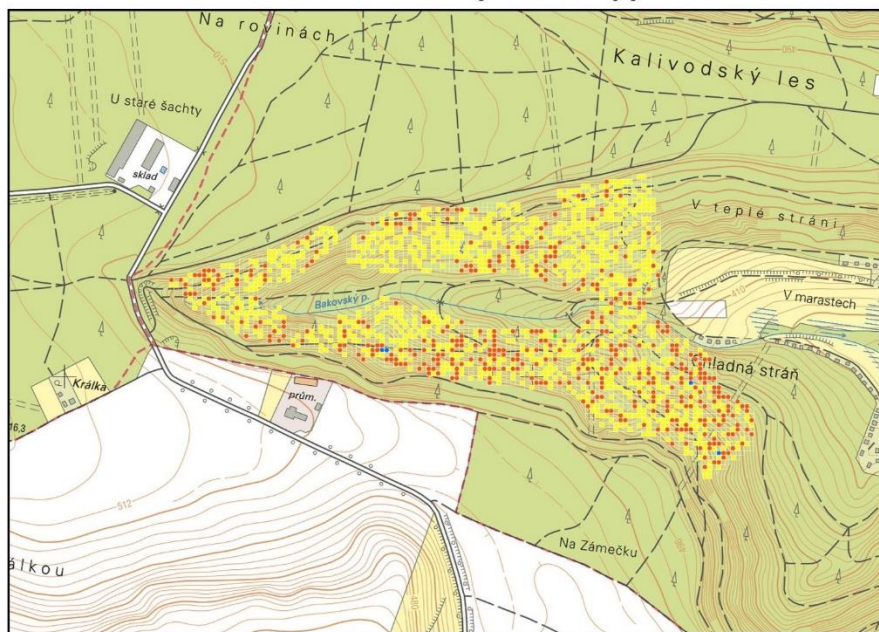
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Buk
- Základní mapa 1 : 10 000

## Lokalita č. 12 (Kalivody)



### Legenda:

Výsledok klasifikácie:

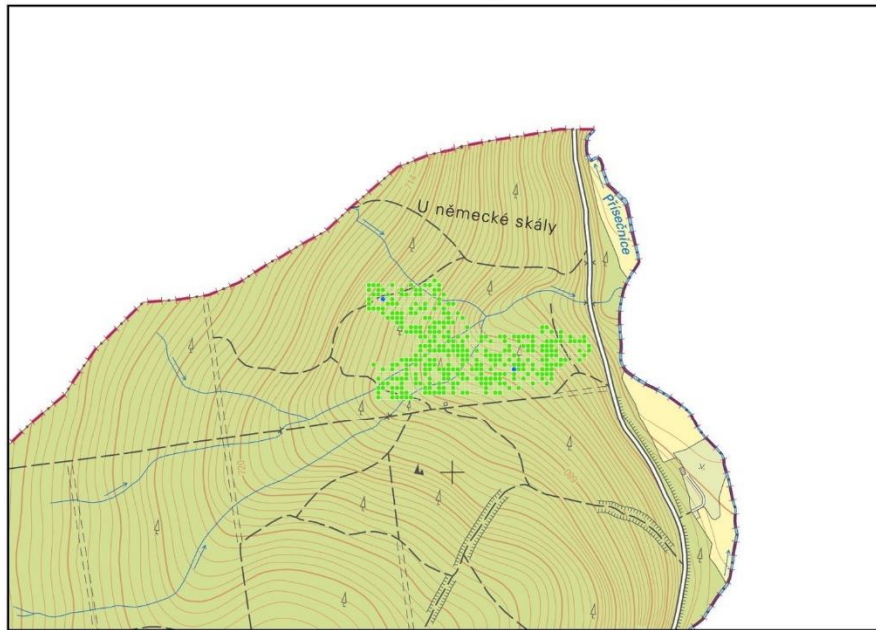
- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Buk
- Základní mapa 1 : 10 000

# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 13 (Přísečnice)



### Legenda:

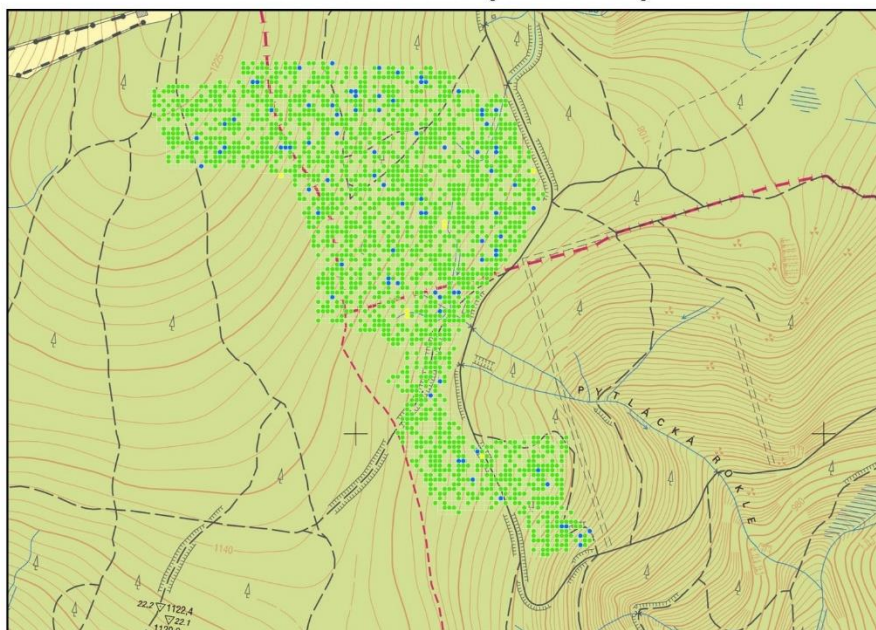
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Smrek
- Základní mapa 1 : 10 000

## Lokalita č. 14 (Klinovec)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

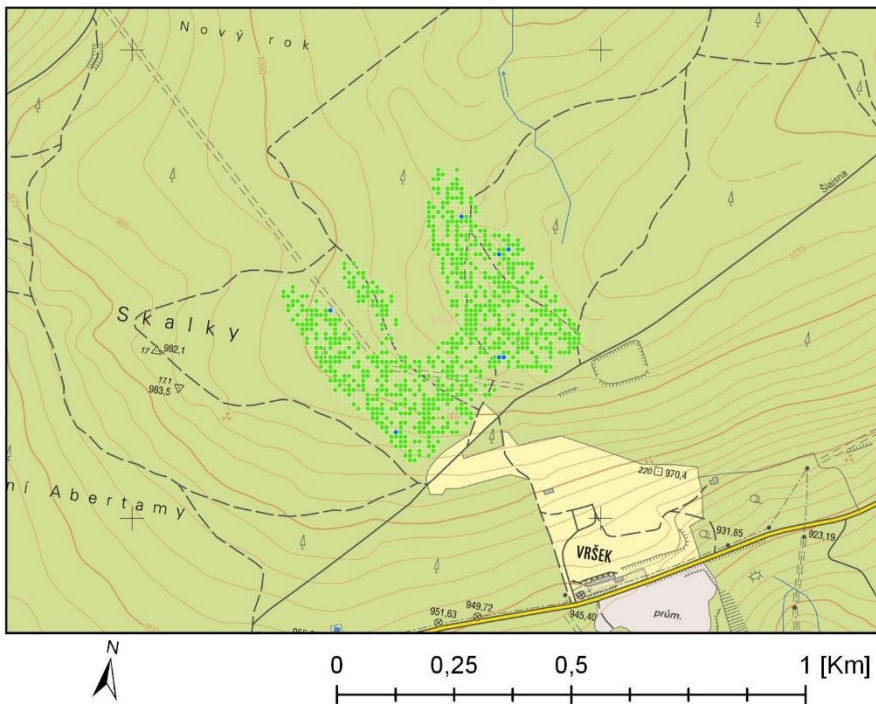
Druh lokality:

- Smrek
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 15 (Vršek)



### Legenda:

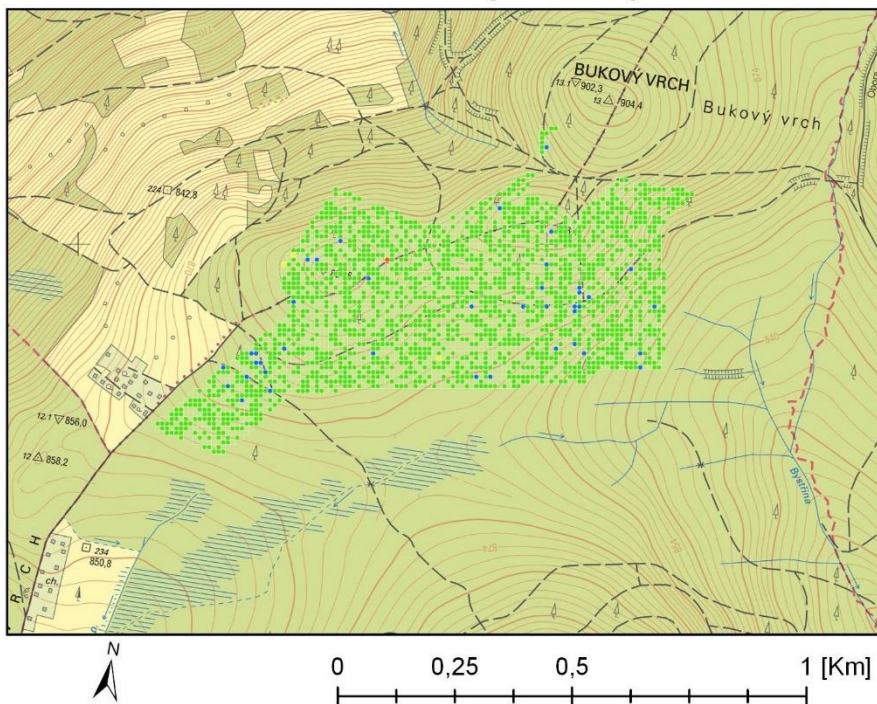
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Smrek
- Základní mapa 1 : 10 000

## Lokalita č. 16 (Kraslice)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

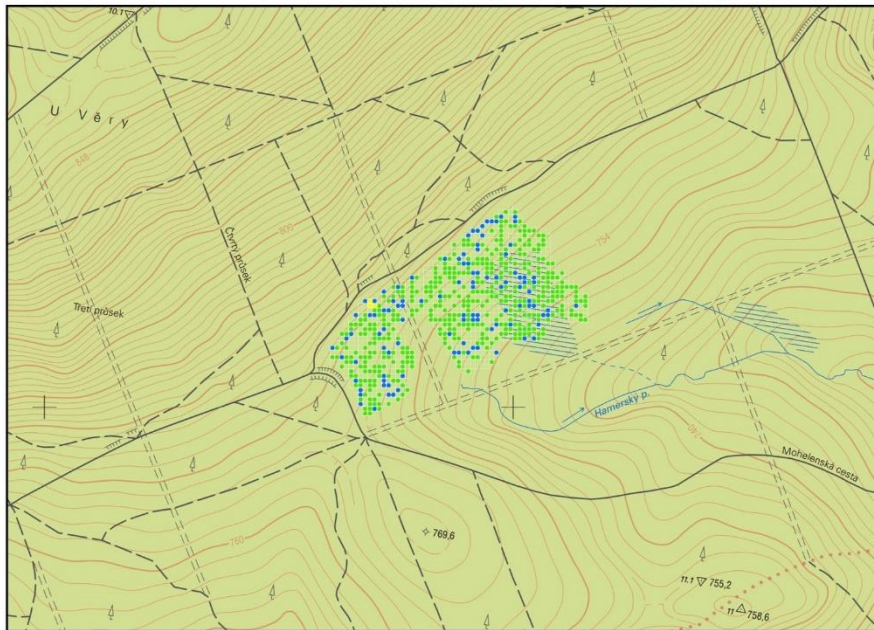
Druh lokality:

- Smrek
- Základní mapa 1 : 10 000



# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 17 (Stará Voda 2)



### Legenda:

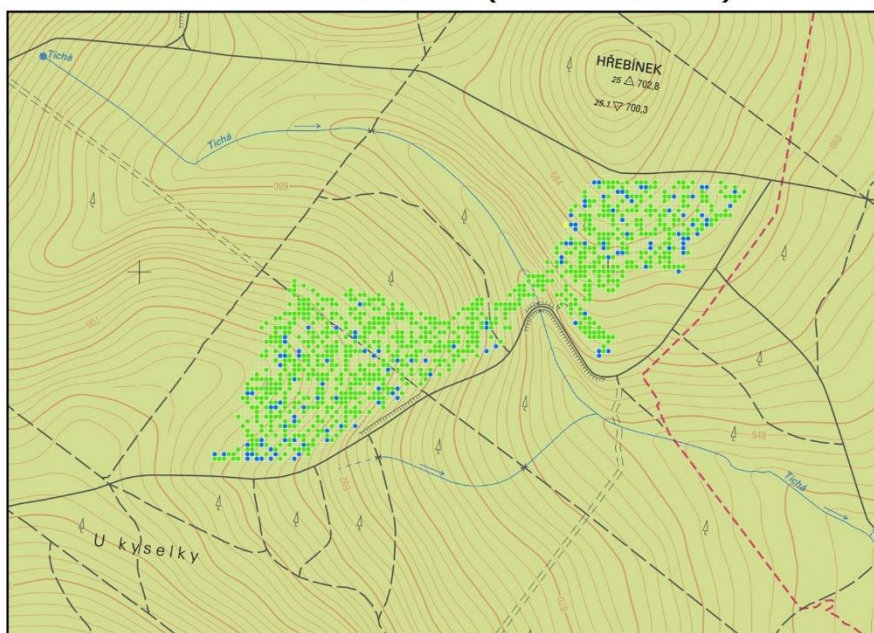
Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

Druh lokality:

- Smrek
- Základní mapa 1 : 10 000

## Lokalita č. 18 (Stará Voda 2)



### Legenda:

Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

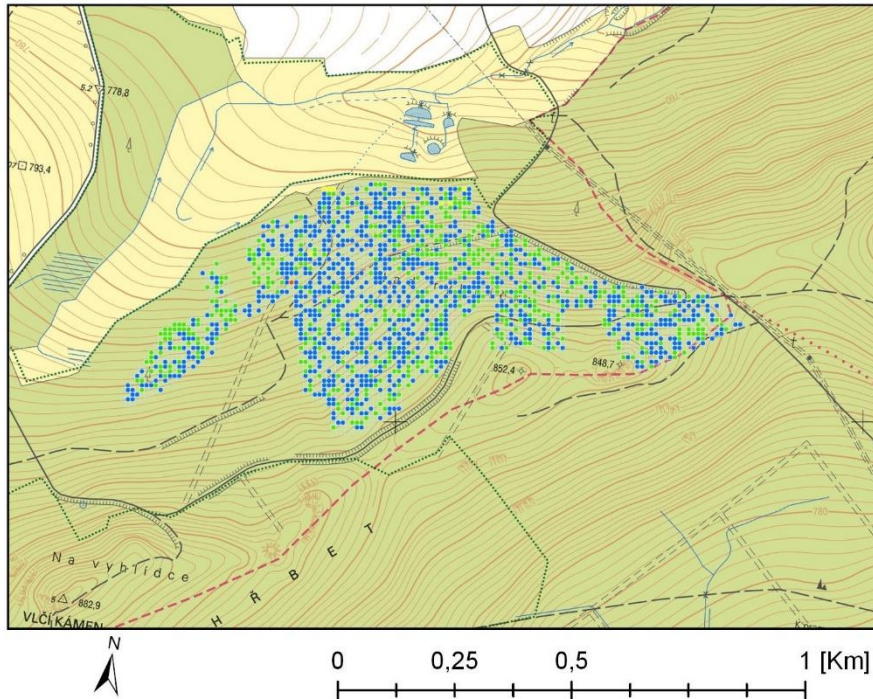
Druh lokality:

- Smrek
- Základní mapa 1 : 10 000

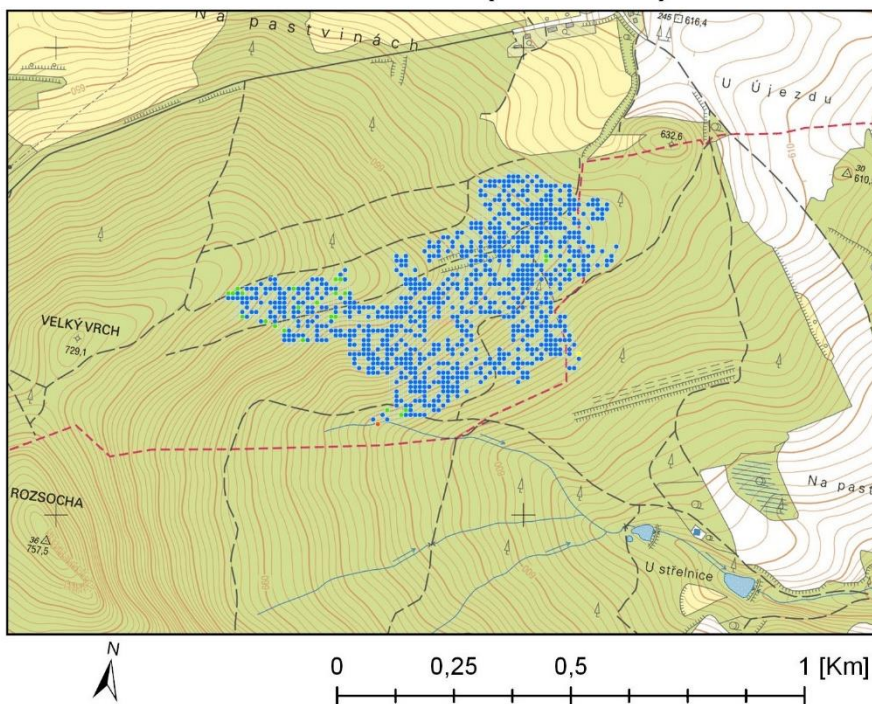


# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 19 (Prameny)



## Lokalita č. 20 (Studánka)



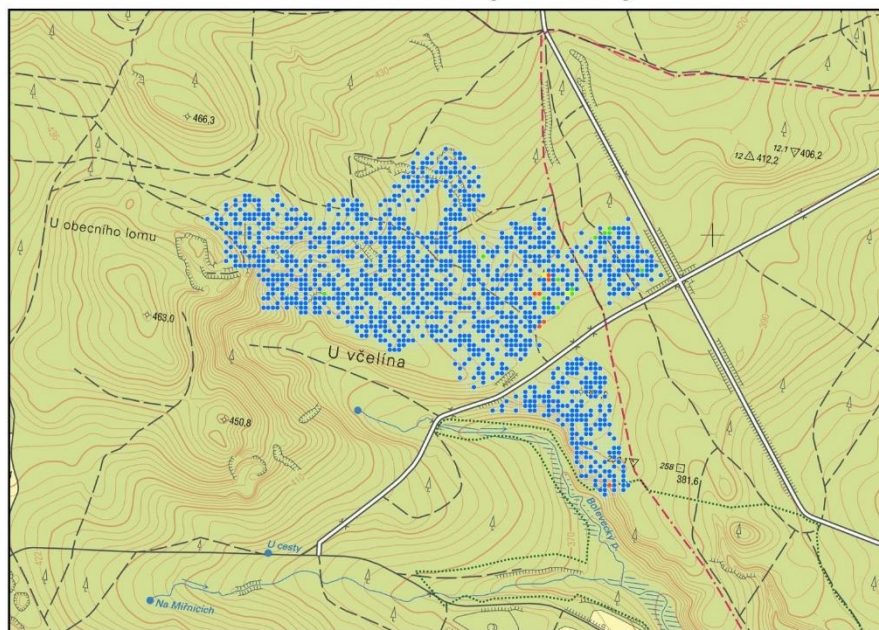






# Klasifikácia lesných porastov z dát Sentinel-2 z 4.7.2015

## Lokalita č. 23 (Plzeň 2)



### Legenda:

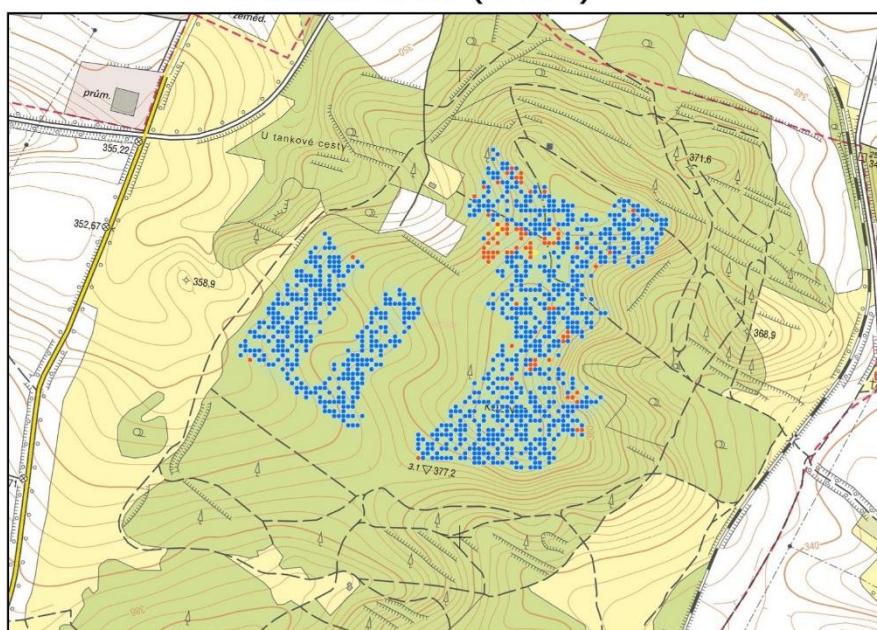
#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Borovica
- Základní mapa 1 : 10 000

## Lokalita č. 24 (Valov)



### Legenda:

#### Výsledok klasifikácie:

- Predikcia Dub
- Predikcia Buk
- Predikcia Smrek
- Predikcia Borovica

#### Druh lokality:

- Borovica
- Základní mapa 1 : 10 000

## PRÍLOHA 4: ZDROJOVÉ KÓDY PROGRAMOV

### 4.1. Program – Overenie správnosti nastavenie parametra n\_estimators

(názov súboru v elektronickej prílohe – „vypocet vplyvu estimatora na presnost.py“)

```
#
# určenie vplyvu estimators na výpočet presnosti
#

from sklearn.ensemble import RandomForestClassifier
import numpy as NUM
import arcpy as ARCPY
import arcpy.da as DA
import pandas as PD
import seaborn as SEA
import matplotlib.pyplot as PLOT
import arcgisscripting as ARC
import SSUtilities as UTILS
import os as OS

# definoveni počiatkových premenných
inputFC = r'Vrstva_body'
predictVars = [ 'blue', 'red', 'NIR', 'green', ]
classVar = ['Trieda'] # [trieda_stromu]
allVars = predictVars + classVar #mená všetkých atribútov

#pole tréningových bodov zo všetkými atribútmi
trainFC = DA.FeatureClassToNumPyArray(inputFC, ["SHAPE@XY"] + allVars)
spatRef = ARCPY.Describe(inputFC).spatialReference

#namerané údaje
data = PD.DataFrame(trainFC, columns = allVars)

#vytvorenie tréningovej a testovacej množiny (tréningová je 50% z celkovej)
fracNum = 0.5
train_set = data.sample(frac = fracNum) # 50% náhodných z data
test_set = data.drop(train_set.index) # zvyšných 50%

# Encode input values as an enumerated type or categorical variable
indicator, labels = PD.factorize(train_set[classVar[0]]) # classVar[0] je 'Trieda'

# cyklus v ktorom sa mení parameter estimators pri rovnakej množine tréningových
# a testovacích bodov
estimators = 15
for i in range(1,9):
    estimators *= 2
    rfco = RandomForestClassifier(n_estimators = estimators, oob_score = True)
    rfco.fit(train_set[predictVars], indicator)
    treesPred = rfco.predict(test_set[predictVars])
    #preklad druhov stromov
    x = []
    for kod in treesPred:
        x.append(labels[kod])
    treesPredLables = NUM.array(x, dtype=NUM.float64)

    test_trees = test_set[classVar].as_matrix()
    test_trees = test_trees.flatten()

#výpočet celkovej presnosti
treesPrd = 0
for i in range(len(treesPredLables)):
```

```
if (test_trees[i] == treesPredLables[i]): treesPrd += 1
error = (NUM.abs(len(treesPredLables) - treesPrd)/len(treesPredLables))*100
print('estimators = {0}    Presnost = {1} % '.format(estimators, (100 - error)))
```

## 4.2. Program – Overenie veľkosti trérovacej množiny na výslednú presnosť

(názov súboru v elektronickej prílohe – „vypocet vplyvu velkosti trenovacej mnoziny.py“)

```
#
# určenie vplyvu veľkosti trérovacej množiny na presnosť
#
from sklearn.ensemble import RandomForestClassifier
import numpy as NUM
import arcpy as ARCPY
import arcpy.da as DA
import pandas as PD

# definoveni počiatocných premenných
inputFC = r'Vrstva_body'
#predictVars = ['gridcode', 'gridcode_1', 'gridcode_2', 'gridcode_3']
predictVars = [ 'blue', 'red', 'NIR', 'green',]
classVar = ['Trieda'] # [trieda_stromu]
allVars = predictVars + classVar #mená všetkých atribútov

#pole trérovacích bodov zo všetkými atribútmi
trainFC = DA.FeatureClassToNumPyArray(inputFC, ["SHAPE@XY"] + allVars)
spatRef = ARCPY.Describe(inputFC).spatialReference

#namerané údaje
data = PD.DataFrame(trainFC, columns = allVars)

# cyklus v ktorom sa mení parameter estimators pri rovnakej množine trérovacích a
# testovacích bodov
for i in range(1,11):
    #vytvorenie trérovacej a testovacej množiny
    fracNum = 0.05 * i
    train_set = data.sample(frac = fracNum) # fracNum% náhodných z data
    test_set = data.drop(train_set.index) # zvyšné dáta

    # Encode input values as an enumerated type or categorical variable
    indicator, labels = PD.factorize(train_set[classVar[0]]) # classVar[0] je
    vlastne 'Trieda'
    # vytvoríme objekt RandomForest
    rfco = RandomForestClassifier(n_estimators = 500, oob_score = True)
    #priradíme trérovaciu množinu
    rfco.fit(train_set[predictVars], indicator)
    treesPred = rfco.predict(test_set[predictVars])
    #preklad druhov stromov
    x = []
    for kod in treesPred:
        x.append(labels[kod])
    treesPredLables = NUM.array(x, dtype=NUM.float64)

    test_trees = test_set[classVar].as_matrix()
    test_trees = test_trees.flatten()

    #výpočet celkovej presnosti
    treesPrd = 0
    for i in range(len(treesPredLables)):
        if (test_trees[i] == treesPredLables[i]): treesPrd += 1
    error = (NUM.abs(len(treesPredLables) - treesPrd)/len(treesPredLables))*100
    print('trérovacia množina = {0} % (počet bodov = {1}) Presnosť = {2} %
    '.format(fracNum*100, train_set.shape[0], (100 - error)))
```

### 4.3. Program – použitý na výpočet

(názov súboru v elektronickej prílohe – „program kvalifikovaný výber.py“)

```
from sklearn.ensemble import RandomForestClassifier
import numpy as NUM
import arcpy as ARCPY
import arcpy.da as DA
import pandas as PD
import seaborn as SEA
import matplotlib.pyplot as PLOT
#import arcgisscripting as ARC
#import SSUtilities as UTILS
import os as OS

# definovanie počiatočných premenných
inputFC = r'Vrstva_body'
#predictVars = ['gridcode_1', 'gridcode_2', 'gridcode_3']
predictVars = [ 'blue', 'red', 'NIR', 'green',]
classVar = ['Trieda'] # [trieda_stromu]
allVars = predictVars + classVar #mená všetkých atribútov

#pole tréningových bodov zo všetkými atribútmi
trainFC = DA.FeatureClassToNumPyArray(inputFC, ["SHAPE@XY"] + allVars)
spatRef = ARCPY.Describe(inputFC).spatialReference

#namerané údaje
data = PD.DataFrame(trainFC, columns = allVars)

#zobrazenie prvých x riadkov v tabuľke
data.head()

#vykreslenie korelačnej tabuľky
corr = data.astype('float64').corr()
ax = SEA.heatmap(corr, cmap=SEA.diverging_palette(220, 10, as_cmap=True),
square=True, annot = True, linecolor = 'k', linewidths = 1)
PLOT.show()

#vytvorenie tréningovej a testovacej množiny (tréningová je 50% z celkovej)
fracNum = 0.5
train_set = data.sample(frac = fracNum) # 50% náhodných z data
test_set = data.drop(train_set.index) # zvyšných 50%
print('Veľkosť množiny: tréningovej = {0}; testovacej =
{1}'.format(train_set.shape[0], test_set.shape[0]))

# Prekódovanie vstupných hodnôt ako vymenovaný typ alebo kategorizovaná premenná
indicator, labels = PD.factorize(train_set[classVar[0]]) # classVar[0] je
vlastne 'Trieda'
# vytvoríme objekt RandomForest
#rfco = RandomForestClassifier(n_estimators = 500, min_samples_split = 40,
min_samples_leaf = 20, oob_score = True)
rfco = RandomForestClassifier(n_estimators = 500, oob_score = True, criterion =
'entropy')
#priradenie tréningovej množiny
#fit(X, y[, sample_weight])Build a forest of trees from the training set (X, y).
rfco.fit(train_set[predictVars], indicator)
#a spolu s testovacov množinou sa urobí predpoklad
treesPred = rfco.predict(test_set[predictVars])

x = []
for kod in treesPred:
    x.append(labels[kod])
treesPredLables = NUM.array(x, dtype=NUM.float64)
```



```

# Vytvorenie confusion matrix
PD.crosstab(test_set[classVar[0]], treesPredLables, rownames=['Actual Species'],
colnames=['Predicted Species'])
# Vytvorenie confusion matrix v %
PD.crosstab(test_set[classVar[0]], treesPredLables, rownames=['Druhy skutočné [%]'],
colnames=['Druhy predpovedané']).apply(lambda r: (r/r.sum())*100, axis=1)

# Vytvorenie poľa z matice testovacích údajov
test_trees = test_set[classVar].as_matrix()
test_trees = test_trees.flatten()
#print('test_trees = ', len(test_trees))
#print('treesPred = ', len(treesPred))

# Vymenovanie počtu skutočných a predpovedaných druhov stromov
druhStromov = ['nil', 'Dub', 'Buk', 'Smrek', 'Borovica']
for i in range(1,5):
    print('{0} Výskyt = {1} Predikcia = {2}'.format(druhStromov[i],
len(NUM.where(test_trees==i)[0]), len(NUM.where(treesPredLables==i)[0])))

# výpočet celkovej presnosti
treesPrd = 0
for i in range(len(treesPredLables)):
    if (test_trees[i] == treesPredLables[i]): treesPrd += 1
error = (NUM.abs(len(treesPredLables) - treesPrd)/len(treesPredLables))*100
print('trénovacia množina = {0} % (počet bodov = {1}) Presnosť = {2} %
'.format(fracNum*100, train_set.shape[0], (100 - error)))

# Zobrazenie zoznamu atribútov a ich dôležitosť na vypočítanie odhadu
list(zip(train_set[predictVars], rfco.feature_importances_))

# Zobrazíť predpokladané pravdepodobnosti prvých 10 bodov
rfco.predict_proba(test_set[predictVars])[0:10]

# zostavenie cesty a názvu projektu
projectName = 'L8_L2_170830'
outputDir = r'D:\gis\{0}\{0}.gdb'.format(projectName)
print(outputDir)
#outputDir = r'D:\gis\L8_L1_160624\L8_L1_160624.gdb'

# vytvorenie vrstiev z predikcie podľa druhu stromov
for i in range(1,5):
    nameFC = 'Predpoved{0}_{1}'.format(i, druhStromov[i])
    print(nameFC, ' - ', outputDir)
    stromy = trainFC[["SHAPE@XY"]][test_set.index[NUM.where(treesPredLables==i)]]
    ARCPY.da.NumPyArrayToFeatureClass(stromy, OS.path.join(outputDir, nameFC),
['SHAPE@XY'], spatRef)

# vytvorenie vrstvy tréningových bodov
nameFC = 'TreningVrstva'
stromy = trainFC[["SHAPE@XY"]][train_set.index]
ARCPY.da.NumPyArrayToFeatureClass(stromy, OS.path.join(outputDir, nameFC),
['SHAPE@XY'], spatRef)

```