

UNIVERZITA KARLOVA

Přírodovědecká fakulta

Katedra aplikované geoinformatiky a kartografie

Studijní program: Geografie (magisterské studium)

Studijní obor: Kartografie a geoinformatika



Bc. Daniela VALCHÁŘOVÁ

**KLASIFIKACE KRAJINNÉHO POKRYVU VE VYBRANÝCH
ÚZEMÍCH ETIOPIE POMOCÍ KLASIFIKÁTORU
STROJOVÉHO UČENÍ**

**LANDCOVER CLASSIFICATION OF SELECTED PARTS OF
ETHIOPIA BASED ON MACHINE LEARNING METHOD**

Diplomová práce

Vedoucí práce: doc. RNDr. Přemysl Štych, Ph.D.

Praha 2021

Vysoká škola: Univerzita Karlova

Fakulta: Přírodovědecká

Katedra: Aplikované geoinformatiky a kartografie

Školní rok: 2020/21

Zadání diplomové práce

pro Danielu Valchářovou

obor Kartografie a geoinformatika

Název tématu: Klasifikace krajinného pokryvu ve vybraných územích Etiopie pomocí klasifikátoru strojového učení

Zásady pro vypracování

Téma a zadání diplomové práce se odvíjí od řešeného projektu České rozvojové agentury „Zajištění udržitelného hospodaření v krajině ve vybraných oblastech Etiopie na základě geovědního mapování“. Cílem práce bude klasifikovat krajinný pokryv vybraných území v Etiopii, a to s využitím družicových snímků Sentinel-2 a PlanetScope. Bude vytvořen algoritmus klasifikace v cloudovém prostředí Google Earth Engine a testovány moderní klasifikační metody strojového učení. Z geoinformačního metodického hlediska bude cílem vyřešit maskování oblačnosti, mozaikování použitých scén, definovat přesnou legendu tříd land cover, zvolit minimální mapovací jednotku a vybrat vhodné období a spektrální pásma používaných družicových dat. Klíčovým úkolem bude výběr vhodného klasifikátoru a nalezení vhodných nastavení parametrů tak, aby bylo docíleno co nejvyšší přesnosti klasifikace (min 80 %). Sestavený algoritmus bude testován na klasifikaci krajinného pokryvu

dvou typů družicových dat (Sentinel-2 a PlanetScope) na dvou měřítkových úrovních. Výsledky klasifikace budou prezentovány pomocí map a také budou k dispozici jako geodatabáze pro další využití. Vytvořené skripty budou volně k dispozici v prostředí Google Earth Engine.

Rozsah grafických prací: cca 4 strany

Rozsah průvodní zprávy: cca 70 stran

Seznam odborné literatury:

EGGEN, M., OZDOGAN, M. ZAITCHIK, F., SIMANE, B. (2016): Land Cover Classification in Complex and Fragmented Agricultural Landscapes of the Ethiopian Highlands. Remote Sensing, 8, 12. Dostupné z: <https://www.mdpi.com/2072-4292/8/12/1020> (cit. 26. 6. 2020).

NOI, P. T., KAPPAS, M. (2017): Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, 18, 1. Dostupné z: <https://www.mdpi.com/1424-8220/18/1/18> (cit. 26. 6. 2020).

OK, A. O., AKAR, O., GUNGOR, O. (2012): Evaluation of random forest method for agricultural crop classification. European Journal of Remote Sensing, 45, 1. Dostupné z: <https://www.tandfonline.com/doi/abs/10.5721/EuJRS20124535> (cit. 28. 6. 2020).

ŠTYCH, P., JEŘÁBKOVÁ, B., LAŠTOVIČKA, J., RIEDL, M., PALUBA, D. (2019): A Comparison of WorldView-2 and Landsat 8 Images for the Classification of Forests Affected by Bark Beetle Outbreaks Using a Support Vector Machine and a Neural Network: A Case Study in the Sumava Mountains. Geosciences, 9, 9. Dostupné z: <https://www.mdpi.com/2076-3263/9/9/396> (cit. 26. 6. 2020).

Vedoucí diplomové práce: doc. RNDr. Přemysl Štych, Ph.D.

Konzultant diplomové práce: -

Datum zadání diplomové práce: 7. 1. 2020

Termín odevzdání diplomové práce: podzim 2021

Platnost tohoto zadání je po dobu jednoho akademického roku.

.....
Vedoucí diplomové práce

.....
Garant studijního programu

V Praze dne

Prohlášení

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

V Praze dne

.....

Bc. Daniela Valchářová

Poděkování

Na tomto místě bych chtěla poděkovat vedoucímu mé diplomové práce doc. RNDr. Přemyslu Štychovi, Ph.D. za téma, trpělivost, věnovaný čas a cenné rady. Dále bych chtěla poděkovat kolegům z katedry Aplikované geoinformatiky a kartografie za plodné diskuze, jmenovitě jsou to: Bc. Gabriela Pinlová, Bc. Mojmír Polák, Mgr. Daniel Paluba, Mgr. Jan Svoboda, Mgr. Josef Laštovička, Ph.D., RNDr. Jakub Lysák, Ph.D. a Ing. Miroslav Čábelka. Dále také děkuji kolegům z České geologické služby, kteří mě vzali na terénní výzkum do Etiopie, především doc. RNDr. Kryštofu Vernerovi, Ph.D., že mě tam nenechal. V neposlední řadě děkuji rodině a přátelům za podporu a pomoc v průběhu celého studia. A nejvíce svému příteli Dominiku Machovi za všechno.

Abstrakt

Diplomová práce se zabývá klasifikací krajinného pokryvu v regionu Sidama v Etiopii a 2 kebelí, Chanco a Dangora Morocho. Využívány jsou družicové snímky vysokého rozlišení Sentinel-2 a velmi vysokého rozlišení PlanetScope. Tvorba klasifikačního algoritmu probíhá v cloudovém prostředí Google Earth Engine. Testováno je 10 kombinací 4 nejdůležitějších parametrů klasifikační metody Random Forest. Definovaná legenda obsahuje 8 tříd land cover, a to zástavba, plodiny, louka/pastvina, les, křoviny, holá půda, mokřad a vodní plocha. Trénovací dataset je sbírán v terénu na podzim roku 2020. Porovnávány jsou výsledky klasifikace dvou typů dat na dvou měřítkových úrovních. Nejvyšší celková přesnost pro klasifikaci land cover regionu Sidama vyšla 84,1 % a kappa index 0,797, a to s parametry metody Random Forest 100 stromů, 4 spektrální pásma vstupující do každého stromu, hodnota 1 pro populaci listu a 40 % trénovacích dat využitých pro každý strom. Pro klasifikaci land cover kebele Chanco a Dangora Morocho se stejným nastavením metody vyšla celková přesnost 66,00 a 73,73 % a kappa index 0,545 a 0,601. Pro klasifikaci kebele Chanco vyšla lépe jiná kombinace parametrů (80, 3, 1, 0,4), a přenositelnost stejného nastavení algoritmu se tak potvrdila pouze z 50 %. Pro klasifikaci kebelí autorka práce v diskuzi navrhuje objektovou klasifikaci či vektorizaci kebelí jako možné lepší řešení.

Klíčová slova: krajinný pokryv, Etiopie, Google Earth Engine, Sentinel-2, PlanetScope, Random Forest

Abstract

Diploma thesis deals with the land cover classification in Sidama region of Ethiopia and 2 kebeles, Chanco and Dangora Morocho. High resolution Sentinel-2 and very high resolution PlanetScope satellite images are used. The development of the classification algorithm is done in the Google Earth Engine cloud based environment. Ten combinations of the 4 most important parameters of the Random Forest classification method are tested. The defined legend contains 8 land cover classes, namely built-up, crops, grassland/pasture, forest, scrubland, bareland, wetland and water body. The training dataset is collected in the field during the fall 2020. The classification results of the two data types at two scales are compared. The highest overall accuracy for land cover classification of Sidama region came out to be 84.1% and kappa index of 0.797, with Random Forest method parameters of 100 trees, 4 spectral bands entering each tree, value of 1 for leaf population and 40% of training data used for each tree. For the land cover classification of Chanco and Dangora Morocho kebele with the same method settings, the overall accuracy came out to be 66.00 and 73.73% and kappa index of 0.545 and 0.601. For the classification of Chanco kebele, a different combination of parameters (80, 3, 1, 0.4) worked out better and thus the transferability of the same algorithm setting was only 50% confirmed. For kebele classification, the author of this thesis in the discussion suggests object-based classification or vectorization of kebeles as a possible better solution.

Keywords: land cover, Ethiopia, Google Earth Engine, Sentinel-2, PlanetScope, Random Forest

Obsah

Seznam obrázků a tabulek	9
Seznam zkratk	11
1 Úvod.....	12
2 Úvod do problematiky.....	13
2.1 Klasifikace krajinného pokryvu.....	13
2.2 Metody strojového učení	14
2.3 Klasifikace krajinného pokryvu v Etiopii	17
2.4 Klasifikace krajinného pokryvu v cloudovém prostředí	21
3 Data a metodika.....	22
3.1 Zájmové území.....	22
3.1.1 Fragmentovaná krajina regionu Sidama.....	24
3.2 Data	28
3.2.1 Sentinel-2.....	28
3.2.2 PlanetScope	29
3.2.3 Další podpůrná data	31
3.3 Metodika	32
3.3.1 Legenda a mapovací jednotka.....	33
3.3.2 Terénní výzkum a sběr trénovacích dat	35
3.3.3 Předzpracování snímků.....	39
3.3.4 Spektrální charakteristiky	42
3.3.5 Volba parametrů klasifikační metody	44
3.3.6 Filtrace obrazu	46
3.3.7 Vizualizace	48
3.3.8 Hodnocení přesnosti a kontrolní množina dat.....	49
3.3.9 Klasifikace land cover s využitím dat PlanetScope ve vybraných kebelích	51
4 Výsledky	56
4.1 Výsledky klasifikace dat Sentinel-2.....	56
4.2 Výsledky klasifikace dat PlanetScope	61
4.3 Porovnání výsledků klasifikace dat Sentinel-2 a PlanetScope	67
5 Diskuze.....	72
6 Závěr	78
7 Použitá literatura a data	80
Seznam příloh	85

Seznam obrázků a tabulek

Obr. 1 – Jednoduchý příklad rozhodovacího stromu	16
Obr. 2 – Prostředí psaní kódu v Google Earth Engine.....	22
Obr. 3 – Zájmové území.....	23
Obr. 4 – Stát Sidama s významnými městy	25
Obr. 5 – Různorodá krajina Sidamy, zleva teplá, horká a horská oblast.....	25
Obr. 6 – Typická krajinná struktura východní části regionu Sidama	26
Obr. 7 – Ukázka mokřadu, vlevo snímek z února 2021, vpravo z listopadu 2020.....	27
Obr. 8 – Vybrané kebele regionu Sidama.....	28
Obr. 9 – Pokrytí regionu Sidama dostupnými snímky PlanetScope 18. 11. 2020	31
Obr. 10 – Schéma metodiky	33
Obr. 11 – Terénní výzkum v regionu Sidama na podzim 2020	36
Obr. 12 – Rozmístění dat z terénu v regionu Sidama	37
Obr. 13 – Vybrané snímky Sentinel-2 ze dne 19. 11. 2020	39
Obr. 14 a 15 – Porovnání původního obrazu a masek oblačnosti bez zařazení hodnoty 7 pásma SCL a s jejím zařazením (červená místa na obrázcích jsou ta odmaskovaná)	41
Obr. 16 – Průměrná spektrální odrazivost trénovacích dat všech tříd land cover v datech Sentinel-2 ..	43
Obr. 17 – Spektrální příznaky: Odlehlé hodnoty trénovacích dat v třídě vodní plochy	43
Obr. 18 – Vztah odrazivosti pásem NIR a Red trénovacích dat všech tříd land cover.....	44
Obr. 19 – Filtrace obrazu.....	48
Obr. 20 – Legenda.....	49
Obr. 21 – Rozmístění trénovacích bodů pro klasifikaci kebelí	53
Obr. 22 – Spektrální příznaky: Odlehlé hodnoty trénovacích dat pro klasifikaci kebelí v třídě vodní plochy	54
Obr. 23 – Vztah odrazivosti pásem NIR a Red trénovacích dat všech tříd land cover pro klasifikaci kebelí.....	55
Obr. 24 – Srovnání výřezu výsledné mapy krajinného pokryvu a snímku Sentinel-2	58
Obr. 25 – Zastoupení jednotlivých tříd land cover v regionu Sidama v roce 2020 [%]	59
Obr. 26 – Výsledná mapa krajinného pokryvu regionu Sidama v roce 2020.....	60
Obr. 27 – Příklad oblasti stínů mraků klasifikovaných jako třída zástavba a mokřad v porovnání se snímkem Sentinel-2.....	60
Obr. 28 - Zastoupení jednotlivých tříd land cover v kebeli Chanco a Dangora Morocho v roce 2020 [%]	64
Obr. 29 – Výsledné mapy krajinného pokryvu kebelí Chanco a Dangora Morocho v roce 2020	65
Obr. 30 – Srovnání výřezu výsledné mapy krajinného pokryvu a snímku PlanetScope	66
Obr. 31 – Srovnání výřezu výsledné mapy krajinného pokryvu a snímku PlanetScope II.....	66

Obr. 32 – Porovnání zastoupení jednotlivých tříd land cover v kebelích Chancho a Dangora Morocho klasifikací nad daty Sentinel-2 a PlanetScope [%]	67
Obr. 33 – Výsledné mapy krajinného pokryvu kebele Chancho s využitím dat Sentinel-2, PlanetScope ve srovnání se snímkem velmi vysokého rozlišení	69
Obr. 34 - Výsledné mapy krajinného pokryvu kebele Dangora Morocho s využitím dat Sentinel-2, PlanetScope ve srovnání se snímkem velmi vysokého rozlišení	71
Tab. 1 – Přehled použitých legend ve vybraných vědeckých studiích	20
Tab. 2 – Přehled použitých legend ve vybraných databázích LU/LC afrického kontinentu	20
Tab. 3 – Využitá pásma datasetu Sentinel-2 MSI Level-2A.....	29
Tab. 4 – Vybrané snímky Sentinel-2	29
Tab. 5 – Pásma datasetu PlanetScope.....	30
Tab. 6 – Vybrané snímky PlanetScope.....	30
Tab. 7 – Ukázka části atributové tabulky dat z terénu	37
Tab. 8 – Počet trénovacích bodů jednotlivých klasifikačních tříd	38
Tab. 9 – Hodnoty pásma SCL dat Sentinel-2	40
Tab. 10 – Parametry metody Random Forest v prostředí GEE.....	45
Tab. 11 – Testované kombinace parametrů metody Random Forest	46
Tab. 12 – Počet vygenerovaných kontrolních bodů.....	50
Tab. 13 – Počet trénovacích dat pro klasifikaci kebelí	52
Tab. 14 – Počet vygenerovaných kontrolních bodů pro klasifikaci kebelí.....	56
Tab. 15 – Výsledná celková přesnost a kappa index klasifikace regionu Sidama pro 10 kombinací parametrů metody Random Forest	57
Tab. 16 – Chybová matice klasifikace krajinného pokryvu regionu Sidama, kombinace 5	58
Tab. 17 - Výsledná celková přesnost a kappa index klasifikace kebelí pro 10 kombinací parametrů metody Random Forest	62
Tab. 18 - Chybová matice klasifikace kebele Chancho, kombinace 1	63
Tab. 19 – Chybová matice klasifikace kebele Dangora Morocho, kombinace 5	63

Seznam zkratek

ČRA	Česká rozvojová agentura
FAO	Food and Agriculture Organization of the United Nations
GEE	Google Earth Engine
GEP	Google Earth Pro
LC	Land Cover
LU	Land Use
MLC	Maximum Likelihood Classification
MMU	Minimální mapovací jednotka
NIR	Near-infrared spektrální pásmo
NP	Národní park
RF	Random Forest
RGB	Red, Green, Blue spektrální pásma
SHP	Shapefile
SVM	Support Vector Machine
SWIR	Shortwave-infrared spektrální pásmo

1 Úvod

Diplomová práce se zaměřuje na klasifikaci krajinného pokryvu (land cover) vybraných oblastí Etiopie, a to pomocí družicových snímků Sentinel-2 a PlanetScope, v cloudovém prostředí Google Earth Engine (GEE) s využitím moderních klasifikačních metod strojového učení. Důkladně je řešeno předzpracování snímků, volba vhodné legendy a nastavení parametrů klasifikační metody tak, aby bylo dosaženo co nejvyšší přesnosti.

Práce se odvíjí od projektu České rozvojové agentury (ČRA) „*Zajištění udržitelného hospodaření v krajině ve vybraných oblastech Etiopie na základě geovědního mapování*“, jehož cílem je detailně zmapovat zájmová území v Etiopii, vytvořit na základě podrobné analýzy krajiny a terénního šetření mapové výstupy a dále poskytnout získané informace místnímu obyvatelstvu. Tyto výstupy (mapy využití krajiny, krajinné plány, databáze dat, analýzy s časovou řadou) by měly pomoci k lepšímu a udržitelnému rozvoji daného území Etiopie, a to především v oblasti hospodaření s přírodními zdroji (pitná voda, půda, nerostné suroviny). Na projektu spolupracuje několik organizací – Česká geologická služba, Etiopská geologická služba, All4Soil a Přírodovědecká fakulta Univerzity Karlovy. Jedná se tak o velice komplexní spolupráci odborníků mnoha oborů. Důraz je kladen na spolupráci s místní samosprávou na úrovni kebelí (malé správní jednotky). Jedná se tak o participativní a kolaborativní mapování a tvorbu plánů využití přírodních zdrojů.

Z metodického hlediska je cílem diplomové práce vyřešení několika zásadních úkolů, které vedou k co nejvíce uživatelsky a výzkumně vhodné a přesné klasifikaci krajinného pokryvu (min 80 %) exotické lokality regionu Sidama v Etiopii v roce 2020, a to nad snímky Sentinel-2 a PlanetScope v cloudovém prostředí za pomoci metod strojového učení. Dílčími cíli je volba vhodné metody maskování oblačnosti a mozaikování, jelikož se jedná o rozsáhlejší území. Dalším úkolem je definice legendy a minimální mapovací jednotky pro klasifikaci na základě terénního průzkumu, používaných dat a požadavků koncového uživatele. S tím se pojí také výběr vhodného období a spektrálních pásem používaných družicových dat. Důležitým cílem je vyvinout klasifikační algoritmus ve volně dostupné platformě GEE a tímto způsobem zprostředkovat výsledky práce široké uživatelské skupině. V tomto ohledu bylo nutné seznámit se s vývojářskou cloudovou platformou GEE, zjistit nabídku a detailně proniknout do podstaty klasifikátorů (jejich parametrů) a dále naprogramovat/modifikovat a otestovat klasifikátor pro vytyčené účely této studie. Klíčovým

úkolem je tedy výběr vhodného klasifikátoru a nalezení vhodných nastavení parametrů tak, aby bylo docíleno co nejvyšší celkové přesnosti klasifikace (min 80 %). Pro účely klasifikace a vyhodnocení přesnosti bylo nezbytné sesbírat kvalitní trénovací a kontrolní množiny. Pro tyto účely proběhl výše zmíněný terénní průzkum v řešené oblasti. Výsledky klasifikace jsou k dispozici jako geodatabáze a dále jsou prezentovány pomocí mapy regionu Sidama v měřítku 1 : 100 000 dle zadání projektu ČRA. Pro tvorbu map krajinného pokryvu vybraných kebelí v měřítku 1 : 10 000 byla využita data s velmi vysokým rozlišením PlanetScope pro klasifikaci lokálního území, tzv. kebelí. Byla tak testována vhodnost a přesnost vytvořeného algoritmu pro klasifikaci dvou typů družicových dat (Sentinel-2 a PlanetScope) na dvou měřítkových úrovních.

2 Úvod do problematiky

2.1 Klasifikace krajinného pokryvu

Klasifikace obrazu je proces, při kterém je datový soubor rozdělen do určitých tříd podle určitého příznaku na základě určitých rozhodovacích pravidel, tzv. klasifikátoru (Červená 2016a). Třídy jsou voleny na základě daného území a účelu klasifikace. Objekty jsou do tříd přiřazovány podle příznaku, který je na objektech naměřen. Základní je spektrální, prostorový, časový a polarizační příznak (Kolář 1990).

Klasifikaci dělíme na řízenou a neřízenou, přičemž hlavní rozdíl spočívá v tom, že u řízené klasifikace potřebujeme předem trénovací data a nadefinovanou legendu (Červená 2016b). U neřízené klasifikace není potřeba předem znát počet tříd a shluků, které jsou během klasifikace vytvořeny, je význam přiřazen až později. Při klasifikaci krajinného pokryvu (LC) nebo využití krajiny (LU) jsou jednotlivým plochám na zemi přiřazovány třídy LC/LU, kdy LC představuje fyzický krajinný pokryv (lesy, vodní plochy, pole apod.) a LU udává, jakým způsobem je území lidmi využíváno (lesnictví, zemědělství, vodní hospodářství apod.) (Mendelu 2020). Pro klasifikaci krajinného pokryvu jsou využívány různé klasifikátory založené na statistice nebo na bázi umělé inteligence, např. Minimum Distance (minimální vzdálenost), Maximum Likelihood (maximální pravděpodobnost), Support Vector Machine (podpůrné vektory), Random Forest (náhodný les) a další.

2.2 Metody strojového učení

Strojové učení patří mezi umělou inteligenci a používaný systém se během procesu učí na základě dat, se kterými pracuje. Umělá inteligence je široký pojem označující systémy, jež napodobují lidskou inteligenci (Oracle 2020). Systém strojového učení se přizpůsobuje na základě prostředí a prolíná se se statistikou a data mining. Příkladem strojového učení je rozpoznávání řeči nebo psaného textu, predikce obchodního úspěchu, označování fotografií na sociálních sítích, doporučování relevantních reklam a další. Mezi modely strojové učení patří např. rozhodovací stromy, algoritmus k-nejbližších sousedů, neuronové sítě a podpůrné vektory.

Pro kvalitní a přesnou klasifikaci je důležitá nejen volba dat, ale též volba správné metody. Noi, Kappas (2017) porovnávali několik metod strojového učení, a to Random Forest (RF), k-Nearest Neighbor (k-NN) a Support Vector Machine (SVM) na příkladu oblasti Red River Delta ve Vietnamu. Pro dosažení maximální celkové přesnosti byla testována různá nastavení parametrů jednotlivých metod, přičemž u SVM metody je volena jádrová funkce (kernel function), váha (cost) a šířka (weight) a u RF metody počet stromů (ntree) a počet proměnných v každém stromě (mtry). U metody RF bylo potvrzeno, že velký počet stromů přinese dobré výsledky. Stejně vyhovující vyšlo nastavení ntree = 100, 200, 500 a 1000 a mtry = 2, 3 a 4. Autoři článku však naznačují, že pro každou zkoumanou oblast a dataset se mohou nastavení lišit.

Součástí studie je také analýza četnosti využívání různých klasifikačních metod ve vědeckých článcích v poslední dekádě. Z analýzy vyplývá, že metody strojového učení si v poslední době získaly velkou pozornost. Dále lze pozorovat vysokou četnost využití metody Maximum Likelihood (MLC). To je však dáno tím, že metody strojového učení jsou často porovnávány s MLC jako pokročilou parametrickou metodou. Autoři také vysvětlují, proč právě SVM a RF patří mezi nejvhodnější metody pro klasifikaci obrazu. Noi, Kappas (2017, s. 2) tvrdí, že podle výsledků ostatních studií „*metoda neuronových sítí obsahuje vysokou úroveň obtížnosti v procesu výpočtu, což způsobuje, že je pro klasifikaci pomocí družicových snímků méně populární. SVM a RF jsou necitlivé k ruchu nebo přetrénování, což ukazuje na jejich schopnost zvládnout nevyvážená data. Proto se SVM a RF, řazené mezi neparametrické metody, stávají velmi populárními a užívanými v případě klasifikaci obrazu.*“

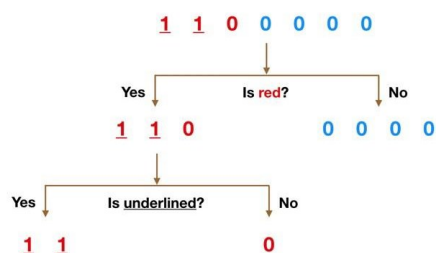
Ok, Akar, Gungor (2012) se věnují klasifikaci plodin pomocí pixelového a objektového přístupu s využitím metod RF a MLC. Nejprve byla provedena pixelová klasifikace jednou i druhou klasifikační metodou, dále byly výsledné mapy překryty parcelami (vektorovými daty) a třída každé parcely byla zvolena na základě největšího zastoupení pixelů v jedné třídě. Postup byl aplikován na území Karacabey Plain v Turecku a vybranými plodinami byly kukuřice, rajče, paprika, pšenice, cukrová řepa a rýže. Kvůli fázím růstu plodin, a tedy odlišným spektrálním charakteristikám, bylo vytvořeno více tříd (krajinného pokryvu) pro každý typ plodiny. Trénovací data byla rozmístěna metodou *stratified random sampling*. Metoda RF byla zvolena na základě svých progresivních vlastností (nepřetrénovává se, tzn. lze tvořit stromů, kolik je potřeba, a příliš velké množství nemá negativní vliv na klasifikaci...) a byla porovnávána s MLC. Dále byla testována různá nastavení parametrů N a m (N je počet stromů a m počet proměnných, které vstupují do každého stromu). Nejefektivnější se v tomto případě ukázalo být nastavení $N = 200$ a $m = 2$. Takovým způsobem bylo dosaženo celkové přesnosti 86 % a kappa indexu 0,79 metodou RF a 78 % a 0,67 metodou MLC. Autoři také upozorňují na skutečnost, že při různém nastavení RF metody lze sledovat konzistentnost výsledků, a metodu tedy považují za vhodnou pro klasifikaci plodin.

Štych a kol. (2019) si ve své studii kladli za cíl otestovat využití dat WorldView-2 a Landsat 8 pro detekci vlivu kůrovce na lesy šumavského národního parku. Jsou zde porovnávána dvojice data – velmi vysokého rozlišení a vysokého rozlišení – a dvě metody klasifikace – SVM a neuronové sítě (NN) – s různými parametry nastavení. Legenda byla zvolena na základě podobných studií, charakteru území a požadavků zaměstnanců NP Šumava, a to 9 kategorií. Trénovací data byla získána terénním výzkumem a kontrolována s ortofoto ze stejného roku. U klasifikace metodou SVM byla volena jádrová funkce. Nejlepší kombinací metod a parametrů se pro tento případ ukázala být metoda SVM s polynomičnou funkcí (s 6 polynomy) a gammou v původním nastavení, která dosáhla 86 % celkové přesnosti a 0,84 kappa indexu (Štych a kol. 2019). Studií bylo prokázáno, že data velmi vysokého rozlišení, konkrétně WorldView-2, mají velký potenciál při klasifikaci lesa. Autoři také poukazují na důležitost rovnoměrnosti rozmístění trénovacích a testovacích dat, která by navíc měla být získávána ve stejnou roční dobu, a dále označují polynomičnou funkci za vhodnější nežli funkce RBF při klasifikaci metodou SVM. Závěrem je doporučeno aplikovat a zkoumat jiné zdroje dat, např. Sentinel-2 nebo Planet klasifikační metodou RF, což by také mohlo přinést velmi slibné výsledky (Štych a kol. 2019).

Na základě výše zmíněných studií vyplývá, že metoda RF je vhodná a perspektivní pro klasifikaci land cover a může dosahovat velmi kvalitních výsledků, pokud jsou správně zvoleny parametry a jsou k dispozici kvalitní referenční data. Právě důležitost kvalitních referenčních dat zmiňuje několik autorů (Noi, Kappas 2017, Štych a kol. 2019); k jejich získání je možné využít regionální mapy, fotografie nebo terénní výzkum. Dále záleží na pravidelném rozmístění testovacích i trénovacích dat. Dosud však nebylo stanoveno žádné univerzální vhodné nastavení metody RF, které by se dalo využít pro jakoukoli oblast. Pro vysokou celkovou přesnost klasifikace je tedy nutné vždy testovat více nastavení parametrů. V následující kapitole je tak detailně popsána metoda RF.

Metoda Random Forest

Metoda Random Forest (náhodný les) vzniká kombinací několika rozhodovacích stromů (Decision Trees) a pixel je vždy zařazen do té kategorie, která měla během klasifikace nejvyšší četnost. Příklad jednoduchého rozhodovacího stromu je na Obr. 1. Metoda rozhodovacích stromů je řízenou klasifikací, kdy vstupem je vektor příznaků daného pixelu a výstupem přiřazení pixelu do jedné z definovaných tříd, a to na základě řady podmínek. Podmínky jsou předem definovány dle příznaků trénovacích dat a jejich přiřazení do tříd. Tak vznikají uzly, resp. podmínky větvení stromů (Potůčková 2020).



Obr. 1 – Jednoduchý příklad rozhodovacího stromu (Zdroj: Yiu 2019)

Metoda navazuje na tzv. bagging (Bootstrap Aggregating), což je algoritmus vybírající pro více modelů různé podmnožiny trénovací datové sady a kombinující předpovědi všech modelů. To vše zcela náhodně (Brownlee 2016).

Parametry metody Random Forest

V metodě Random Forest lze volit počet stromů, které tvoří les. Lesy fungují jako spojování úsilí každého stromu a týmová práce mnoha stromů tak zlepšuje výkon jednoho náhodného stromu. Příliš velký počet však může zpomalovat proces výpočtu. Dalším základním

parametrem je volba počtu proměnných v každém stromě. Výchozí nastavení odpovídá odmocnině celkového počtu vstupních proměnných. Maximální možný počet odpovídá počtu vstupních proměnných. Pro klasifikaci krajinného pokryvu proměnné znamenají jednotlivá spektrální pásma vstupních dat (Srivastava 2015).

Existují další parametry, kterými lze metodu přizpůsobit konkrétnímu úkolu a klasifikaci tak zpřesnit nebo proces urychlit. Jedním je volba části (frakce) prvků v každém stromě. Možné je využít všechna trénovací data v každém stromě nebo jen jejich část, např. 50 %. V každém stromě náhodně zvolenou část, tedy 2 stromy mohou využívat i stejnou část dat. Tzv. větvení udává maximální počet pater každého stromu. Souvisí také s minimálním počtem populace listu, tedy ukončení větvení. Je-li nastaven počet populace listu, větvení skončí pod touto prahovou hodnotou. Např. je-li nastaven minimální počet populace listu 10, v případě, že je v daném souboru 9 prvků, větvení se ukončí. Výchozí nastavení je bez limitu počtu populace listu (Srivastava 2015).

2.3 Klasifikace krajinného pokryvu v Etiopii

Na základě literární rešerše bylo zjištěno několik relevantních studií, které se zabývaly klasifikací krajinného pokryvu v Etiopii. Dle zjištěných informací se autoři potýkali s mnoha překážkami, které fragmentovaná a různorodá krajina Etiopie přináší. Mekuyie, Jordaan, Melka (2018) se zabývají sledováním změn LU, LC v oblasti jižního Afaru v Etiopii mezi roky 1985 a 2015 a hledáním jejich příčin se zaměřením na změny v pastevectví. Pro klasifikaci LU, LC byly využity snímky Landsat stažené z portálu USGS. Třídy byly zvoleny na základě standardních tříd definovaných USGS – křovinaté porosty, obdělávaná pole, travnaté porosty, vodní plochy a holá půda. Byla využita metoda MLC. Během klasifikace se autoři potýkali s problémem oblačnosti. Musely tak být vybírány snímky z období sucha, to však vedlo k obtížné identifikaci listnatých stromů a křovin, které jsou v období sucha bez listů. Oblasti s velkým množstvím vegetace tedy byly klasifikovány na základě snímků z přelomového období sucha a dešťů.

Hishe, Bewket, Nyssen, Lyimo (2019) se ve studii zaměřovali na analýzu LU, LC v Middle Sulah Valley v severní Etiopii, a to jak do minulosti, tak i s predikcí do budoucna. Klasifikace proběhla v softwaru ENVI metodou MLC na základě Landsat snímků z let 1985 až 2015 a leteckých snímků z let 1935, 1965 a 1994. Letecké snímky byly georeferencovány a rektifikovány pomocí podkladových dat na Google Earth. Jako klasifikační třídy byly zvoleny

travnaté porosty, zástavba, křovinaté porosty, holá půda, lesy, orná půda a voda. Byly vytvořeny mapy odpovídající stavu v letech 1985, 2000, 2015 a simulace stavu v roce 2030 podle modelu CA-Markov.

Degife, Zabel, Mauser (2018) zkoumali změny krajinného pokryvu mezi lety 1987 a 2017 v oblasti Gambella, Etiopie, na základě satelitních snímků Landsat 5 a Sentinel-2 Level-2A. Legenda byla zvolena dle CORINE land cover tříd – urbanizovaná území, zemědělská půda, lesní plochy, tropické travnaté porosty, mokřady, holá půda či půda s řídkou vegetací, vodní plochy. Nejprve byla provedena neřízená klasifikace metodou ISODATA a dále řízená MLC, a to s přesností 88 %. Nejvýznamnějším zjištěním byly pozorované změny v NP Gambella a degradace této oblasti. Příčinou byly především velkoplošné investice bez zjištění možných dopadů na životní prostředí.

Fetene, Hilker, Yeshitela, Prasse, Cohen, Yang (2016) ve své studii využili analýzu změn krajinného pokryvu pro vysvětlení změn v ekosystémech v NP Nech Sar v Etiopii. Pro klasifikaci byla využita metoda MLC, snímky Landsat z let 1985 až 2013 a 7 tříd land cover – les, křoviny, dřevnaté travní porosty, dřeviny, traviny, zemědělské plochy, plochy s řídkou vegetací. Kromě družicových snímků byly dále využity fotografie, rozhovory s místními, již provedené analýzy v minulosti, terénní výzkum a meteorologické údaje. Výzkum ukázal, že změny nebyly způsobeny klimatickou změnou, avšak antropogenní činností, a to především odlesňováním a vysazováním nepůvodních druhů. Více než 90% přesnost klasifikace ukázala na využitelnost satelitních snímků pro klasifikaci krajinného pokryvu v podobné geografické oblasti.

Studie autorů Eggen, Ozdogan, Zaitchik, Simane (2016) se věnuje klasifikaci land cover v komplexní a fragmentované zemědělské krajině v etiopské vrchovině. Jedná se o velkou a segmentačně složitou oblast, jejíž klasifikace není jednoduchým úkolem. Právě z toho důvodu byla také pro klasifikaci zvolena metoda SVM, která je schopna zvládnout menší množství trénovacích dat, nepřetrénovává se a nepředpokládá normální rozdělení dat. Aby autoři článku dokázali vhodnost metody, testovány byly další, a to MLC, Decision Tree, k-NN a neuronové sítě). Jejich celková přesnost se od výsledku SVM lišila o 5–25 %. Klasifikováno bylo postupně 14 snímků Landsat do 8 tříd – zemědělská půda, traviny, křoviny/dřeviny, les, holá půda, zástavba/nepropustné povrchy, voda, sezónní voda/bažiny. Trénovací data byla manuálně získávána z Landsat snímků, kdy byly vždy pozorovány 1–2 snímky z období sucha a 1–2 snímky z období dešťů a srovnávány s aplikací Google Earth.

Po několika iteracích, kdy byla pokaždé vytvořena nová trénovací data, bylo dosaženo 55 % celkové přesnosti. Největším problémem klasifikace byla nedostatečná referenční data, dále oblačnost a velká rozmanitost hospodaření s půdou. Právě kvalitnější referenční data by mohla přispět k větší přesnosti klasifikace, i tak považují autoři studie výzkum za přínosný.

Z výše uvedeného plyne, že pro analýzu změn krajinného pokryvu v Etiopii bylo využíváno maximálně 7 klasifikačních tříd. Přičemž celková přesnost klasifikací dosahovala relativně vysokých čísel (80–98 %) v případě lokálních studií. Plošně rozlehlejší studie Eggen, Ozdogan, Zaitchik, Simane (2016) dosáhla 55% celkové přesnosti. Za hlavní problém označují autoři nedostatečná referenční data, dále také oblačnost v období dešťů a rozmanitost zemědělství. Důležitou proměnnou při klasifikaci je tak volba období, ze kterého jsou vybírány družicové snímky, a to především proto, aby byly snímky bez oblačnosti. Mekuyie, Jordaan, Melka (2018) zmiňují, že v období sucha bylo nesnadné rozlišení listnatých stromů a křovin, a tak byly vybírány snímky z přelomu období dešťů a sucha, kdy lze získat bezoblačné snímky a zároveň není porost suchý. Je tedy nutné seznámit se s klimatickými podmínkami daného území a výběr snímků tomu přizpůsobit. Další důležitou proměnnou je kvalitní testovací i kontrolní množina dat. Možnost získat data z terénu v tak rozmanité krajině se jeví jako klíčová. Co se týče klasifikátorů, ve studiích bylo použito metod MLC a SVM. Bohužel není k dispozici studie z posledních let, kdy došlo k rozmachu nových klasifikačních metod.

Volba legendy

Pro přesnou a tematicky správnou klasifikaci krajinného pokryvu je klíčové správně zvolit legendu, resp. klasifikační třídy. Z tohoto hlediska je nutné seznámit se s daným územím, a to pomocí terénního výzkumu, provést vizuální interpretaci leteckých či satelitních snímků např. v aplikaci Google Earth, dále je možné vyhledat regionální mapy a fotografie. Důležité je též konzultovat výběr tříd s koncovými uživateli výstupů klasifikace. Nápomocné mohou být i studie věnující se klasifikaci krajinného pokryvu na podobném území, nebo již vytvořené databáze krajinného pokryvu.

Autoři zmíněných článků volili různou cestu k výběru klasifikačních tříd, jejichž výčet je v Tab. 1. Mekuyie, Jordaan a Melka (2018) založili svůj výběr na terénním výzkumu, vizuální interpretaci satelitních snímků s vysokým rozlišením a pro definici tříd využili tradiční USGS legendu LU/LC a FAO, které přizpůsobili zájmovému území. Hishe, Bewket,

Nyssen a Lyimo (2019) vybrali třídy podle rešerše již provedených výzkumů. Degife, Zabel, Mauser (2018) definují třídy na základě databáze CORINE, přestože byla vyvinuta pro Evropu. Fetene, Hilker, Yeshitela, Prasse, Cohen a Yang (2016) pro zvolení legendy procházeli lokalizované fotografie různých typů krajiny zájmového území, dále vedli rozhovory s místními a provedli rešerši již provedených klasifikací krajinného pokryvu v podobných oblastech v letech 2008 až 2011.

Mekuyie, Jordaan, Melka (2018)	Hishe, Bewket, Nyssen, Lyimo (2019)	Degife, Zabel, Mauser (2018)	Fetene, Hilker, Yeshitela, Prasse, Cohen, Yang (2016)	Eggen, Ozdogan, Zaitchik, Simane (2016)
cultivated land	settlement	artificial areas	cultivated land	agriculture
grassland	agricultural land	farmland	grassland	grassland
bush/shrub land	grass land	forest land	wooded grassland	woodland/shrub
bare land	forest land	tropical grassland	woddland	forest
water body	shrub-bush land	wetlands	forest	bare ground
	bare land	barren or sparsely vegetated land	bush/shrubland	urban/impervious surfaces
	water body	water bodies	area under encroaching plants	water
				seasonal water/marsh areas

Tab. 1 – Přehled použitých legend ve vybraných vědeckých studiích (Zdroj: vlastní)

Dostupné databáze krajinného pokryvu v Africe jsou CCI Land Cover z roku 2016 (CCI 2016), ETH Land Cover EOS4 SD z roku 2017 (ETH 2017) a Africa Land Cover Demo z roku 2018 (ALC 2018). Kategorie, které autoři zvolili, jsou vypsány v Tab. 2. Databáze slouží pro celou Afriku a tomu jsou kategorie přizpůsobeny.

CCI (2016)	ETH (2017)	ALC (2018)
built up areas	agriculture	built-up
cropland	grassland	cropland
grassland	open woodland	shrubland
trees cover areas	dense woodland	open forest - evergreen broadleaved
strubs cover areas	non-vegetated	open forest - deciduous broadleaved
vegetation aquatic or regually flooded		open forest - unknown type
lichen mosses/ sparse vegetation		closed forest - evergreen broadleaved
bare areas		closed forest - deciduous broadleaved
snow and/or ice		closed forest - unknown type
open water		herbaceous vegetation
		herbaceous wetland
		bare/ sparse vegetation
		permanent water bodies

Tab. 2 – Přehled použitých legend ve vybraných databázích LU/LC afrického kontinentu (Zdroj: vlastní)

2.4 Klasifikace krajinného pokryvu v cloudovém prostředí

Pod pojmem tzv. *cloud computing* si lze představit poskytování hostovaných služeb přes internet, ke kterým lze vzdáleně přistupovat pomocí webového prohlížeče. Tento přístup může být veřejný nebo soukromý. Cílem je poskytovat snadný a přizpůsobitelný přístup k výpočetním zdrojům a IT službám. *Cloud computing* funguje na principu poskytnutí přístupu uživatele k datům ze vzdálených serverů, databází a počítačů. Vlastnostmi takového prostředí je sdílení zdrojů se všemi uživateli, tzv. *Pay per use*, tedy platba pouze za skutečné využívání aplikace, dále elasticita a modifikovatelnost výpočetních zdrojů, aktuálnost a dostupnost na internetu odkudkoli. Výhodou je jednoduché uživatelské rozhraní, rychlé přizpůsobení požadavkům uživatele, snížení ceny za software i hardware, zvýšení výkonnosti procesu a další. Nevýhodou je závislost na internetovém připojení, rychle se rozvíjející nástroje, s nimiž je náročné držet krok, nejasné právní podmínky, nejistota zabezpečení, uchování dat a další (Chai 2020).

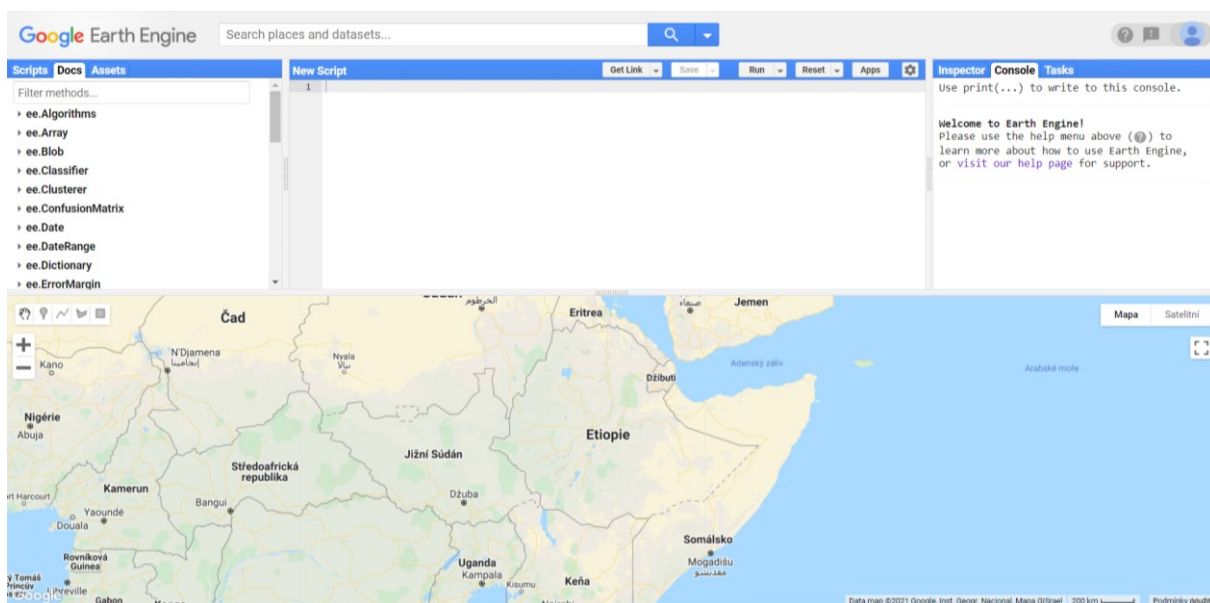
V dálkovém průzkumu Země, kde se pracuje s Big daty, je možnost využití cloudového prostředí pro zpracování dat zásadním usnadněním a zrychlením práce. Příkladem cloudové platformy pro dálkový průzkum jsou v poslední době se rozvíjející Sentinel-Hub nebo GEE, který byl využit v této diplomové práci.

Google Earth Engine

Google Earth Engine je webová platforma, která umožňuje analýzu a vizualizaci geoprostorových databází, a to pro akademické, vládní, komerční i nekomerční uživatele. Pracuje na principu sdílení dat, která není nutné stahovat, jen využívat, a umožňuje tak zpracovávat data velmi rychle a efektivně. Aplikace uchovává množství databází se satelitními snímky za více než 40 let a nabízí celou škálu předpřipravených funkcí pro analýzu a získávání informací (GEE 2020). Funkce umožňují detekovat změny či pozorovat trendy na zemském povrchu. Aplikace je ovládána pomocí skriptů v programovacím jazyce JavaScript a Python a pro využívání je nutné se zdarma registrovat (López 2018).

Prostředí psaní kódu je rozděleno do několika oken (viz Obr. 2). Horní lišta slouží k vyhledávání datových sad. Po rozkliknutí vybrané sady lze číst metadata a následně sadu do kódu přidat. Levé horní okno je děleno na *Scripts*, kde jsou uloženy všechny skripty uživatele nebo skripty připojené od ostatních uživatelů, *Docs* je seznam funkcí včetně nápovědy a

Assets slouží jako úschovna pro připojená data (GeoTIFF/ SHP/ CSV/ Image Collection/ Folder). Středové horní okno je určeno pro psaní samotného kódu, který lze sdílet (*Get Link*), uložit (*Save*), spustit (*Run*) a další. Pravé horní okno je děleno na *Inspector*, kde lze zobrazit informace o určitém bodě na mapě v dolním okně, v záložce *Console* jsou vypisovány výsledky a chybová hlášení a *Tasks* ukazuje seznam úkonů z proběhlého skriptu, které lze spustit tlačítkem Run, např. exportování výsledného obrazu. Své kódy i nahraná data lze sdílet s dalšími uživateli a pracovat tak společně na větších projektech. Výsledky lze vizualizovat na interaktivní mapě v dolním okně nebo je exportovat do *Assets*, na Google Disk nebo jiný cloudový prostor. Platforma GEE nabízí 9 různých klasifikátorů, např. Decision Tree, SVM, Naive Bayes nebo RF.



Obr. 2 – Prostor psaní kódu v Google Earth Engine (Zdroj: vlastní)

3 Data a metodika

3.1 Zájmové území

Etiopie je velká země rozkládající se na více než 1 120 000 km² na východě Afriky v oblasti známé jako Africký roh. Celé zemi dominuje náhorní plošina, která je rozdělená Velkou příkopovou propadlinou udávající ráz okolní krajiny. Výška plošiny se průměrně pohybuje mezi 1 500 a 3 000 m n. m. a nejvyšší vrchol *Ras Dašen* dosahuje výšky 4 620 m n. m. Charakteristické jsou horské hřebeny rozdělené hlubokými roklemi a říčními údolními. Země se nachází v tropické monzunové oblasti s velmi rozmanitým podnebím v závislosti

na nadmořské výšce. Střídají se období dešťů a sucha. Občasné deště obvykle trvají od ledna do února a silné deště přicházejí v polovině června a trvají do poloviny září.

Etiopie se dělí na 10 etnický založených federálních států a 2 samosprávné městské oblasti. Státy se dělí na okresy (woredy) a dále na obce (kebele) (Wikipedie 2020). Federální stát Sidama (viz Obr. 3) je zájmovým územím této diplomové práce.



Obr. 3 – Zájmové území (Zdroj dat: Natural Earth 2021, upraveno)

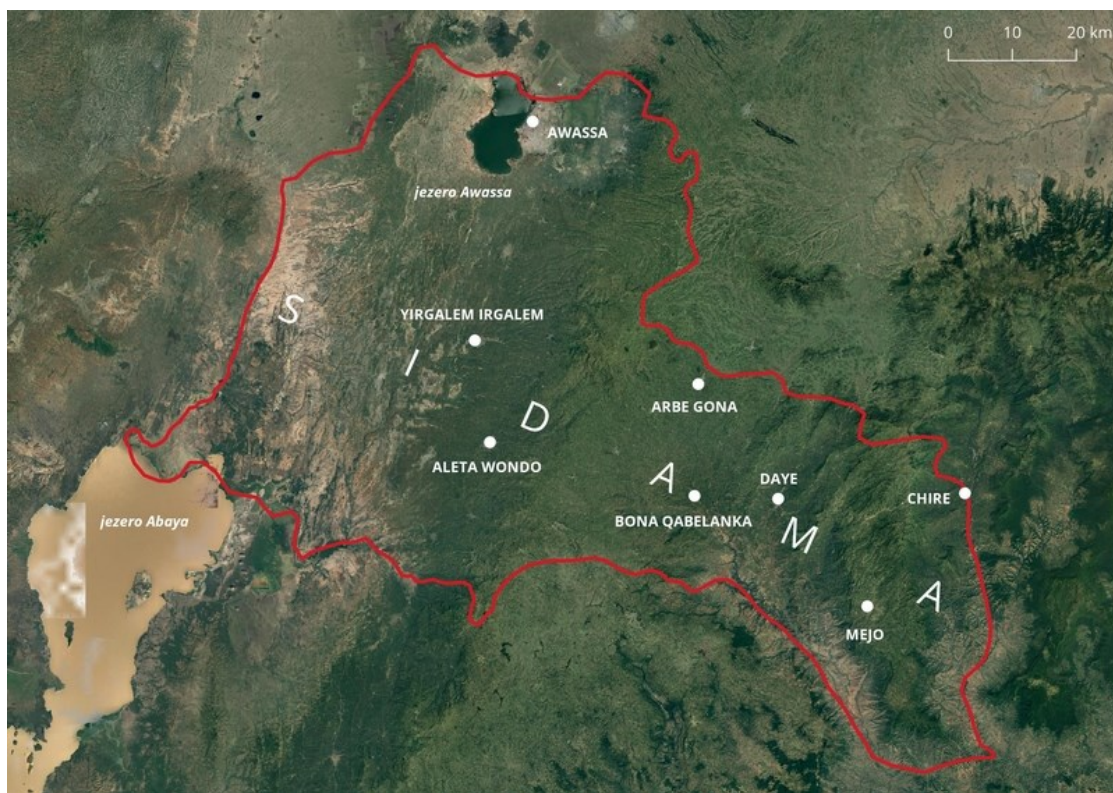
Pro celý svět je Etiopie synonymem hladomoru a vyprahlé pouště (Briggs 2012). Je to však země s bohatou a dlouhou historií, zajímavou kulturou a dechberoucí přírodou. Pouštní oblast je jen velmi málo osídlena a většina obyvatelstva žije poblíž úrodné půdy na náhorní plošině. Ta se pyšní vysokou biodiverzitou i endemickými druhy, např. vlček etiopský, nyala horská nebo pavián dželada (Poláček 2012). Etiopská kultura byla v historii ovlivněna židovskou, egyptskou i různými africkými kulturami. Typická je také velká etnická i náboženská rozmanitost obyvatelstva. Nejvíce zastoupena je zde pravoslavná církev, dále islám a protestantismus. V Etiopii žije více než 100 mil. obyvatel a roční přírůstek je 2,6 %. Zdravotní systém je nedostačující a 26 % obyvatelstva na vesnicích nemá přístup k pitné vodě. Ženy jsou diskriminovány, často prodávány jako otrokyně. Mnoho dětí je nuceno pracovat nebo žije na ulici. Negativní vliv na fungování země měly také války se Somálskem

a Eritreou v 90. letech 20. století. Etiopie patří mezi nejchudší státy světa, a přestože se v poslední době rychle rozvíjí, je závislá na zahraniční pomoci (Wikipedie 2020).

Mnohé vyspělé země nabízejí Etiopii humanitární pomoc v podobě lékařské či sociální péče. Ze strany Česka jsou to organizace Člověk v tísni, ADRA či Česká rozvojová agentura. Tyto organizace realizují projekty na podporu lékařské péče, vzdělávání v různých oblastech a zapojují se do boje proti suchu a nedostatku pitné vody. Jsou to například projekty „*Postavme školu v Africe*“, „*Obnova studen v regionu Somali*“ nebo „*Zlepšení kvality života obyvatel zóny Sidama cestou zajištění dostupnosti a udržitelného hospodaření s vodními zdroji*“ (Člověk v tísni 2020, ADRA 2014, ČRA 2020).

3.1.1 Fragmentovaná krajina regionu Sidama

Sidama je od roku 2020 jedním ze svazových států Etiopie a leží na jihozápadě země. Více než 3 mil. obyvatel žije na více než 6 000 km². Hlavním a největším městem je Awassa, které leží na severu státu u stejnojmenného jezera (viz Obr. 4). Obyvatelstvo tvoří z 93 % Sidamové, mateřským jazykem je sidamština a 84 % věřících jsou protestanté. Je to hornatá oblast průměrně v 1 800 m n. m. (Wikipedie 2021). Klimaticky se dá region rozdělit na 3 oblasti (viz Tab. 4). Horká oblast se nachází na západě regionu a na většině území se rozkládá rozsáhlá pustina, která je jen řídce osídlena. Na východě leží horská oblast, která je vlhká a úrodná. Centrální, teplá část je nížinná a sušší (viz Obr. 5).



Obr. 4 – Stát Sidama s významnými městy (Zdroj dat: Google Earth 2021, upraveno)



Obr. 5 – Různorodá krajina Sidamy, zleva teplá, horká a horská oblast (Zdroj: vlastní)

Zpevněné cesty spojují větší města. Cesty mimo města jsou nezpevněné, prašné a v období dešťů blátivé. V oblastech hustšího osídlení lemují cesty pravoúhlé domy z hlíny s plechovými střechami. Ve větších městech jsou stěny domů z cihel a mají několik pater.

Budovy v roztroušené zástavbě v zemědělských oblastech mají kruhový tvar a střechu ze slámy. Pro nížinnou oblast jsou typická rozsáhlejší pole plodin jako kukuřice, teff (místní obilovina), cukrová třtina nebo jiné těžko identifikovatelné plodiny. Jihozápadní část je suchá a často se zde vyskytují tzv. bad landy, sesuvy půd, strže, popraskaná půda a koryta sezónních řek. Východní oblast je nejvíce fragmentovaná. Zdejší typická krajina znamená nepočetná hliněná obydlí se slaměnou střechou, v okolí několik políček s různými druhy plodin (banánovníky, kukuřice, luštěniny...) v rozloze pár desítek čtverečních metrů, pastvina a okolo plot (viz Obr. 6). Ploty mohou mít různou podobu – ze dřeva, z kaktusů, z keřů a výjimečně z plechu. Podél cest často vede řada keřů či stromů.



Obr. 6 – Typická krajinná struktura východní části regionu Sidama (Zdroj: Maxar 2021)

Řeky bývají v období dešťů rozlité do okolních luk. V přechodném období je těsné okolí řek a potoků podmáčené a vznikají tak mokřady (viz Obr. 7). V období sucha bývají některé řeky vyschlé. Kvůli častým sesuvům jsou řeky a jezera znečišťována půdou, lze je tedy dělit na vodní plochy a toky s čistou a kalnou vodou (viz Obr. 4 jezero Awassa a jezero Abaya).

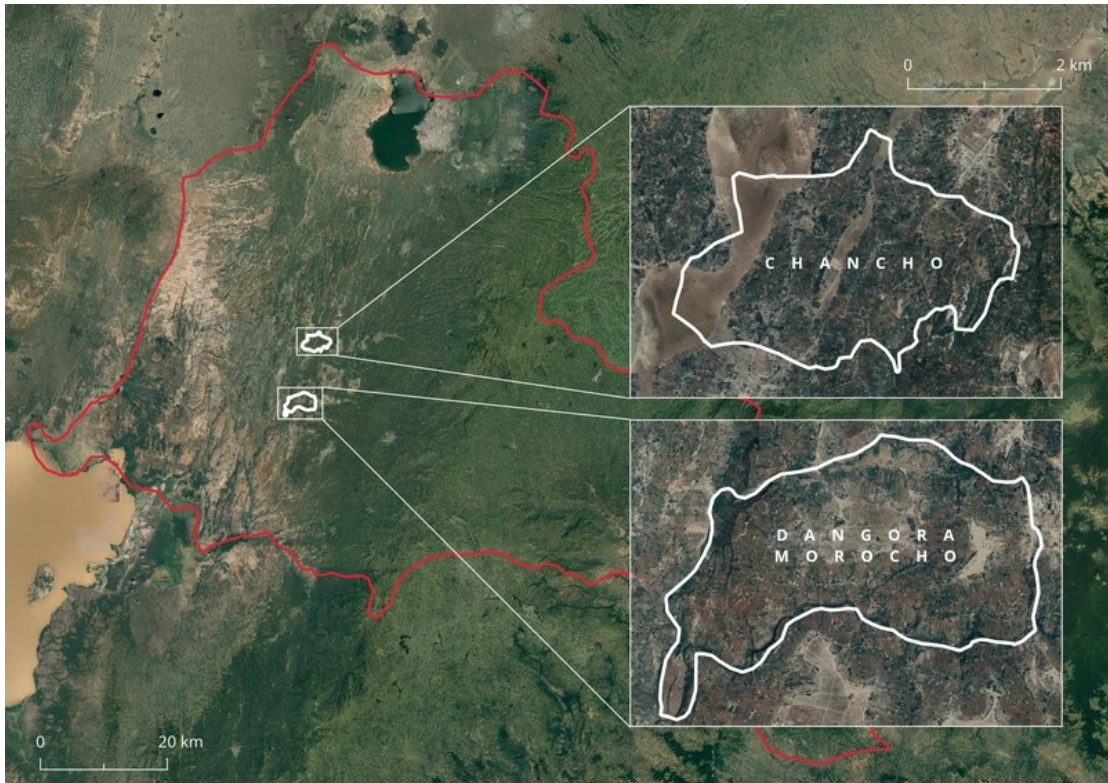


Obr. 7 – Ukázka mokřadu, vlevo snímek z února 2021, vpravo z listopadu 2020 (Zdroj: Maxar 2021)

Region Sidama se nachází v klimatické oblasti monzunů, kdy období sucha trvá od října do února, potom nastává období občasných dešťů a od června do poloviny září je období velmi vydatných dešťů. V období dešťů dozrává nejvíce plodin, zároveň je však území téměř neustále zahaleno mraky. Nejméně oblačnosti vychází na leden, únor, to však už dlouho panuje v zemi sucho a mnoho plodin je sklizeno. Z toho vyplývá, že ideální doba pro klasifikaci land cover pomocí optických družicových dat je říjen, listopad, kdy už tolik neprší, ale zároveň je vegetace ještě vzrostlá a zelená.

Kebele Chanco a Dangora Morocho

Kebele (= obce) jsou nejmenší správní celky Etiopie. V projektu ČRA je analýza vybraných kebelů jedním z hlavních cílů a na jejím základě budou tvořeny krajinné plány se zaměřením na protierozní opatření. Pro tuto práci byly vybrány 2 kebele v regionu Sidama. Kebele Chanco z woredy Loko Abaya a kebele Dangora Morocho z woredy Chuko byly vybrány na základě sběru in-situ dat autorkou práce přímo v těchto oblastech. Obě kebele se nacházejí v teplé klimatické oblasti v západní části regionu (viz Obr. 8).



Obr. 8 – Vybrané kebele regionu Sidama (Zdroj dat: Google Earth 2021, Open AFRICA 2020)

3.2 Data

3.2.1 Sentinel-2

Mise programu Copernicus s označením Sentinel-2 tvoří od roku 2017 dvě družice na stejné oběžné dráze s posunem o 180° , které poskytují multispektrální snímky s vysokým prostorovým i časovým rozlišením. Totožné místo je zaznamenáno minimálně jednou za 5 dní. Optický senzor na družicích snímá 13 pásem (viditelné až střední infračervené spektrum) s rozlišením 10, 20 a 60 m a šířkou záběru 290 km. Cílem mise je především monitorování krajinného pokryvu a jeho změn. Produkt Level-2A je volně dostupný uživatelům a obsahuje atmosférickou korekci dat (Žubrientovský, Švábová, Šváb 2017).

Do GEE byla kolekce Sentinel-2 MSI Level-2A připojena pomocí příkazu `ee.ImageCollection("COPERNICUS/S2_SR")`. Kolekce obsahuje snímky od 28. 3. 2017 po současnost s 24 pásmy a 57 atributy. Kromě spektrálních pásem jsou to pomocná pásma zaměřená na oblačnost, land cover a RGB kompozit. V tabulce 3 jsou uvedena jen ta pásma, která vstupovala do klasifikace. Prvních 10 pásem jsou spektrální pásma, která vstupovala do klasifikace obrazu, poslední 2 (SCL a QA60) jsou pomocná pásma, která sloužila k tvorbě masky oblačnosti.

Název	Velikost pixelu [m]	Vlnová délka S2A [nm]	Vlnová délka S2B [nm]	Popis
B2	10	496,6	492,1	Blue
B3	10	560,0	559,0	Green
B4	10	664,5	665,0	Red
B5	20	703,9	703,8	Red Edge 1
B6	20	740,2	739,1	Red Edge 2
B7	20	782,5	779,7	Red Edge 3
B8	10	835,1	833,0	NIR
B8A	20	864,8	864,0	Red Edge 4
B11	20	1613,7	1610,4	SWIR 1
B12	20	2202,4	2185,7	SWIR 2
SCL	20			Scene Classification Map (The "No Data" value of 0 is masked out). 1-11
QA60	60			Cloud mask

Tab. 3 – Využitá pásma datasetu Sentinel-2 MSI Level-2A (Zdroj: Sentinel-2 MSI 2021)

V práci bylo využíváno 6 snímků ze dne 19. 11. 2020 (viz Tab. 4). Datum bylo vybráno tak, aby snímky obsahovaly co nejméně oblačnosti, existovaly pro celé zájmové území a zároveň aby datum bylo blízké sběru in-situ dat. In-situ data byla sbírána v regionu Sidama v Etiopii na podzim roku 2020, více je popsáno v kapitole 3.3.2.

	ID snímku
1	20201119T074221 20201119T074219 T37NDH
2	20201119T074221 20201119T074219 T37NDG
3	20201119T074221 20201119T074219 T37NCH
4	20201119T074221 20201119T074219 T37NEG
5	20201119T074221 20201119T074219 T37NEH
6	20201119T074221 20201119T074219 T37NCG

Tab. 4 – Vybrané snímky Sentinel-2 (Zdroj: vlastní)

3.2.2 PlanetScope

Planet Lab Inc. je soukromý poskytovatel prostorových dat již od roku 2010. Jeho cílem je každodenní zobrazení celé Země tak, aby bylo možné sledovat změny povrchu, a to ve velmi vysokém prostorovém i časovém rozlišení. Společnost navrhuje a vyrábí malé satelity Triple-CubeSat s názvem Doves (překl. z ang. holubice), které společně v tzv. hejnech nasnímají za jeden den celý povrch Země v rozlišení 3 až 5 m. K dispozici jsou snímky RapidEye, SkySat a PlanetScope, přičemž snímky RapidEye mají rozlišení 5 m v 5 pásmech (RGB, Red-Edge, NIR), SkySat 0,5 m v 5 pásmech (RGB, NIR, panchromatické) a PlanetScope 3 m ve 4

pásmech (viz Tab. 5). Data PlanetScope se distribuují pomocí dlaždic o rozměru 24 x 8 km. V roce 2021 společnost provozuje více než 200 aktivních satelitů (Planet 2021).

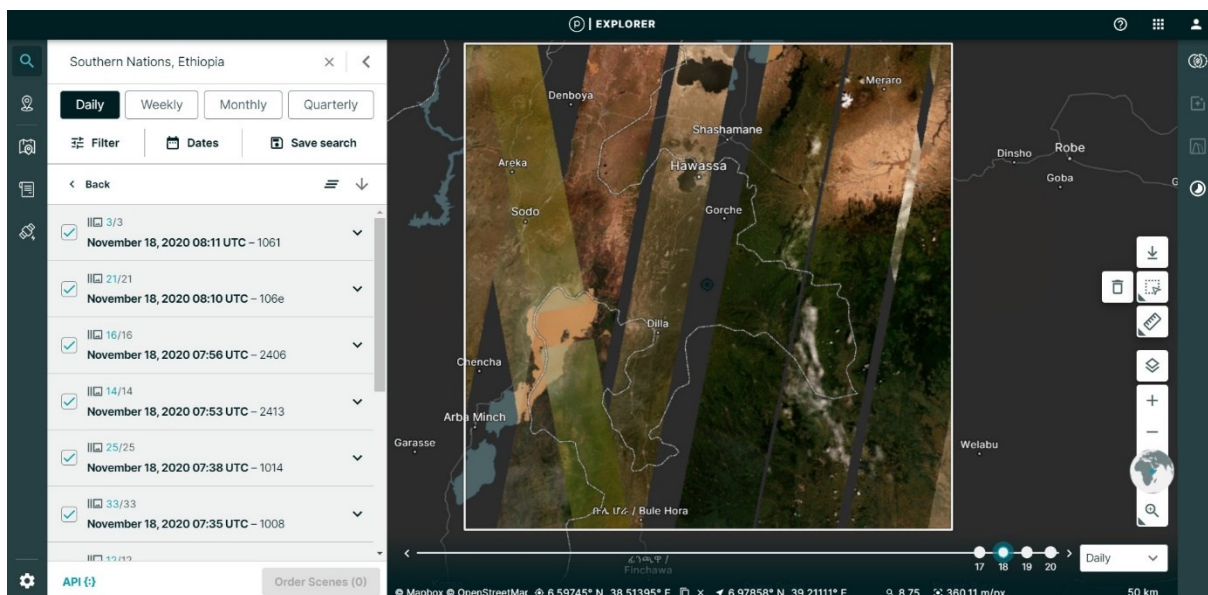
Název	Vlnová délka [nm]
Blue	455–515
Green	500–590
Red	590–670
NIR	780–860

Tab. 5 – Pásma datasetu PlanetScope (Zdroj: Planet 2021)

Snímky PlanetScope byly využívány pro klasifikaci land cover 2 menších lokálních území (tzv. kebelí). Dva vybrané snímky PlanetScope ze dne 18. 11. 2020 (viz Tab. 6) byly staženy z portálu Planet.com (viz Obr. 9) ve formátu GeoTIFF pod studentskou licenci. Snímky tohoto data byly použity z důvodu nedostatečné kvality snímků ze dne 19. 11. 2020. Ze dne 19. 11. 2020 (stejně datum, kdy byly nasnímány snímky Sentinel-2 využité pro klasifikaci regionu Sidama) byly snímky dostupné, ale byly nasnímány v 5 hodin 42 minut ráno. Slunce ten den vycházelo v 6 hod 12 minut. Snímky byly tedy pořízeny 30 minut před východem slunce. Obraz nebyl ostrý a v údolí řeky byl tmavý stín. Snímky pořízené týden před a po 19. listopadu byly nasnímány mezi 7. a 8. hodinou ranní. Vybrané snímky ze dne 18. 11. 2020 nasnímané v 7 hodin 35 minut byly ostřejší a vhodnější pro klasifikaci. Do GEE byly snímky nahrány do záložky *Assets*. Do skriptu klasifikace byly postupně načteny příkazem `ee.Image('users/valcharda/DP/planetscope_527')`.

	ID snímku
1	20201118_073527_1008_3B_AnalyticMS_SR
2	20201118_073528_1008_3B_AnalyticMS_SR

Tab. 6 – Vybrané snímky PlanetScope (Zdroj: vlastní)



Obr. 9 – Pokrytí regionu Sidama dostupnými snímky PlanetScope 18. 11. 2020 (Zdroj: vlastní)

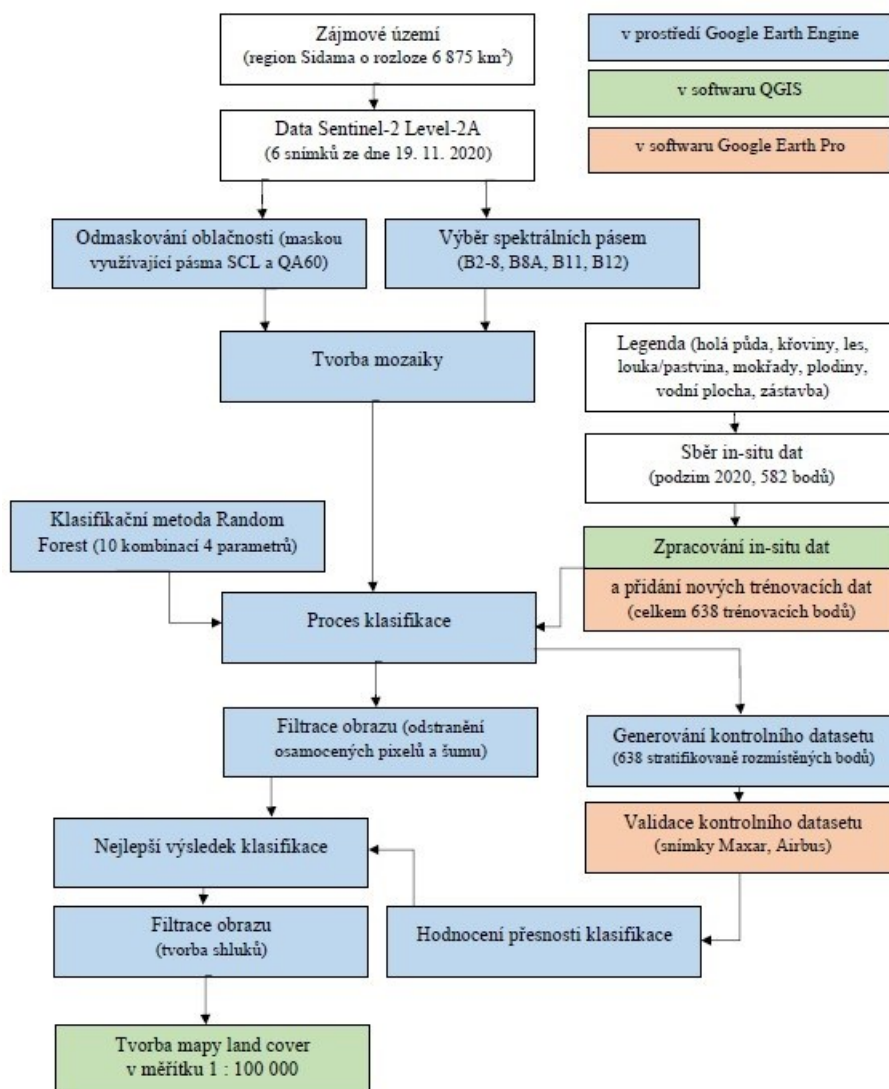
3.2.3 Další podpůrná data

Hranice administrativních jednotek Etiopie byly staženy z otevřené databáze Open Africa z roku 2016 ve formátu SHP (Open AFRICA 2020). Pro terénní výzkum byl pro lepší orientaci v prostoru a porovnání změn v průběhu roku připraven mapový projekt, který obsahoval družicové snímky Sentinel-2 Level-1C ze dne 31. 10. 2019 a 8. 2. 2020 v RGB syntéze v pravých i nepravých barvách (USGS 2020) a snímky Landsat 8 OLI ze dne 19. 1. 2020 (USGS 2020). Jako další podpůrná data pro seznámení se s územím a tvorbu kontrolní množiny dat byl využit mapový portál Google Earth a software Google Earth Pro (GEP). Google Earth zobrazuje podle úrovně přiblížení postupně mozaiky satelitních snímků MODIS, dále snímků Landsat 8 a pro největší přiblížení letecké snímky, jejichž rozlišení a aktuálnost se liší místo od místa (15 cm – 15 m, 3–5 let) (Google Earth 2021). Pro tvorbu kontrolní množiny dat byl využit software GEP a nabízené snímky s velmi vysokým rozlišením, a to buď CNES Airbus nebo Maxar Technologies s přesností 30 cm/px (Airbus 2021, Maxar 2021). Data pořízení a dostupnost snímků se na území regionu Sidama lišila. Nejvíce dat bylo dostupných v okolí velkých měst, v odlehlých oblastech bylo snímkování méně frekventované. Pro velkou část území byly dostupné snímky z podzimu 2020, dále také 4 roky do minulosti a ty nejaktuálnější pocházejí z února 2021.

3.3 Metodika

Hlavním cílem práce je klasifikace krajinného pokryvu regionu Sidama v roce 2020 pro tvorbu mapy land cover v měřítku 1 : 100 000. Pro klasifikaci byla využita data Sentinel-2 a PlanetScope, byla zvolena vhodná legenda na základě rešerše a terénního výzkumu a byla zvolena MMU podle požadavků koncových uživatelů a vstupních dat. Na podzim roku 2020 byla sbírána trénovací množina dat do předpřipraveného projektu v aplikaci ArcGIS Collector. Data byla zpracovávána v programu QGIS a GEP. Tvorba algoritmu klasifikace byla realizována v otevřené cloudové platformě GEE, kde byly snímky naimportovány z knihovny, byla odmaskována oblačnost a snímky sjednoceny do mozaiky. Pomocí grafů sestavených v GEE byly zkoumány spektrální charakteristiky sledovaných land cover tříd z trénovacího datasetu. Byly odstraněny tzv. mixed-pixels, aby bylo dosaženo větší přesnosti. Klasifikace byla natrénována a proběhla nad vybranými snímky Sentinel-2 ze dne 19. 11. 2020.

Na základě vlastního testování a doporučení odborné literatury (Ok, Akar, Gungor 2012, Noi, Kappas 2017, Štych a kol. 2019) byla vybrána klasifikační metoda RF. Nejdůležitějším výzkumně-metodickým úkolem byla volba nastavení parametrů klasifikátoru, které vedlo k co nejvyšší celkové přesnosti klasifikace. Z uživatelského hlediska se též bral v potaz aspekt časové délky výpočtů, tzn. aby proces výpočtů trval přiměřeně dlouhou dobu. Během experimentální činnosti bylo testováno 10 kombinací 4 různých parametrů. Dále byla provedena filtrace obrazu, kdy byly odstraněny osamocené pixely a ostatní šumy. Také byly pixely sjednoceny do shluků, aby výsledná klasifikace byla v měřítku 1 : 100 000 a odpovídala tak požadavkům koncového uživatele. Kromě toho byla hodnocena přesnost klasifikace, a to pomocí chybové matice a kappa indexu, jejichž výpočty byly odvozeny od kontrolního datasetu. Kontrolní dataset byl vytvořen v prostředí GEE, kde byly generovány náhodné body na základě rozlohy jednotlivých tříd klasifikace. Kontrola těchto bodů proběhla v softwaru GEP pomocí obsažených snímků s velmi vysokým rozlišením Maxar a Airbus, které časově co nejvíce odpovídaly datu pořízení družicových snímků použitých pro klasifikaci. Stejným způsobem proběhla klasifikace dat velmi vysokého rozlišení PlanetScope s požadavkem koncového uživatele na měřítko 1 : 10 000. Výsledný obraz klasifikace byl v GEE exportován ve formátu GeoTIFF a upraven do podoby mapy v softwaru QGIS. Jednotlivé metodické kroky jsou znázorněny na obrázku 10 a podrobně popsány v následujících kapitolách.



Obr. 10 – Schéma metodiky (Zdroj: vlastní)

3.3.1 Legenda a mapovací jednotka

Na základě provedené literární rešerše (viz kapitola 2.3) a terénního výzkumu (viz kapitola 3.3.2) byla definována legenda s 8 třídami, a to zástavba (settlement), plodiny (cropland), louka/pastvina (grassland/pasture), les (forest), křoviny (shrubland), holá půda (bareland), mokřad (wetland) a vodní plocha (water body) s přesnější definicí:

1. Zástavba zahrnovala všechny budovy s plechovou či jinak pevnou střechou, zpevněné silnice a jiné umělé objekty.

2. Do třídy plodiny byly zařazeny všechny zemědělské plochy, kde byly v době klasifikace pěstovány plodiny, jmenovitě kukuřice, chat, teff, cukrová třtina, banánovníky a další.
3. Louky a pastviny byly definovány jako hustý travnatý porost, kde dominují traviny, často využívaný pro pastvu. Za známku pastvy byl považován pasoucí se dobytek, stopy dobytka nebo výkaly.
4. Les byl definován jako souvislá vegetace složená převážně ze stromů či vyšších keřů s minimální výškou 5 m a pokrytí korunami min 30 % plochy. Na území se vyskytovaly eukalyptové, bambusové, jiné listnaté, jehličnaté či smíšené lesy.
5. Pod třídou křoviny byly všechny plochy smíšené vegetace složené převážně z nižších keřů, bylin či travin s občasnými rozptýlenými stromy. Hlavním faktorem bylo zarůstání ploch bez zásahu člověka, jednalo se tedy o pustiny s řídkou vegetací, zarostlé svahy, či pruhy vegetace kolem cest.
6. Holá půda byla definována jako oblast bez pokrytí vegetací, zahrnující vypálené oblasti, skály, tzv. bad landy a nezpevněné cesty. Do této třídy spadaly také oblasti povrchové těžby nerostných surovin.
7. Třída mokřadů zahrnovala nezalesněné oblasti částečně, sezónně nebo trvale podmáčené, a to včetně rákosí, mangrove a vlhkých luk. Tyto oblasti se nacházely především v okolí řek a jezer.
8. Vodní plochy byly definovány jako přírodní nebo umělé vodní toky a plochy, kde přítomnost vody je po většinu roku, přičemž minimální šířka vodních toků byla stanovena na 10 m. Konkrétně to byla jezera, řeky a vodní nádrže. Příklady fotografií všech tříd land cover jsou v příloze 1.

Byla také stanovena MMU. Její stanovení bylo založeno na těchto dvou faktorech – na požadavcích koncových uživatelů (mapové výstupy v měřítku 1 : 100 000) a na vlastnostech použitých dat Sentinel-2 (10–20 m/px). Vzhledem k vlastnostem snímků Sentinel-2 bylo počítáno s pixelem o velikosti 10–20 m. Klasifikace tedy proběhla s rozlišením 10 m. Následně byl obraz filtrován, tak aby byly odstraněny šумы (např. osamocené pixely) a následně kvůli požadovanému měřítku 1 : 100 000 zachovány prvky o velikosti 2500 m², což odpovídá Toblerovu pravidlu (Nagi 2010). Více o filtraci v kapitole 3.3.6.

3.3.2 Terénní výzkum a sběr trénovacích dat

Terénní výzkum zájmové oblasti proběhl na podzim roku 2020 pod záštitou projektu ČRA. Expedice byla naplánována na 4 týdny v oblasti Sidama a Gedeo, Etiopie. Mapování zahrnovalo kromě sběru dat pro klasifikaci krajinného pokryvu také geologické, hydrologické, pedologické a geomorfologické mapování, na kterém se podílel tým vědeckých pracovníků z České geologické služby, Etiopské geologické služby, Přírodovědecké fakulty Univerzity Karlovy a společnosti All4Soil. V této kapitole je popsána příprava na terénní výzkum, způsob mapování v terénu a následné zpracování získaných dat.

Příprava na terénní výzkum

Pro seznámení se s územím byl proveden průzkum oblasti Sidama pomocí aplikace Google Earth, Google Street View, Mapy.cz a byla stanovena legenda (viz kapitola 3.3.1).

Předem byly stanoveny tyto cíle terénního výzkumu:

- Sbírat body (trénovací data) pro všechny předdefinované třídy v zájmovém území, a to pomocí programu ArcGIS Collector v tabletu s GPS zařízením.
- Ke každému záznamu vést informaci o poloze, třídě a datu sběru. K některým záznamům byly navíc přidány fotografie.
- Posbírat co největší počet kvalitních trénovacích dat rozmístěných, pokud možno, co nejvíce rovnoměrně po celém zájmovém území.
- Všimnout si podobností, struktury, charakteristiky jednotlivých tříd a jejich přechodů – zaznamenat pomocí poznámek a fotografií.
- Definovat spektrální charakteristiky klasifikačních tříd.
- Stanovit vhodné RGB kompozity pro zvýraznění daných tříd.
- Zpřesnit legendu na základě výzkumu a stanovit MMU (rozloha, šířka).

Pro účely sběru dat byl vytvořen mapový projekt v aplikaci ArcGIS Online, který obsahoval několik vrstev pro orientaci v prostoru a pro zhrubé porovnání změn v průběhu roku. Vstupními podklady byly satelitní snímky území z různých ročních období a v barevných syntézách, dále hranice zóny Sidama a jednotlivých wored (blíže jsou tato data

popsána v kapitole 3.2.3) a vytvořená bodová vrstva pro sběr dat. Mapový projekt byl otevřen v aplikaci ArcGIS Collector v tabletu Samsung Galaxy Tab Active 2. Dle výrobce je přesnost GPS zařízení v tabletu 5 m.

Sběr dat v terénu

Terénní výzkum v regionu Sidama (viz Obr. 11) probíhal na podzim roku 2020 po dobu 4 týdnů. Stanovení místa, kde bylo možné sbírat data, bylo tehdy ovlivněno aktuálními podmínkami v daném území a rozhodnutími ostatních vědců ve výzkumném týmu pro účely naplnění potřeb projektu. Dataset trénovacích dat tak neobsahoval rovnoměrně rozmístěná data, pouze data z míst, kam se tzv. „dalo dostat“. Mnohá místa v regionu byla dostupná pouze pěšky nebo na motorce a jiná zcela vůbec – neprostupný terén nebo nezpevněné cesty, které se po dešti staly neprůjezdnými. Velká část mapovaného území byla zastavěná a obydlí lemovaly ploty, nebylo tedy možné vstupovat přímo dovnitř jednotlivých plošek krajinného pokryvu. Tento problém byl řešen manuálním zpřesněním lokalizace bodů, povětšinou o několik metrů doprostřed sledovaných plošek.



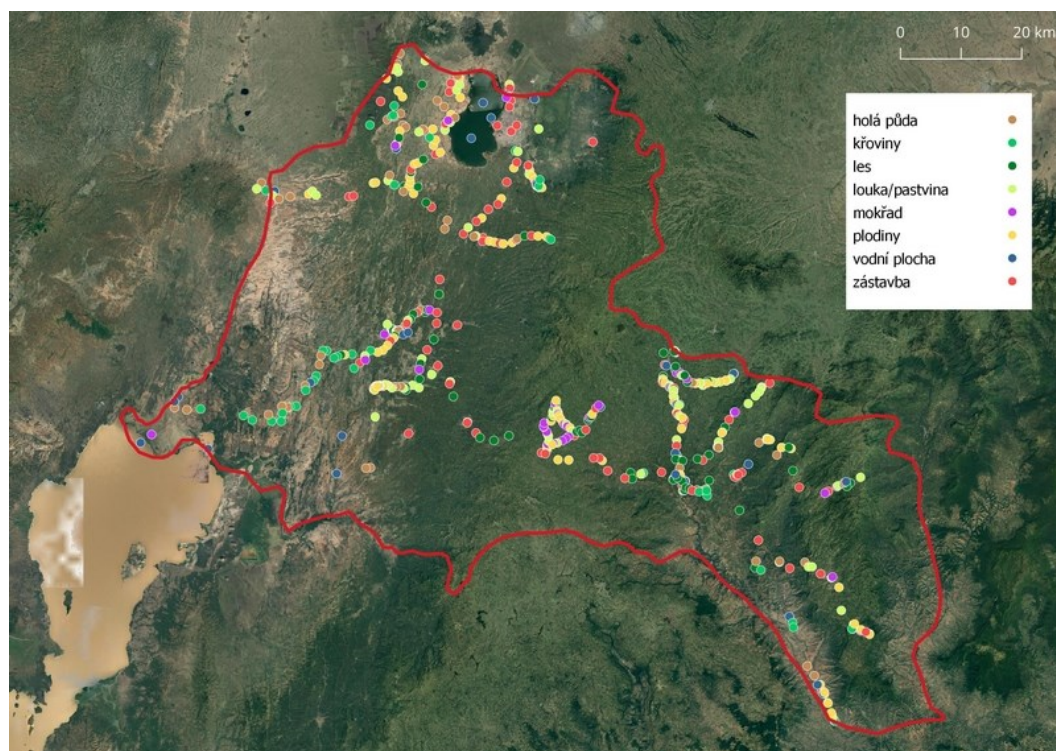
Obr. 11 – Terénní výzkum v regionu Sidama na podzim 2020 (Zdroj: Jan Ureš)

Každý bod byl zaznamenán do aplikace ArcGIS Collector, a to s parametry souřadnic X a Y, třídou (1–8), poznámkami a případně fotografiemi. Do poznámek byly psány bližší specifikace, např. druh plodiny, druh lesa, odhadovaná výška stromů, odhadovaná šířka

vodního toku, materiál budov apod. (viz Tab. 7). V terénu v regionu Sidama a blízkém okolí bylo nasbíráno 582 bodů (viz Tab. 6 a Obr.12).

ID	Třída	Poznámky	X	Y
300	1	vesnice s hlíněnými domky a plechovými střechami, asfaltová silnice středem	38,53268	6,557561667
301	4	vysoké stromy jehličnaté, 10 m a více	38,54157	6,56152863
303	2	falešné banánovníky	38,54147	6,568129432

Tab. 7 – Ukázka části atributové tabulky dat z terénu (Zdroj: vlastní)



Obr. 12 – Rozmístění dat z terénu v regionu Sidama (Zdroj podkladových dat: Google Earth 2021)

Zpracování naměřených dat

Kvůli nestabilnímu internetovému připojení v Etiopii nebylo bezpečné pracovat s daty již během terénního výzkumu nebo v průběhu expedice. Synchronizace aplikace ArcGIS Collector s ArcGIS Online aplikací je velmi citlivá na dobré internetové připojení a mohou nastat komplikace nebo dojít až ke ztrátě všech dat. Je proto nutné data pravidelně zálohovat.

Po příjezdu z expedice byla data zpracovávána dle tohoto postupu:

- Synchronizace jednoho tabletu a kontrola zobrazení dat v aplikaci ArcGIS Online.

- Synchronizace druhého tabletu a kontrola zobrazení dat z obou tabletů v aplikaci ArcGIS Online.
- Stažení obou souborů SHP z internetu.
- Nahrání obou souborů SHP do programu QGIS a jejich následná úprava.

Data byla zobrazena v programu QGIS, kde byly nejprve opravovány překlady a dále všechny poznámky přeloženy do angličtiny pro případné využití kýmoli dalším v rámci projektu. Polohová přesnost každého bodu byla zkontrolována v softwaru GEP, kde byly některé body polohově a tematicky zpřesněny. Dále byly odstraněny tzv. mixed-pixels, tedy pixely, které obsahovaly více klasifikačních tříd.

Co se týče potřebného počtu trénovacích ploch, pak Červená (2016b) uvádí minimální počet jako $N + 1$ pro každou třídu, kdy N je počet klasifikovaných spektrálních pásem. Do klasifikace vstupovalo 10 pásem snímků Sentinel-2 (více popsáno v kapitole 3.2.1). Z terénního šetření nebyla data rovnoměrně rozmístěna v prostoru celého regionu Sidama a nebyl nasbírán dostatečný počet vzorků pro každou klasifikační třídu. Prázdná místa byla proto doplněna nad aktuálními snímky s velmi vysokým rozlišením Maxar a Airbus v softwaru GEP, a to především třídy, které obsahovaly nejméně vzorků – křoviny, mokřady a vodní plochy. Ve výsledku čítal celkový počet trénovacích dat vstupujících do klasifikace 638 bodů (viz Tab. 8).

Třída	Počet in-situ nasbíraných bodů	Celkový počet bodů s využitím snímků
Holá půda	55	61
Křoviny	44	65
Les	82	110
Louka/pastvina	93	91
Mokřady	40	64
Plodiny	131	110
Vodní plocha	45	56
Zástavba	92	81
Celkem	582	638

Tab. 8 – Počet trénovacích bodů jednotlivých klasifikačních tříd (Zdroj: vlastní)

3.3.3 Předzpracování snímků

Pro klasifikaci krajinného pokryvu regionu Sidama bylo vybráno 6 snímků Sentinel-2 Level-2A ze dne 19. 11. 2020, které pokrývají celé zájmové území. Snímky byly načteny a zobrazeny v RGB a ořezány polygonem regionu Sidama ve formátu SHP (viz Obr. 13).

Kód pro výběr konkrétního snímku 1/6 a jeho vizualizace:

```
var collection1 = ee.ImageCollection('COPERNICUS/S2_SR')
  .filterBounds([38.43,6.71])
  .filterDate('2020-11-18', '2020-11-20')
  .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE',20))
  .sort('CLOUDY_PIXEL_PERCENTAGE')
  .select(['B2','B3','B4','B5','B6','B7','B8',
          'B11','B12','B8A','QA60','SCL']);
var region = ee.FeatureCollection("users/valcharda/DP/region_sidama");
Map.addLayer(collection1.clip(region),{min:0,max:3000,bands:"B4,B3,B2",
gamma: 1.25},'collection1', 0);
Map.setCenter(38.5,6.65, 9);
```



Obr. 13 – Vybrané snímky Sentinel-2 ze dne 19. 11. 2020 (Zdroj dat: Sentinel-2 MSI 2021, upraveno)

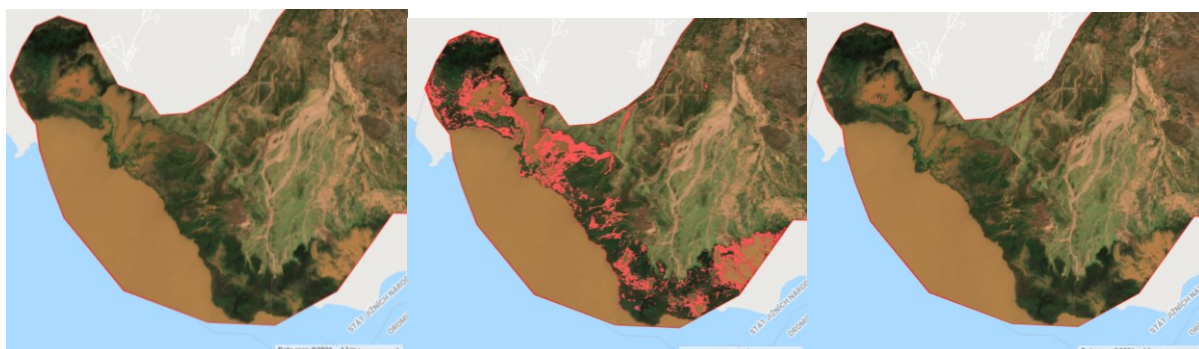
Vybrané snímky Sentinel-2 byly téměř bezoblačné. Řada několika oblaků se nacházela ve středu zájmového území a několik mraků pak v oblasti na východě území (viz Obr. 12). Pro odmaskování oblačnosti byla využita maska skládající se z pásme QA60 a SCL. QA60 je

tzv. *Cloud mask* s 60m pixelem a nabývá hodnot 0 a 1, kdy 0 je pro žádnou oblačnost a 1 pro oblačnost. Pásmo SCL je tzv. *Scene Classification Map* s 20m pixelem a nabývá hodnot 1 až 11 (viz Tab. 9). Maska byla sestavena z hodnoty 0 pro pásmo QA60 a hodnot 1, 2, 4, 5, 6 a 7 pro pásmo SCL, jedná se tedy o hodnoty, které mají ve snímcích zůstat.

Hodnota	Barva	Popis (původní)	Popis (překlad)
1	ff0004	Saturated or defective	Hodně zabarvené nebo vadné pixely
2	868686	Dark Area Pixels	Tmavé pixely
3	774b0a	Cloud Shadows	Stíny mraků
4	10d22c	Vegetation	Vegetace
5	ffff52	Bare Soils	Holá půda
6	0000ff	Water	Vodstvo
7	818181	Clouds Low Probability / Unclassified	Nízká pravděpodobnost oblačnosti/ neklasifikováno
8	c0c0c0	Clouds Medium Probability	Střední pravděpodobnost oblačnosti
9	f1f1f1	Clouds High Probability	Vysoká pravděpodobnost oblačnosti
10	bac5eb	Cirrus	Cirry
11	52fff9	Snow / Ice	Sníh/ led

Tab. 9 – Hodnoty pásma SCL dat Sentinel-2 (Zdroj: Sentinel-2 MSI 2021)

Testováno bylo využití jednotlivých hodnot pásma SCL, klíčová byla hodnota 7 – nízká pravděpodobnost výskytu oblačnosti nebo neklasifikováno. Na obrázcích 14 a 15 lze pozorovat odmaskování oblačnosti, a to porovnáním původního snímku a maskami bez zařazení hodnoty 7 a s jejím zařazením. V první sérii, kde se jedná o okolí jezera Abaya, lze vidět velké rozdíly. Pokud není do oblačné masky zařazena hodnota 7, velké oblasti v okolí jezera jsou odstraněny (na obrázku zvýrazněno červenou barvou). O oblačnost se však nejedná. Některá místa zástavby byla odstraněna jako součást oblačnosti, pokud hodnota 7 nebyla zařazena do masky. Na druhé sérii lze pozorovat odmaskování části přístávací dráhy letiště v Awasse, což je nežádoucí.





Obr. 14 a 15 – Porovnání původního obrazu a masek oblačnosti bez zařazení hodnoty 7 pásma SCL a s jejím zařazením (červená místa na obrázcích jsou ta odmaskovaná) (Zdroj: vlastní)

Kód¹ pro tvorbu masky, která zanechává vše kromě oblačnosti, jak bylo popsáno výše:

```
var cloudmask = function (image) {
  var SCL = image.select('SCL');
  var QA60 = image.select('QA60');
  var SCLmask = SCL.eq(1).or(SCL.eq(2)).or(SCL.eq(4)).or(SCL.eq(5))
    .or(SCL.eq(6)).or(SCL.eq(7));
  var QA60mask = QA60.eq(0);
  return image.updateMask(SCLmask).updateMask(QA60mask);
};
```

Příkaz pro aplikaci masky oblačnosti 1/6:

```
var withoutclouds1 = collection1.map(cloudmask);
```

Snímky s odmaskovanou oblačností byly spojeny do mozaiky funkcí `mosaic()`. Snímky se navzájem překrývají a tato funkce dává na sebe snímky podle toho, jak jsou za sebou řazeny v kolekci (poslední nahoru). Výslednou hodnotou mozaiky v místech, kde se snímky překrývají, je tedy hodnota z posledního snímku v seznamu.

¹ Pozn. Tato část kódu byla vytvořena ve spolupráci s Mgr. Danielem Palubou.

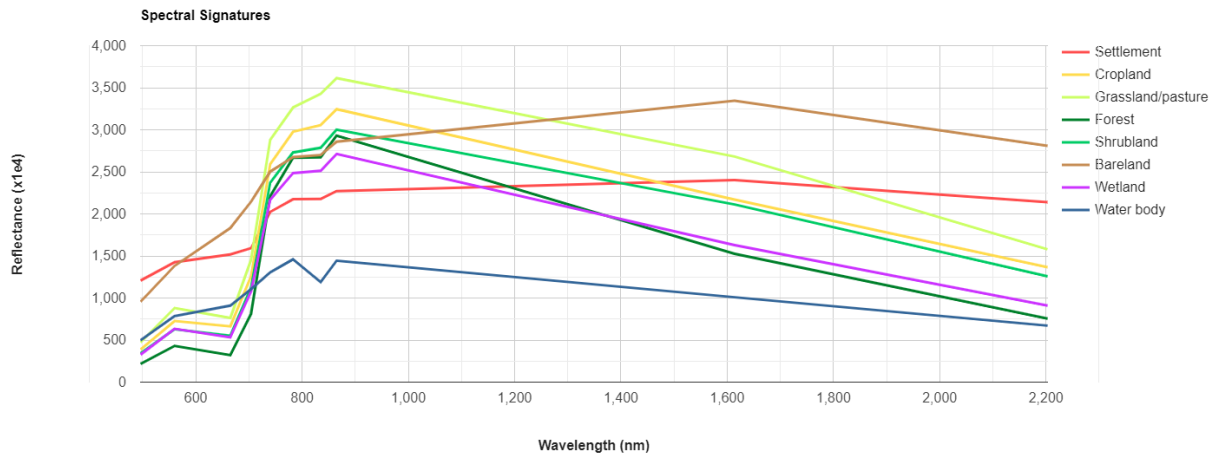
Kód pro tvorbu mozaiky:

```
var mosaicFirst = ee.ImageCollection([
  withoutclouds1.first(),
  withoutclouds2.first(),
  withoutclouds3.first(),
  withoutclouds4.first(),
  withoutclouds5.first(),
  withoutclouds6.first(),
]).mosaic();
```

3.3.4 Spektrální charakteristiky

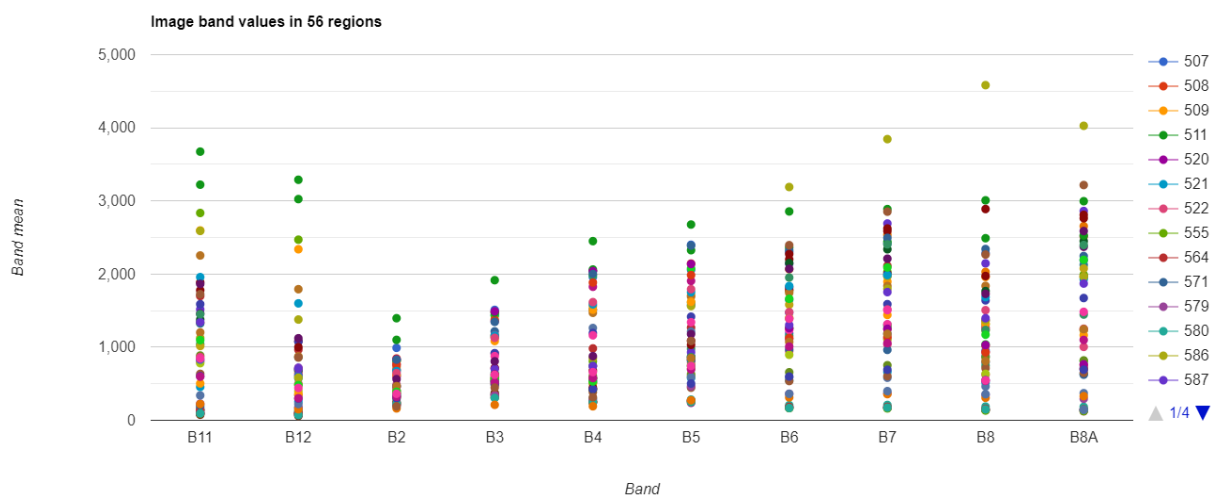
Pro klasifikaci bylo zvoleno 10 spektrálních pásem snímků ze dne 19. 11. 2020 Sentinel-2, a to B2-B8, B8A, B11 a B12, tedy RGB, Red Edge 1, 2, 3 a 4, NIR, SWIR 1 a 2. Spektrální pásma, která měla jiné než zadané prostorové rozlišení (10 m/px), byla v GEE automaticky převzorkována metodou Nearest Neighbor. V tomto prostředí byla vytvořena série grafů, která ukázala spektrální příznaky trénovacích dat a zároveň separabilitu jednotlivých tříd. Byl vytvořen graf průměrné odrazivosti klasifikačních tříd ve vlnových délkách 400 nm až 2200 nm, grafy odlehklých hodnot v datech každé klasifikační třídy zvlášť a grafy vztahu odrazivosti dat v pásmech NIR a Red pro každou třídu. Kód pro tvorbu grafů v GEE je v příloze 2.

Z grafu průměrných hodnot odrazivosti tříd land cover (viz Obr. 16) je patrné, že 5 tříd má relativně podobnou křivku (louka/pastvina, plodiny, křoviny, les, mokřad), tedy s potenciálem, že bude nesnadné je od sebe rozlišit během klasifikace. Třída zástavba a vodní plocha se nejvíce liší od ostatních. Třída vodní plocha má podobnou průměrnou hodnotu odrazivosti jako louka/pastvina v pásmu Blue, Green a Red, ale v ostatních pásmech se liší. V pásmu NIR, kde lze pozorovat překrývající se hodnoty tříd les, křoviny a holá půda, je hodnota odrazivosti vodních ploch nižší o 15 %. Třídy les, křoviny a holá půda lze potenciálně lépe rozlišit v pásmech SWIR 1, případně Green a Red. Relativně podobnou křivku mají třídy křoviny, plodiny a louka/pastvina, nejvíce se liší v pásmu Red Edge 3, NIR a Red Edge 4. V těchto pásmech mají podobnou průměrnou hodnotu odrazivosti třídy les a křoviny, liší se však o 7 % v pásmu SWIR 1.



Obr. 16 – Průměrná spektrální odrazivost trénovacích dat všech tříd land cover v datech Sentinel-2
(Zdroj: vlastní)

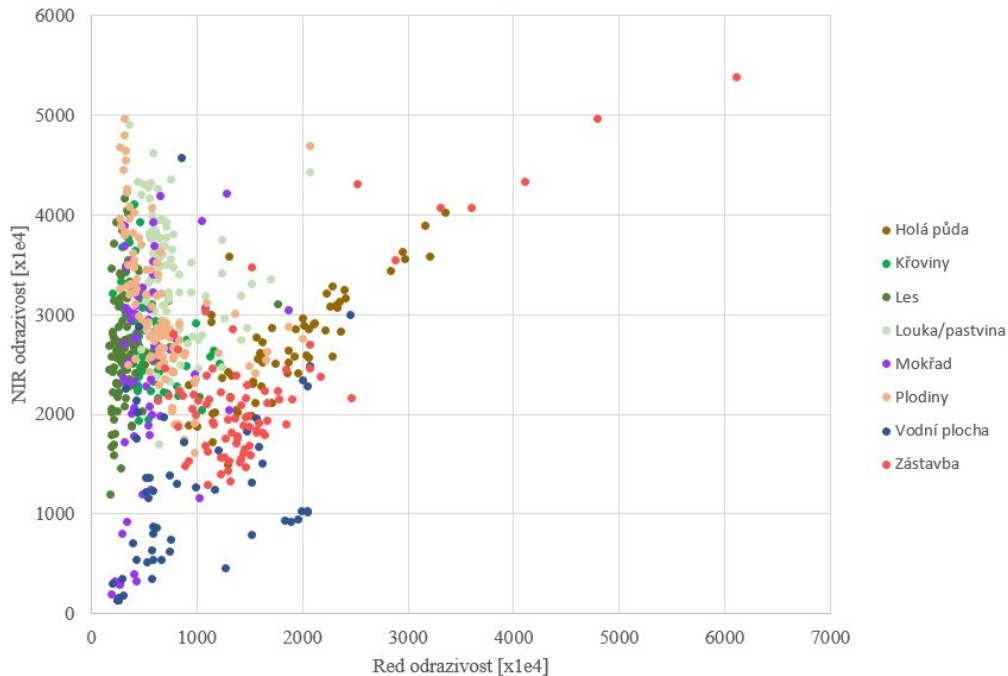
Grafy odlehlých hodnot (příklad třídy zástavba viz Obr. 17) ukazují na body/lokality, které se v jednotlivých pásmech příznakově velmi liší od ostatních, a je tedy na zvážení jejich odstranění z datasetu. Konkrétní místa s odlehlými hodnotami jednotlivých tříd byla přezkoumána na podkladu aktuálních snímků s velmi vysokým rozlišení v softwaru GEP a odstraněna z trénovacího datasetu, jednalo-li se o tzv. mixed-pixels (více tříd v jednom pixelu). Dále byla trénovací data prostorově i početně upravována tak, aby bylo docíleno co nejvyšší celkové přesnosti klasifikace. Důraz byl kladen na to, aby počet trénovacích bodů byl dostatečný a umístění shodné s vypovídající hodnotou pro danou třídu.



Obr. 17 – Spektrální příznaky: Odlehlé hodnoty trénovacích dat v třídě vodní plochy (Zdroj: vlastní)

Grafy vztahu odrazivosti v pásmech NIR a Red (propojení těchto grafů viz Obr. 18) ukazují separabilitu jednotlivých tříd. Na obrázku 18 lze vidět, že trénovací body třídy les se seskupují do linie kolem hodnoty 3 % v pásmu Red a 11–42 % v pásmu NIR. Třída

louka/pastviny se shromažďuje ve stejném okolí pásma NIR, ale 7–20 % v pásmu Red. Třída vodní plochy se nachází v nízkých hodnotách obou pásem. Třídy zástavba a holá půda procházejí středem grafu, tedy hodnoty Red a NIR jednotlivých bodů jsou si podobné. Třídy křoviny a mokřad jsou relativně podobné, a tak potenciálně špatně oddělitelné během klasifikace.



Obr. 18 – Vztah odrazivosti pásem NIR a Red trénovacích dat všech tříd land cover (Zdroj: vlastní, vytvořeno v programu Excel)

Kód pro přiřazení spektrálních příznaků z klasifikovaného obrazu jednotlivým bodům trénovací množiny dat:

```
var training = image
  .select(bands)
  .sampleRegions({
    collection: training_data,
    properties: [label],
    scale: scale});
print(training, 'training');
```

3.3.5 Volba parametrů klasifikační metody

Ve výše zmíněných výzkumných studiích (např. Noi, Kappas 2017, Ok, Akar, Gungor 2012) bylo detailně řešeno nastavení parametrů v metodě RF, např. počet stromů a počet

spektrálních pásem, které vstupují do každého stromu. V GEE je možné nastavit parametrů šest: jedná se o počet stromů (*numberOfTrees*), počet proměnných v každém stromě (*variablesPerSplit*), minimální populace listu (*minLeafPopulation*), část prvků v každém stromě (*bagFraction*), počet pater každého stromu (*maxNodes*) a náhodnost (*seed*) (viz Tab. 10). Metodu RF lze najít pod příkazem `ee.Classifier.smileRandomForest()`.

Název		Datový typ	Výchozí nastavení	Funkce
numberOfTrees (povinné)	Počet stromů v lese	Integer	-	Určuje počet stromů v náhodném lese
variables-PerSplit	Počet proměnných v každém stromě	Integer	Odmocnina z počtu vstupních proměnných	Určuje počet proměnných (pásem), které vstupují do každého stromu
minLeaf-Population	Minimální populace listu	Integer	1	Ukončuje větvení pod touto prahovou hodnotou
bagFraction	Část prvků v každém stromě	Float	0,5	Určuje část (frakci) vstupních trénovacích dat, která vstupují do každého stromu
maxNodes	Větvení	Integer	Bez limitu	Určuje maximální počet pater každého stromu, tedy počet rozhodovacích pravidel
seed	Náhodnost	Integer	0	Určuje typ náhodnosti

Tab. 10 – Parametry metody Random Forest v prostředí GEE (Zdroj: GEE 2021, upraveno)

V této studii bylo testováno nastavení 4 nejdůležitějších parametrů – *numberOfTrees*, *variablesPerSplit*, *minLeafPopulation* a *bagFraction*. Nastavení zbylých dvou parametrů se netestovalo, jelikož parametr *maxNodes* je užitečný, pokud je pracováno s velkým množstvím dat a je potřeba proces urychlit, a parametr *seed* volí pouze jinou náhodnost. Postupně byly vybrané parametry testovány a byla sledována celková přesnost klasifikace. Bylo zvoleno 10 kombinací 4 parametrů (viz Tab. 11), a to na základě rešerše (Noi, Kappas 2017, Ok, Akar, Gungor 2012) a vlastního testování. Pro první parametr (*numberOfTrees*) byly zvoleny hodnoty 80, 100, 120 a 500 stromů. Pro parametr počtu proměnných (= spektrálních pásem) byly zvoleny hodnoty 3 a 4. Hodnota 3 jako výchozí, protože výchozí nastavení parametru odpovídá odmocnině z celkového počtu, tedy $\sqrt{10} \approx 3$. Parametr větvení byl na základě testování ponechán na výchozí hodnotě 1. Vyšší hodnoty by byly užitečné pouze v případě většího množství dat ke zpracování a čas výpočtů by se tak urychlil. Poslední parametr (*bagFraction*) udává, jak velká část trénovací množiny dat bude vstupovat do klasifikace. Testovány byly hodnoty 0,4 a 0,6, tedy 40 a 60 %.

	numberOfTrees	variablesPerSplit	minLeafPopulation	bagFraction
Kombinace 1	80	3	1	0,4
Kombinace 2	80	4	1	0,4
Kombinace 3	80	4	1	0,6
Kombinace 4	100	3	1	0,4
Kombinace 5	100	4	1	0,4
Kombinace 6	100	4	1	0,6
Kombinace 7	120	3	1	0,4
Kombinace 8	120	4	1	0,4
Kombinace 9	120	4	1	0,6
Kombinace 10	500	4	1	0,4

Tab. 11 – Testované kombinace parametrů metody Random Forest (Zdroj: vlastní)

Kód pro natrénování klasifikátoru RF s kombinací parametrů 5 a provedení klasifikace:

```
var classifier = ee.Classifier.smileRandomForest(100,4,1,0.4).train({
  features: training,
  classProperty: label,
  inputProperties: bands});

var classified = image.classify(classifier);
```

3.3.6 Filtrace obrazu

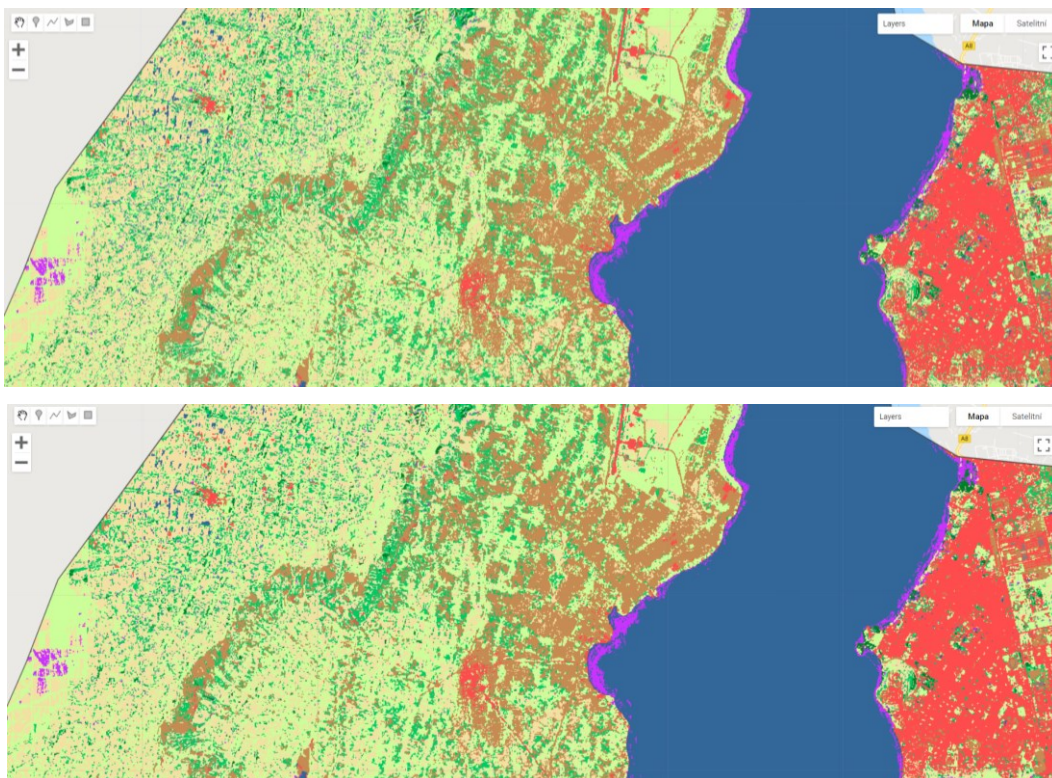
Obraz byl klasifikován s rozlišením 10 m a následně filtrován tak, aby byly odstraněny osamocené pixely a jiné šумы. Byl použit většinový filtr, který zachovává hodnotu pixelu, jsou-li v okolí min 2 pixely se stejnou třídou, a nejsou-li, třída je zvolena na základě většiny. Na základě požadavků koncových uživatelů byl dále obraz filtrován tak, aby jednotlivé prvky byly čitelné na mapě v měřítku 1 : 100 000. Podle Toblerova pravidla (Nagi 2010) má mít pixel na takové mapě rozlišení 50 m, tedy 5 x 5 pixelů Sentinel-2 (10 m/px). To odpovídá 25pixelovým shlukům. MMU je 2500 m². Obraz byl sjednocen do tzv. shluků (clusters) využitím segmentace s nastavením 5 x 5 pixelů, kompaktnost 1 a konektivita 4 (viz Obr. 19). Tak byly vytvořeny 2 varianty výsledku klasifikace. První varianta s odstraněnými osamocenými pixely a s redukcí šumu, u které byla hodnocena přesnost (viz kapitola 3.3.8), a druhá varianta vytvořená dle požadavků koncových uživatelů (mapa v měřítku 1 : 100 000).

Kód² pro filtraci obrazu – odstranění osamocených pixelů:

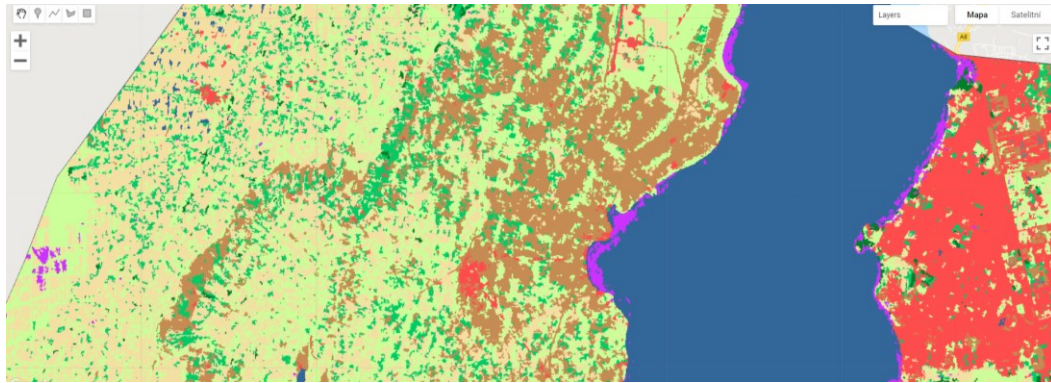
```
var patchsize = classified.connectedPixelCount(2, false);  
var majorityfilter = classified.focal_mode();  
var filtered = classified.where(patchsize.lt(2),majorityfilter);
```

Kód pro filtraci obrazu – vytvoření shluků:

```
var composite = filtered.clip(region);  
var snic = ee.Algorithms.Image.Segmentation.SNIC({  
  image: composite,  
  size: 5,  
  compactness: 1,  
  connectivity: 4});  
var clusters = snic.select('clusters');  
var smoothed = filtered.addBands(clusters);  
var clustered = smoothed.reduceConnectedComponents({  
  reducer: ee.Reducer.mode(),  
  labelBand: 'clusters'});
```



² Pozn. Tato část kódu byla vytvořena ve spolupráci s Mgr. Danielem Palubou.



Obr. 19 – Filtrace obrazu (shora obraz bez filtrace, obraz s odtraněnými osamocenými pixely, obraz s vytvořenými shluky) (Zdroj: vlastní)

3.3.7 Vizualizace

Výslednou mapu krajinného pokryvu lze zobrazit v GEE na mapě v dolním okně. K zobrazení byl využit příkaz `Map.addLayer()`, jehož jeden parametr je barevná paleta.

Kód pro vizualizaci výsledného obrazu:

```
var igbpPalette = [
  'ff4d4d', // settlement
  'ffdb4d', // cropland
  'ccff66', // grassland/pasture
  '00802b', // forest
  '00cc66', // shrubland
  'c68c53', // bareland
  'cc33ff', // wetland
  '336699' // waterbody];
```

```
Map.addLayer(classified.clip(region), {palette: igbpPalette, min: 0, max:
7}, 'classified image');
```

GEE však není vhodné prostředí pro tvorbu mapy k tisku či prezentaci. Vhodnější je výsledek exportovat do GeoTIFFu a mapu dotvořit v některém z GIS softwarů.

Během exportu se musel vyřešit problém, kdy pro klasifikaci byly třídy číslovány od 0 (0 = zástavba), při exportu se však jako třída 0 označí i pixely „no data“, což odpovídá odmaskované oblačnosti a okolí zájmového území, pokud se nejedná o pravoúhlé území. Oblačnost a zástavba by se zobrazovaly stejnou barvou. Proto předtím, než byl výsledek

klasifikace exportován a upravován v softwaru QGIS do podoby mapy k tisku, byly původní hodnoty tříd klasifikace zvýšeny o řád, aby se nevyskytovala 0, a to funkcí `.remap`:

```
var clustered_remapped = clustered.remap(  
    [0, 1, 2, 3, 4, 5, 6, 7],  
    [1, 2, 3, 4, 5, 6, 7, 8]);
```

A dále exportován příkazem:

```
Export.image.toDrive({  
    image: clustered_remapped,  
    description: "Landcover_classification_Sidama_remapped_clustered",  
    region: region,  
    scale: scale,  
    maxPixels: 1e13,  
    crs: 'EPSG: 32637'});
```

Takto byly exportovány obě varianty výsledku klasifikace. Výsledné GeoTIFFy byly otevřeny v softwaru QGIS a upraveny do podoby map. Druhé variantě, která byla vytvořena dle požadavků koncových uživatelů (mapa v měřítku 1 : 100 000), byl kromě legendy a grafického měřítko přidán také nadpis a tiráž. Měřítko mapy bylo zvoleno podle zadání projektu, tedy 1 : 100 000. Velikost papíru pro tisk odpovídá 1400 x 1200 mm, což je 1,68 rozměru největšího běžného formátu A0. Barvy jednotlivých klasifikačních tříd byly vybrány tak, aby asociovaly daný krajinný pokryv (viz Obr. 20).

	holá půda		oblačnost
	křoviny		plodiny
	les		vodní plocha
	louka/pastvina		zástavba
	mokřad		

Obr. 20 – Legenda (Zdroj: vlastní)

3.3.8 Hodnocení přesnosti a kontrolní množina dat

Kontrolní množina dat je pro přesnost klasifikace zásadní proměnná. Pro rozdělení in-situ dat na trénovací a kontrolní množiny nebylo nasbíráno in-situ dat dostatečné množství. Z toho důvodu byla vytvořena přidaná kontrolní množina dat. Kontrolní body byly v počtu 638 vygenerovány náhodně v GEE, a to na základě rozlohy každé třídy výsledné klasifikace

(viz Tab. 12). Kód pro generování náhodných stratifikovaně rozmístěných kontrolních bodů viz příloha 2.

Třída	Počet vygenerovaných kontrolních bodů
Holá půda	45
Křoviny	78
Les	174
Louka/pastvina	96
Mokřad	21
Plodiny	197
Vodní plocha	14
Zástavba	13
Celkem	638

Tab. 12 – Počet vygenerovaných kontrolních bodů (Zdroj: vlastní)

Tato kolekce bodů byla exportována ve formátu SHP na Google disk, stažena a naimportována do programu QGIS. Zde byly přidány atributy *ID*, *valid*. Soubor byl uložen ve formátu KML, přičemž atribut *ID* byl označen jako *Name* a *valid* jako *description*, kam bude vypisováno číslo skutečné třídy daného bodu. Data byla validována v programu GEP na podkladě aktuálních snímků s velmi vysokým rozlišením Maxar a Airbus. Snímky byly na většině území dostupné z podzimu 2020, což odpovídá době sběru trénovacích dat (10 až 11/2020) i době nasnímání klasifikovaných snímků Sentinel-2 (19. 11. 2020). Některé odlehlejší oblasti však nebyly v této době dostupné. V takovém případě bylo k jednotlivým třídám přistupováno individuálně. Třídy, které by se neměly měnit dynamicky (během sezóny), jako les, zástavba, velké vodní plochy, velké plochy holé půdy (pustiny), byly validovány pomocí nejaktuálnějších dostupných leteckých snímků, většinou z ledna a února 2021. V případě proměnných tříd jako je třída plodiny a mokřady byly zhlédnuty všechny dostupné snímky ze stejného ročního období, tedy měsíce říjen, listopad, prosinec 2019, 2018, 2017. Pakliže se mokřad na místě vyskytoval v předchozích letech, bralo se v potaz pravidlo, že pravděpodobně se zde vyskytoval i v listopadu 2020. Nebylo-li možné s jistotou rozhodnout o třídě daného pixelu, validační bod byl z datasetu odstraněn.

Kód pro přiřazení hodnot klasifikace kontrolním bodům:

```
var validation = filtered
.sampleRegions({
collection: valid_data,
properties: [label_valid],
```

```
scale: scale,  
geometries: true  
});
```

První varianta výstupu klasifikace, tedy obraz filtrovaný většinovým filtrem pro odstranění osamocených pixelů a šumu, byla podrobena hodnocení přesnosti klasifikace. Pro hodnocení přesnosti klasifikace byla využita matice chyb a kappa index, do jejichž výpočtů vstupovala přidaná kontrolní množina dat s atributem *valid*, jehož hodnota odpovídala skutečné třídě daného pixelu.

Kód pro výpočet matice chyb a kappa indexu:

```
var testAccuracy = validation.errorMatrix(label_valid, 'classification');  
print('Validation error matrix: ', testAccuracy);  
print('Validation overall accuracy: ', testAccuracy.accuracy());  
print('Kappa index: ', testAccuracy.kappa());
```

Proces klasifikace, od natrénování klasifikátoru po výpočet chybové matice, resp. přesnosti klasifikace, byl testován a detailně vyhodnocen. Byly zkoumány jednotlivé kombinace nastavení nejdůležitějších parametrů a jejich výsledné přesnosti. Cílem bylo dosažení přesnosti vyšší než 80 %. Pro dosažení této přesnosti byly určovány jak vhodné kombinace parametrů klasifikátoru, tak i vhodná sada trénovacích dat. Celý skript klasifikace v GEE je v příloze 2.

3.3.9 Klasifikace land cover s využitím dat PlanetScope ve vybraných kebelích

Vytvořený algoritmus pro klasifikaci krajinného pokryvu regionu Sidama pro tvorbu mapy v měřítku 1 : 100 000 s 8 klasifikačními třídami nad snímky Sentinel-2 byl aplikován na lokální území 2 kebelí v regionu Sidama pro tvorbu mapy v měřítku 1 : 10 000 nad snímky PlanetScope. Z uživatelského hlediska bylo žádoucí, aby metodika klasifikace vyvinutá nad daty Sentinel-2 (popsaná v předchozích kapitolách) byla totožná či velmi podobná. Klasifikováno bylo území 2 kebelí, Chancho a Dangora Morocho, o rozloze 8 a 10 km².

Legenda a minimální mapovací jednotka

Legenda byla zachována stejná jako pro celý region Sidama, aby bylo možné výsledné obrazy porovnat, tedy 8 klasifikačních tříd, a to zástavba (settlement), plodiny (cropland), louka/pastvina (grassland/pasture), les (forest), křoviny (shrubland), holá půda (bareland), mokřad (wetland) a vodní plocha (water body). Stanovení MMU bylo založeno na stejných

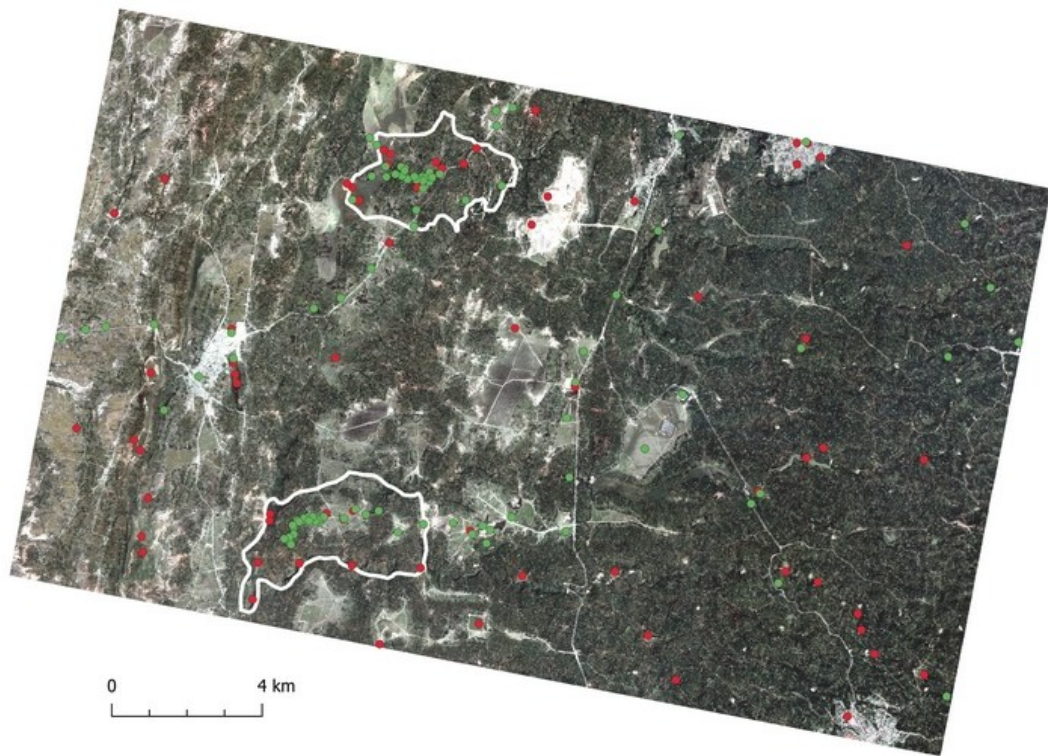
pravidlech, tedy 1) na požadavcích koncových uživatelů, a sice, že mapové výstupy mají být v měřítku 1 : 10 000, a 2) na vlastnostech použitých dat PlanetScope (3 m/px). Klasifikace proběhla s rozlišením 3 m a následně byl obraz filtrován tak, aby byly odstraněny osamocené pixely pomocí většinového filtru. Na základě požadavků koncových uživatelů byl obraz dále filtrován pomocí segmentace tak, aby byly zachovány prvky o velikosti 27 m², což odpovídá Toblerovu pravidlu (Nagi 2010). Tak byly vytvořeny 2 varianty výstupů.

Trénovací data

Pro klasifikaci kebelí byl využit trénovací dataset vycházející z trénovacího datasetu pro klasifikaci regionu Sidama. Přestože obě kebele byly detailně mapovány v terénu, počet i rozmístění sesbíraných dat bylo nedostatečné pro účely kvalitní klasifikace (viz Tab. 13 a Obr. 21). Trénovací dataset byl tedy doplněn v softwaru GEP sběrem nad daty s velmi vysokým rozlišením Maxar a Airbus ze stejného ročního období, jako byl sběr in-situ dat (podzim 2020), tedy 12/2019, 9/2020 a 11/2020. Celkový počet trénovacích dat vstupujících do klasifikace byl 143 bodů pro obě kebele.

Třída	Počet trénovacích bodů z in-situ výzkumu	Celkový počet trénovacích bodů s využitím snímků
Holá půda	4	14
Křoviny	3	10
Les	24	24
Louka/pastvina	20	29
Mokřad	6	7
Plodiny	25	29
Vodní plocha	2	13
Zástavba	8	17
Celkem	92	143

Tab. 13 – Počet trénovacích dat pro klasifikaci kebelí (Zdroj: vlastní)



Obr. 21 – Rozmístění trénovacích bodů pro klasifikaci kebelí (červeně přidaná data) (Zdroj podkladových dat: Planet 2021)

Předzpracování snímků

Pro klasifikace krajinného pokryvu 2 kebelí byly vybrány 2 snímky PlanetScope 3B AnalyticMS SR ze dne 18. 11. 2020, které na sebe v pásu navazovaly a pokrývaly zcela obě území. Data PlanetScope verze Analytic byla již atmosféricky korigována, a to modelem 6SV2.1 (Planet 2021). Snímky byly transformovány do souřadnicového systému WGS 84 (EPSG: 4326) v softwaru QGIS funkcí *Deformovat*. Dále byly snímky nahrány do GEE do složky `Assets` a do skriptu postupně příkazy: `ee.Image('users/valcharda/DP/planetscope_527')`. Snímky byly zobrazeny v RGB a byly zobrazeny také hranice kebelí, které byly do GEE nahrány ve formátu SHP. Vybrané snímky byly zcela bezoblačné a byly mozaikovány stejným způsobem jako snímky Sentinel-2, tedy funkcí `mosaic()`. Dále byla v okolí zájmových území vytvořena obalová zóna se šířkou 200 m.

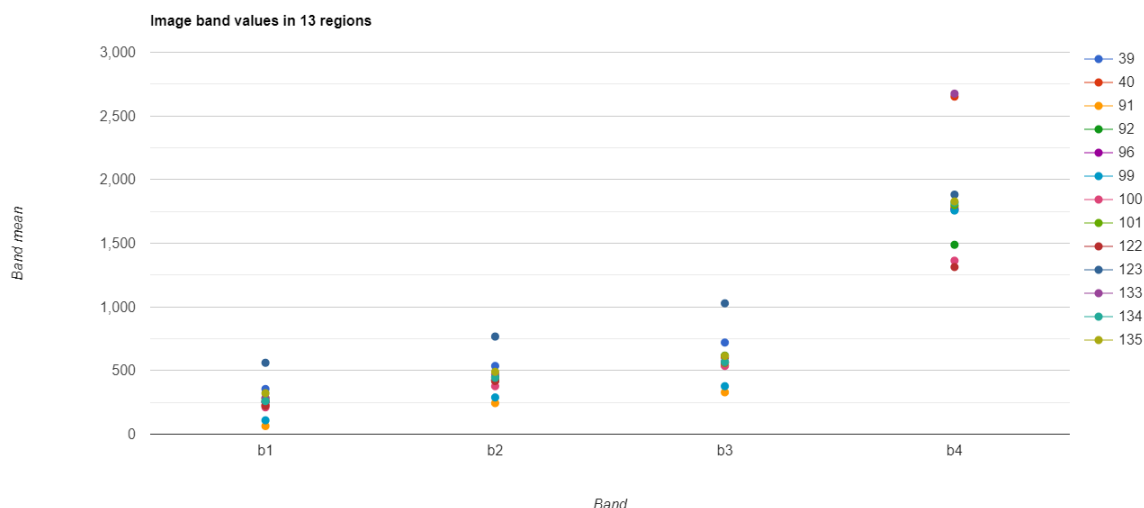
Kód pro tvorbu obalové zóny:

```
var kebele_geom = ee.Geometry.MultiPolygon
    (kebele.geometry().coordinates());
var buffered_kebele = kebele_geom.buffer({'distance': 200});
```

Spektrální charakteristiky

Pro identifikaci spektrálních příznaků jednotlivých tříd land cover byla vytvořena série grafů nad 2 snímky PlanetScope ze dne 18. 11. 2020. Byly vytvořeny grafy odlehlých hodnot v datech každé klasifikační třídy zvlášť a grafy vztahu odrazivosti dat v pásmech NIR a Red pro každou třídu land cover. Jelikož spektrální rozlišení dat PlanetScope není tak vysoké jako v případě Sentinel-2, pro klasifikaci dat PlanetScope bylo testováno využití dvou indexů, a to NDVI (Normalized Difference Vegetation Index) a NDWI (Normalized Difference Water Index), přesné vzorce viz Sentinel-Hub (2021a, 2021b).

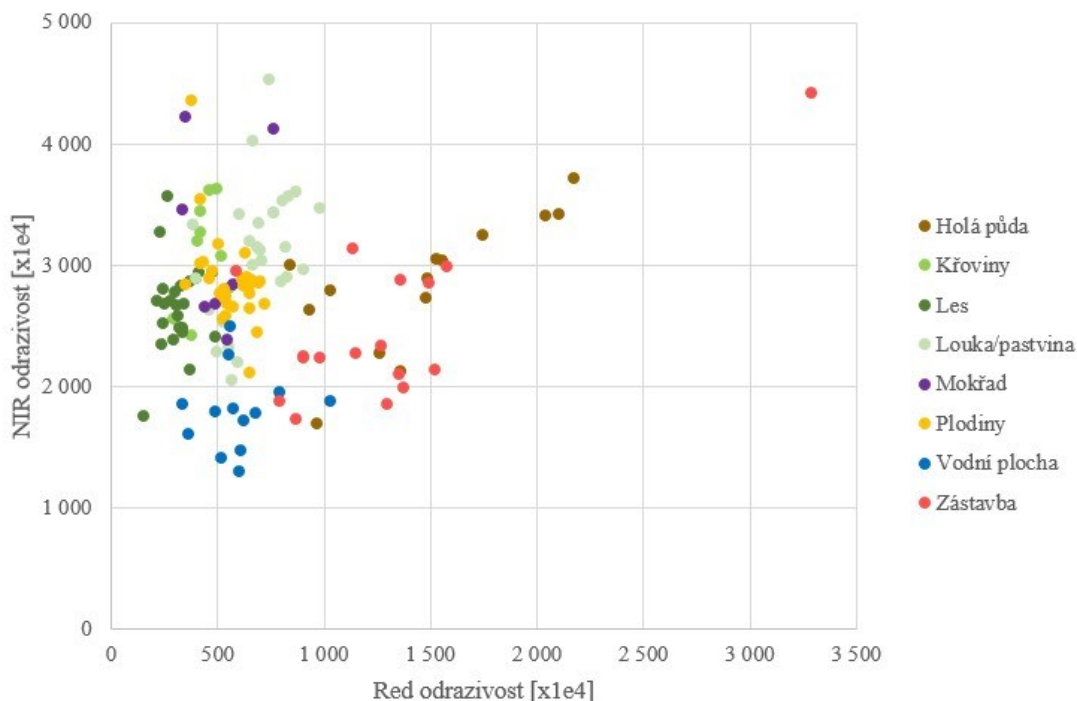
Grafy odlehlých hodnot (příklad třídy vodní plocha viz Obr. 22) ukazují na body, které se v jednotlivých pásmech příznakově velmi liší od ostatních. Tato místa byla přezkoumána na podkladu aktuálních snímků s velmi vysokým rozlišením v softwaru GEP a odstraněna z trénovacího datasetu se stejnými pravidly jako u trénovacího datasetu pro klasifikaci snímků Sentinel-2.



Obr. 22 – Spektrální příznaky: Odlehlé hodnoty trénovacích dat pro klasifikaci kebelů v třídě vodní plochy (Zdroj: vlastní)

Grafy vztahu odrazivosti dat v pásmech NIR a Red pro každou třídu land cover (propojení těchto grafů viz Obr. 23) ukazují separabilitu tříd. Rozmístění jednotlivých tříd

v prostoru je podobné jako v grafu nad snímky Sentinel-2. Třída zástavba a holá půda se vyskytují v okolí diagonály grafu, tedy hodnoty odrazivosti v pásmu Red jsou podobné pásmu NIR. Třída les se nachází v grafu nejvíce vlevo, tedy v nejnižších hodnotách pásma Red. Třída vodní plocha se nachází v rohu grafu, tedy odpovídá nejnižším hodnotám obou spektrálních pásem. Podobný vztah pásem Red a NIR vykazují třídy plodiny a louka/pastvina.



Obr. 23 – Vztah odrazivosti pásem NIR a Red trénovacích dat všech tříd land cover pro klasifikaci kebelí (Zdroj: vlastní, vytvořeno v programu Excel)

Parametry metody Random Forest, filtrace a vizualizace

Bylo testováno stejné nastavení parametrů, jako pro klasifikaci snímků Sentinel-2, tedy 10 kombinací parametrů *numberOfTrees*, *variablesPerSplit*, *minLeafPopulation* a *bagFraction*. Obraz byl klasifikován s rozlišením 3 m a následně filtrován se stejným metodickým postupem. Tedy byly vytvořeny 2 varianty výstupů: první s využitím většinového filtru a zachováním největší podrobnosti (3 m/px) a druhá dle požadavků koncových uživatelů (mapy v měřítku 1 : 10 000). Hodnoty výsledných obrazů byly zvýšeny o řád, aby byla odstraněna hodnota 0. Dále byly exportovány ve formátu GeoTIFF a v programu QGIS upraveny do podoby map krajinného pokryvu kebelí Chanco a Dangora Morocho v roce 2020 (viz příloha 3). Byla zachována barevnost pro porovnání s mapou krajinného pokryvu regionu Sidama.

Hodnocení přesnosti

Přesnost klasifikace snímků PlanetScope byla hodnocena výpočtem matice chyb a kappa indexu, a to porovnáním kontrolní množiny dat a klasifikovaného obrazu první varianty. Kontrolní množina dat byla náhodně generována v prostředí GEE zvlášť pro každou kebeli. Bylo vygenerováno náhodně 100 bodů pro každou kebeli a počet bodů v jednotlivých třídách byl stanoven na základě rozlohy klasifikovaných tříd (viz Tab. 14). Kontrolní body byly určovány a validovány v softwaru GEP nad snímky s velmi vysokým rozlišením Maxar a Airbus s datem pořízení blízkým datu pořízení dat PlanetScope a proběhlému terénnímu šetření. Skript pro klasifikaci dat PlanetScope je v příloze 2.

Kebele	Chancho	Dangora Morocho
Holá půda	1	1
Křoviny	13	10
Les	21	18
Louka/pastvina	15	19
Mokřad	6	3
Plodiny	43	46
Vodní plocha	0	1
Zástavba	1	1
Celkem	100	99

Tab. 14 – Počet vygenerovaných kontrolních bodů pro klasifikaci kebelí (Zdroj: vlastní)

4 Výsledky

4.1 Výsledky klasifikace dat Sentinel-2

Pro klasifikaci dat Sentinel-2 bylo vybráno 6 snímků ze dne 19. 11. 2020. Pro tuto klasifikaci land cover regionu Sidama v roce 2020 byl vytvořen algoritmus v prostředí GEE s využitím klasifikátoru RF. Bylo testováno 10 kombinací 4 nejdůležitějších parametrů klasifikátoru. Kombinace, která vedla k nejvyšší celkové přesnosti, byla vybrána jako výsledná, byly vytvořeny výsledné mapy a pomocí grafů vyjádřeno zastoupení jednotlivých tříd land cover.

Hodnocení přesnosti klasifikace dat Sentinel-2

Klasifikace krajinného pokryvu regionu Sidama na podkladě dat Sentinel-2 proběhla v prostředí GEE pomocí klasifikátoru RF a zde byla také spočítána přesnost klasifikace, a to pomocí chybové matice a kappa indexu. Bylo provedeno testování nastavení vybraných parametrů klasifikátoru v 10 kombinacích, jak bylo popsáno v kapitole 3.3.5. Celková přesnost a kappa index jsou dokumentovány v Tab. 15.

	NumberOfTrees	VariablesPerSplit	MinLeafPopulation	BagFraction	Celková přesnost [%]	Kappa index
Kombinace 1	80	3	1	0,4	79,21	0,734
Kombinace 2	80	4	1	0,4	81,11	0,758
Kombinace 3	80	4	1	0,6	79,21	0,734
Kombinace 4	100	3	1	0,4	79,52	0,738
Kombinace 5	100	4	1	0,4	84,10	0,797
Kombinace 6	100	4	1	0,6	78,89	0,730
Kombinace 7	120	3	1	0,4	80,63	0,753
Kombinace 8	120	4	1	0,4	82,70	0,778
Kombinace 9	120	4	1	0,6	78,10	0,720
Kombinace 10	500	4	1	0,4	79,52	0,737

Tab. 15 – Výsledná celková přesnost a kappa index klasifikace regionu Sidama pro 10 kombinací parametrů metody Random Forest (Zdroj: vlastní)

Nejvyšší celková přesnost vyšla pro kombinaci 5, tedy s nastavením 100, 4, 1 a 0,4 (počet stromů, počet spektrálních pásem, populace listu a část trénovacích dat), a to 84,1 % a kappa index 0,797, což se dá považovat za uspokojivý výsledek a naplňuje to předem stanovené cíle. Pokud se podíváme blíže na dosažené výsledky, lze tvrdit, že celková přesnost klasifikace není zcela závislá na počtu stromů, tzn. přesnost se nezvyšovala se zvyšujícím se počtem stromů. V kombinaci 10 byla nastavena extrémní hodnota, a to 500 stromů. Ukázalo se, že tak vysoký počet nejen zpomaluje proces výpočtu, ale ani celková přesnost není lepší než u nižšího počtu stromů. Klasifikace se 100 stromy vyšla s vyšší přesností než s 80 stromy, ale 120 stromů ukázalo opět nižší přesnost. Testováno bylo 40 a 60 % trénovacích dat (*BagFraction*), která vstupovala do každého stromu z lesa. V případě této práce se prokázalo, že přesněji vychází klasifikace s využitím pouze 40 % dat. V kombinaci s ostatními parametry vyšla pro počet spektrálních pásem (*VariablesPerSplit*), která vstupovala do každého stromu lesa, lépe hodnota 4 než 3.

Na základě provedeného testování klasifikace krajinného pokryvu regionu Sidama v roce 2020 byla vybrána kombinace parametrů s nejvyšší celkovou přesností, tedy kombinace parametrů 5 (viz Obr. 24).



Obr. 24 – Srovnání výřezu výsledné mapy krajinného pokryvu a snímku Sentinel-2 (Zdroj satelitního snímku: Sentinel-2 MSI 2021)

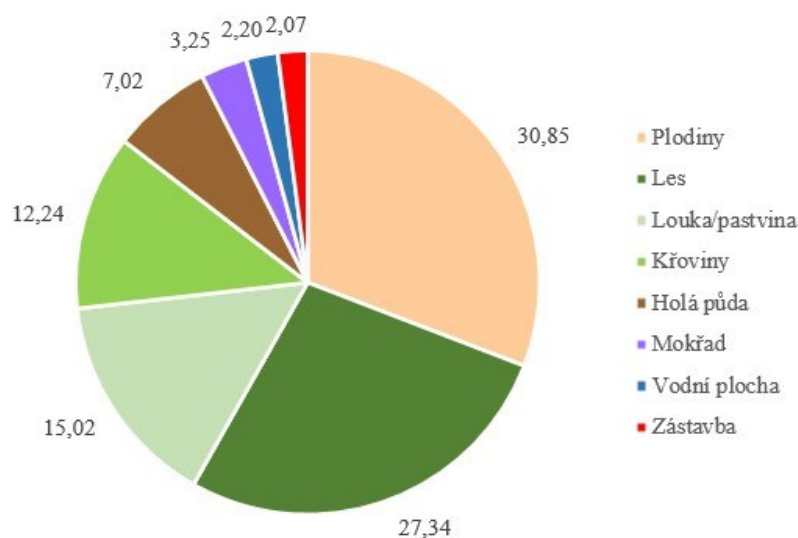
V tabulce matice chyb této klasifikace (viz Tab. 16) lze číst, které třídy byly nejproblématictější a které naopak nejméně problematické. Počty bodů na hlavní diagonále odpovídají správně zařazeným. Ostatní hodnoty jsou chyby, a to chyby z nesprávného zařazení (řádky, tedy uživatelská přesnost) a z opomenutí (sloupce, tedy zpracovatelská přesnost) (Červená 2018). Průměrně vyšla zpracovatelská přesnost (82,22 %) lépe než uživatelská (78,62 %). Nejnižší uživatelskou i zpracovatelskou přesnost měla třída mokřad (37,50 % a 42,86 %). Dále třída křoviny dosáhla pouze 51,14 % uživatelské a 70,31 % zpracovatelské přesnosti. Všechny ostatní třídy vykazovaly vyšší než 84% přesnost obojího typu.

	Zástavba	Plodiny	Louka/ pastvina	Les	Křoviny	Holá půda	Mokřad	Vodní plocha	Celkem	Uživatel. přesnost
Zástavba	9	0	0	0	1	0	1	0	11	81,82
Plodiny	0	165	8	3	14	2	3	0	195	84,62
Louka/pastvina	0	4	83	0	0	0	2	0	89	93,26
Les	0	1	1	166	1	0	0	0	169	98,22
Křoviny	0	20	3	19	45	0	1	0	88	51,14
Holá půda	0	3	1	0	1	43	0	0	48	89,58
Mokřad	0	1	2	4	2	0	6	1	16	37,50
Vodní plocha	0	0	0	0	0	0	1	13	14	92,86
Celkem	9	194	98	192	64	45	14	14	630	
Zpracovatel. přesnost	100,00	85,05	84,69	86,46	70,31	95,56	42,86	92,86		84,13

Tab. 16 – Chybová matice klasifikace krajinného pokryvu regionu Sidama, kombinace 5 (Zdroj: vlastní)

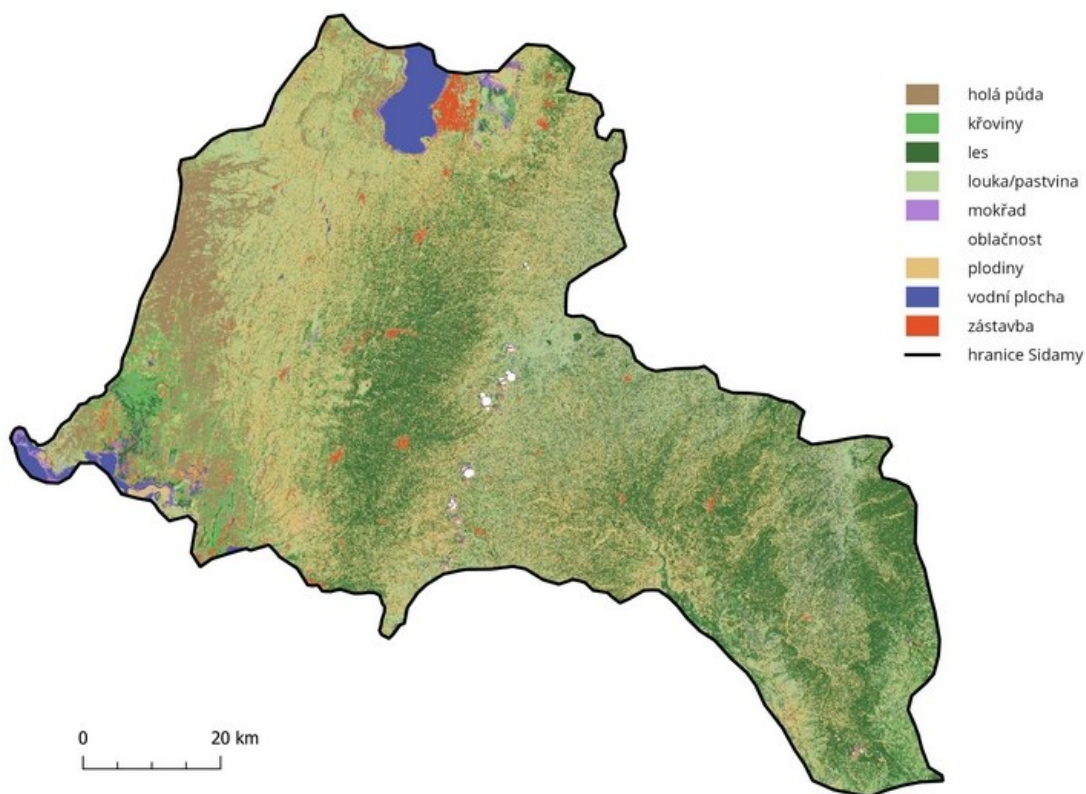
Výsledná klasifikace dat Sentinel-2

Ve zvolené kombinaci parametrů proběhla klasifikace land cover. Obr. 25 dokumentuje rozlohu jednotlivých tříd land cover. Dle dosažených výsledků je zřejmé, že nejvíce byla v regionu v roce 2020 zastoupena třída plodiny (30,85 %) a lesy (27,34 %). Nejméně území regionu zabírala zástavba (2,07 %) a vodní plochy (2,20 %).



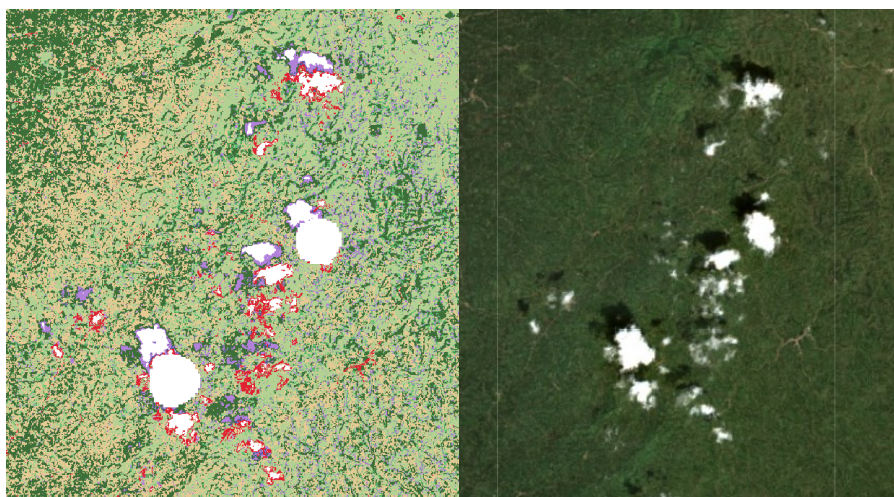
Obr. 25 – Zastoupení jednotlivých tříd land cover v regionu Sidama v roce 2020 [%] (Zdroj: vlastní, vytvořeno v programu Excel)

Výsledná klasifikace je prezentována pomocí mapy, a to ve dvou variantách. První varianta s pixelem o velikost 10 m (viz Obr. 26) a druhá varianta přizpůsobená požadavkům koncových uživatelů, tedy mapa v měřítku 1 : 100 000 (viz Příloha 3). Výsledný obraz klasifikace byl podle požadavků koncových uživatelů generalizován na MMU 2500 m². Pokud by byl obraz pouze převzorkován na 50m pixel, zanikly by silnice či vodní toky, které jsou úzké, ale dlouhé. Použitím metody shluků byly tyto prvky zachovány. Na obou mapách lze pozorovat prostorové rozložení jednotlivých tříd land cover. V západní části zájmového území se nachází převážně holá půda a křoviny. Centrální částí vede zalesněná oblast a také se zde nachází několik větších měst (zástavby). Od severu jsou to města Awassa, Yirgalem Irgalem, Aleta Wondo a Chucko, která jsou propojena zpevněnou cestou. Východní část je horská oblast a je z velké části zalesněná. Výrazné jsou na mapě dvě modré oblasti, které odpovídají jezerům Awassa na severu a Abaya na jihozápadě. Větší oblasti mokřadů se nacházejí v okolí těchto jezer. Největší zastoupení plodin najdeme v nížinné oblasti na severozápadě.



Obr. 26 – Výsledná mapa krajinného pokryvu regionu Sidama v roce 2020 (Zdroj: vlastní)

Při pohledu na výsledné mapy se jeví jako problémové lokality místa s oblačností. Použitý algoritmus bohužel neodstranil některé stíny mraků nebo jemná oblaka (viz Obr. 27). Klasifikační algoritmus klasifikoval tato místa jako zástavba (červeně) nebo mokřad (fialově). Tyto kategorie tak byly nadhodnoceny.



Obr. 27 – Příklad oblasti stínů mraků klasifikovaných jako třída zástavba a mokřad v porovnání se snímkem Sentinel-2 (Zdroj satelitního snímku: Sentinel-2 MSI 2021)

Další problémová místa se vyskytují na jihozápadě zájmového území, kde se nachází říční delta a v jejím okolí vznikají mokřady. Zdejší řeka Bilate byla v některých místech klasifikována jako holá půda, pravděpodobně proto, že byla zanesena erodovaným materiálem, který s sebou nese do jezera Abaya. Severně od jezera Abaya bylo větší území klasifikováno jako zástavba, přestože se tam velké zastavěné území nenachází. Dle náhledů na dostupné snímky lze usuzovat, že se zde vyskytuje pustina s holou půdou a křovinami. Diskutabilní se též jeví klasifikace vysokého množství mokřadů malých rozměrů na východě zájmového území. Mokřady jsou v této oblasti sezónním fenoménem, ohledně jejich výskytu ve sledovaném čase by bylo potřeba provést terénní šetření.

4.2 Výsledky klasifikace dat PlanetScope

Pro klasifikaci dat PlanetScope byly vybrány 2 snímky ze dne 18. 11. 2020. Klasifikace dat PlanetScope pro určení krajinného pokryvu kebelí byla založena na již vytvořeném algoritmu pro klasifikaci dat Sentinel-2. Taktéž nastavení kombinací parametrů bylo inspirováno již otestovaným algoritmem. Navíc bylo využito vegetačních indexů NDVI a NDWI.

Hodnocení přesnosti klasifikace dat PlanetScope

Nejdříve bylo provedeno 10 iterací s kombinacemi parametrů RF bez využití indexů NDVI a NDWI. Celková přesnost klasifikace kebele Chancho se pohybovala od 57 do 64 % a byla průměrně o 6,60 % nižší než s využitím indexů. U klasifikace kebele Dangora Morocho vyšla celková přesnost mezi 59,60 a 66,67 % a byla průměrně o 5,35 % nižší. Extrémní rozdíl byl 10,10 % celkové přesnosti klasifikace kebele Dangora Morocho s parametry kombinace 2 s využitím a bez využití indexů.

Dále bylo testováno 10 kombinací parametrů RF s využitím indexů NDVI a NDWI a výsledky jsou dokumentovány v tab. 17. Nejvyšší celková přesnost klasifikace kebele Chancho vyšla 68 % a kappa index 0,569, a to s kombinací 1 (80, 3, 1 a 40 %). V případě kebele Dangora Morocho vyšla nejvyšší přesnost 73,74 % a kappa index 0,601, a to s kombinací 5 (100, 4, 1 a 40 %). Průměrná celková přesnost klasifikace kebele Chancho vyšla o 2 % nižší a kappa index o 0,022 nižší než v případě kebele Dangora Morocho. Kombinace parametrů metody RF č. 5 vyšla pro klasifikaci kebele Dangora Morocho s nejvyšší celkovou přesností, stejně jako pro klasifikaci regionu Sidama nad snímky Sentinel-2. U klasifikace kebele Chancho se předpokládá přenositelnost algoritmu

s přednastavenými parametry nepotvrdil, nejvyšší přesnosti dosáhla kombinace 1. Kombinace 5 dosáhla v případě této lokality 66 %.

Parametry	Kebele				Chancho		Dangora Morocho	
	NumberOfTrees	VariablesPerSplit	MinLeafPopulation	BagFraction	Celková přesnost [%]	Kappa index	Celková přesnost [%]	Kappa index
Kombinace 1	80	3	1	0,4	68,00	0,569	67,68	0,511
Kombinace 2	80	4	1	0,4	67,00	0,561	73,74	0,598
Kombinace 3	80	4	1	0,6	64,00	0,526	64,65	0,488
Kombinace 4	100	3	1	0,4	66,00	0,544	67,68	0,514
Kombinace 5	100	4	1	0,4	66,00	0,545	73,74	0,601
Kombinace 6	100	4	1	0,6	65,00	0,537	64,65	0,503
Kombinace 7	120	3	1	0,4	67,00	0,556	69,70	0,544
Kombinace 8	120	4	1	0,4	67,00	0,557	68,69	0,520
Kombinace 9	120	4	1	0,6	64,00	0,524	65,66	0,499
Kombinace 10	500	4	1	0,4	69,00	0,586	66,67	0,509

Tab. 17 - Výsledná celková přesnost a kappa index klasifikace kebelí pro 10 kombinací parametrů metody Random Forest (Zdroj: vlastní)

Jako výsledné klasifikace krajinného pokryvu kebelí v roce 2020 byly vybrány kombinace parametrů, které dosáhly nejvyšší celkové přesnosti, tedy kombinace 1 pro kebeli Chancho a kombinace 5 pro kebeli Dangora Morocho s využitím indexů. Pro tyto klasifikace byly vygenerovány matice chyb (viz Tab. 18 a 19). Zpracovatelská i uživatelská přesnost obou klasifikací vyšla velmi nevyváženě u každé třídy land cover. Obecně se nejvíce problematické třídy ukázaly být holá půda, mokřad, vodní plocha a zástavba. Uživatelská přesnost třídy zástavby vyšla pro klasifikaci kebele Chancho 100 %, ale pro klasifikaci kebele Dangora Morocho 0 %. Nejvyšší uživatelskou přesnost vykazovala v obou případech třída les (88,89 a 100 %), zpracovatelskou u kebele Chancho třída křoviny (84,62 %), v druhém případě třída plodiny (86 %).

	Zástavba	Plodiny	Louka/ pastvina	Les	Křoviny	Holá půda	Mokřad	Vodní plocha	Celkem	Uživatelská přesnost
Zástavba	1	0	0	0	0	0	0	0	1	100,00
Plodiny	0	30	2	4	2	0	0	0	38	78,95
Louka/pastvina	1	6	9	0	0	0	0	0	16	56,25
Les	0	0	0	16	0	0	2	0	18	88,89
Křoviny	0	5	0	4	11	0	1	0	21	52,38
Holá půda	1	0	1	0	0	0	0	0	2	0,00
Mokřad	0	2	0	0	0	0	0	0	2	0,00
Vodní plocha	0	1	0	0	0	0	0	1	2	50,00
Celkem	3	44	12	24	13	0	3	1	100	
Zpracovatelská přesnost	33,33	68,18	75,00	66,67	84,62	0,00	0,00	100,00		68,00

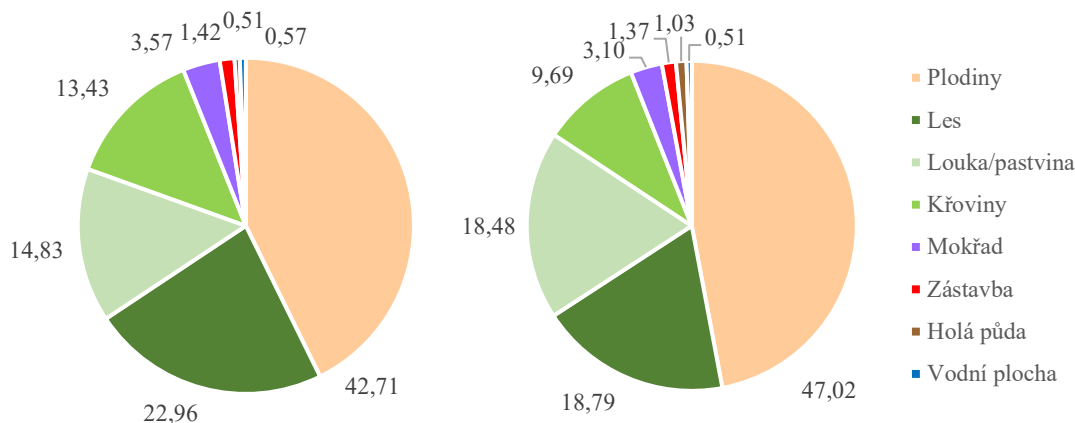
Tab. 18 – Chybová matice klasifikace kebele Chancho, kombinace 1 (Zdroj: vlastní)

	Zástavba	Plodiny	Louka/ pastvina	Les	Křoviny	Holá půda	Mokřad	Vodní plocha	Celkem	Uživatelská přesnost
Zástavba	0	0	0	0	0	0	0	0	0	0,00
Plodiny	0	43	5	4	2	0	1	0	55	78,18
Louka/pastvina	0	2	11	1	0	0	0	0	14	78,57
Les	0	0	0	12	0	0	0	0	12	100,00
Křoviny	0	5	1	3	6	0	0	0	15	40,00
Holá půda	1	0	0	0	0	1	0	1	3	33,33
Mokřad	0	0	0	0	0	0	0	0	0	0,00
Vodní plocha	0	0	0	0	0	0	0	0	0	0,00
Celkem	1	50	17	20	8	1	1	1	99	
Zpracovatelská přesnost	0,00	86,00	64,71	60,00	75,00	0,00	0,00	100,00		73,74

Tab. 19 – Chybová matice klasifikace kebele Dangora Morocho, kombinace 5 (Zdroj: vlastní)

Výsledné klasifikace dat PlanetScope

Na základě výsledků klasifikace dat PlanetScope byla spočítána rozloha jednotlivých tříd land cover (viz Obr. 28). Největší zastoupení měla v obou kebelech třída plodiny (47,02 % Dangora Morocho a 42,71 % Chancho) a dále les (22,96 a 18,79 %). Na území kebele Dangora Morocho byl v roce 2020 větší podíl luk a pastvin (18,48 %) než na území kebele Chancho (14,83 %) a také podíl holé půdy (1,03 a 0,51 %). Relativně nízké rozlohy čítala v obou případech třída zástavba (1,42 % Chancho a 1,37 % Dangora Morocho). Nejmenší podíl obou území zabírala třída vodní plocha (0,57 % Chancho a 0,51 % Dangora Morocho).



Obr. 28 - Zastoupení jednotlivých tříd land cover v kebeli Chanco a Dangora Morocho v roce 2020 [%] (Zdroj: vlastní, vytvořeno v programu Excel)

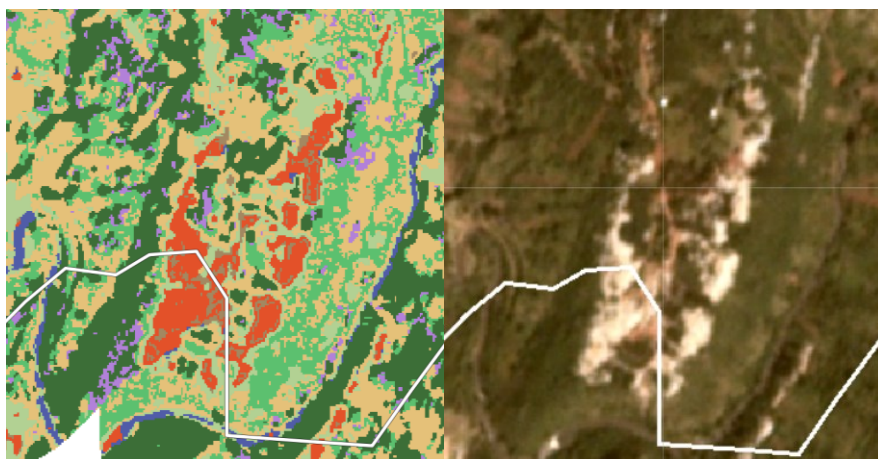
Výsledné mapy land cover byly vytvořeny v softwaru QGIS, a to opět ve dvou variantách. První varianta byla ponechána s nejvyšším rozlišením (3 m/px) (viz Obr. 29) a druhá varianta byla filtrována pomocí shluků tak, aby míra podrobnosti odpovídala požadavkům koncových uživatelů, tedy mapy v měřítku 1 : 10 000 (viz Příloha 3). Z mapových výstupů je zřejmé, že krajinný pokryv obou kebelí tvořila mozaika plodin, křovin, luk a lesů. V západní části kebele Chanco se nacházely dominantně plodiny a louky/pastviny přecházející v centrální části v mozaiku plodin, luk/pastvin a lesů. V jihovýchodní části protéká řeka Gidabo. Z mapy je zřejmé, že území protíná zpevněná komunikace vedoucí napříč celým zájmovým územím, na kterou se váže zástavba na severovýchodě území. Jedná se o jižní část vesnice Dale. Ve východní části jsou též hojně zastoupeny mokřady. V centrální části kebele Dangora Morocho se nachází rozlehlá oblast luk a pastvin. Centrální částí od východu na západ vede částečně zpevněná komunikace propojující zastavěné oblasti. Od západního cípu na jih teče řeka Gidabo spojující se v jižním cípu s řekou Kola tekoucí od východního cípu. V údolích řek rostla hustší lesní vegetace.



*Obr. 29 – Výsledné mapy krajinného pokryvu kebelí Chancho a Dangora Morocho v roce 2020
(Zdroj: vlastní)*

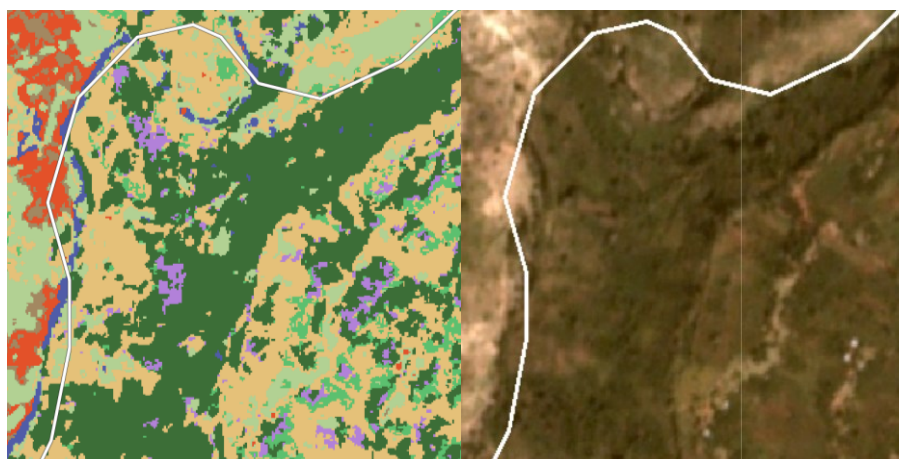
Kterak bylo výše uvedeno v souvislosti s hodnocením přesnost, problematické byly především holá půda, mokřad, vodní plocha a zástavba. Klasifikace vodních ploch byla ztížena znečištěním vodních toků splaveninami a organickou složkou, čímž vznikala záměna s loukami/pastvinami či holou půdou. Mokřady byly zaměňovány s loukami/pastvinami i

s plodinami. Tato třída byla tak nadhodnocena (viz Obr. 28). Skutečnou polohu mokřadů je kvůli sezónní proměnlivosti této třídy nesehnat ověřit. Problematická se též jeví záměna třídy zástavba a holá půda. Na západě a na jihovýchodě kebele Chancho se nachází větší území vyklasifikovaná jako zástavba (viz Obr. 30), jedná se však ve skutečnosti o degradovanou plochu, tedy třídu holá půda.



Obr. 30 – Srovnání výřezu výsledné mapy krajinného pokryvu a snímku PlanetScope (Zdroj satelitního snímku: Planet 2021)

V neposlední řadě nutno zmínit záměnu křovin a lesa s příkladem na severozápadě kebele Dangora Morocho. Zde se nachází větší území klasifikované jako les (viz Obr. 31), jedná se však ve skutečnosti o mozaiku křovin a plodin. Díky času snímání a terénním vlastnostem bylo toto území zastíněno.

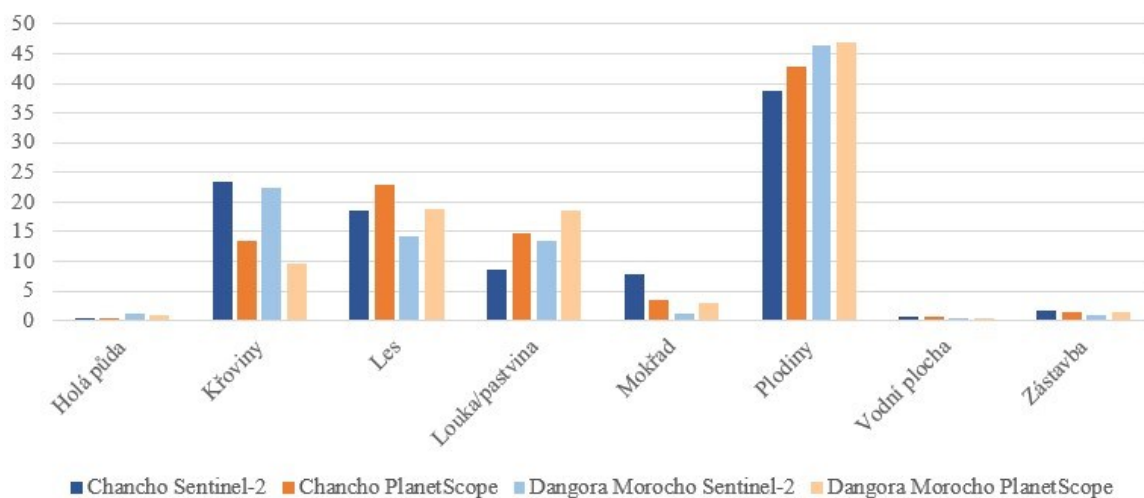


Obr. 31 – Srovnání výřezu výsledné mapy krajinného pokryvu a snímku PlanetScope II (Zdroj satelitního snímku: Planet 2021)

4.3 Porovnání výsledků klasifikace dat Sentinel-2 a PlanetScope

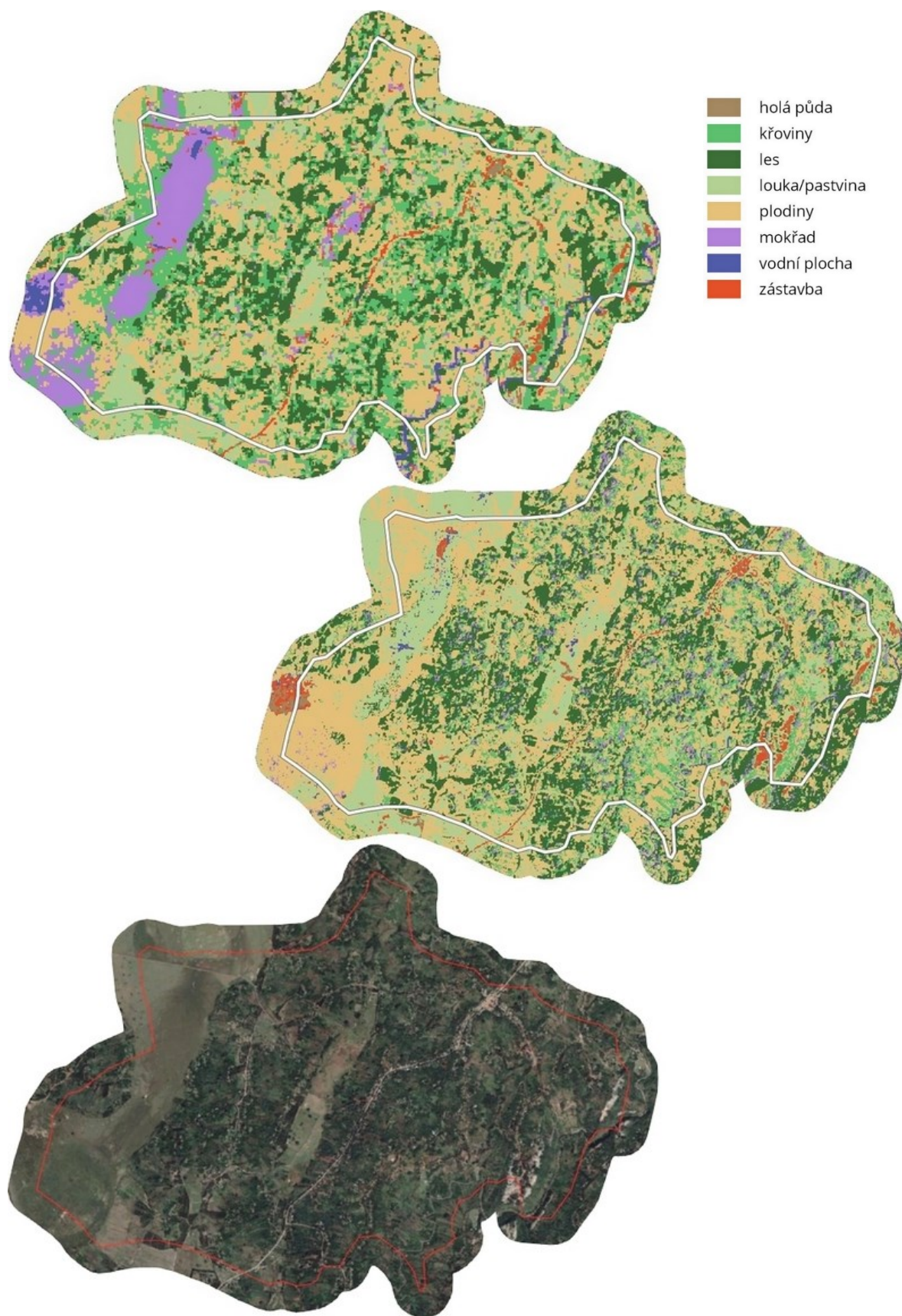
V rámci diplomové práce proběhla klasifikace regionu Sidama nad daty Sentinel-2 s prostorovým rozlišením 10 m/px pořízenými 19. 11. 2020. S využitím sestaveného algoritmu v prostředí GEE proběhla také klasifikace kebelí Chancho a Dangora Morocho nad daty PlanetScope s prostorovým rozlišením 3 m/px pořízenými dne 18. 11. 2020. Kebele se nacházejí na území regionu Sidama a výsledky těchto klasifikací lze porovnat. Pro porovnání výsledků byly zvoleny klasifikace land cover s nejvyšší celkovou přesností, a to klasifikace dat Sentinel-2 s kombinací parametrů 5 s dosaženou přesností 84,1 %. Dále klasifikace dat PlanetScope pro kebeli Chancho s kombinací parametrů 1 s dosaženou přesností 68 % a pro kebeli Dangora Morocho s kombinací parametrů 5 s přesností 73,74 %.

Byly spočítány rozlohy jednotlivých tříd land cover na území kebelí, a to z výsledků klasifikace nad daty Sentinel-2 a PlanetScope (viz Obr. 32). V případě kebele Chancho lze největší rozdíly v zastoupení tříd land cover pozorovat u třídy křoviny (rozdíl 10,01 %) a louka/pastvina (rozdíl 6,19 %). Dále zastoupení třídy mokřad vyšlo nad daty PlanetScope poloviční, a to 3,57 % oproti 7,84 % nad daty Sentinel-2. Velmi podobné hodnoty vyšly pro třídu holá půda, vodní plocha a zástavba. V případě kebele Dangora Morocho je největší rozdíl opět v třídě křoviny (rozdíl 12,58 %) a louka/pastvina (rozdíl 5,04 %). V tomto případě je trend klasifikace třídy mokřad opačný, tedy zastoupení třídy nad daty Sentinel-2 je méně než poloviční (1,09 %) oproti klasifikaci nad daty PlanetScope (3,10 %). Podíl tříd holá půda, vodní plocha a zástavba vyšel podobně pro oba typy dat.



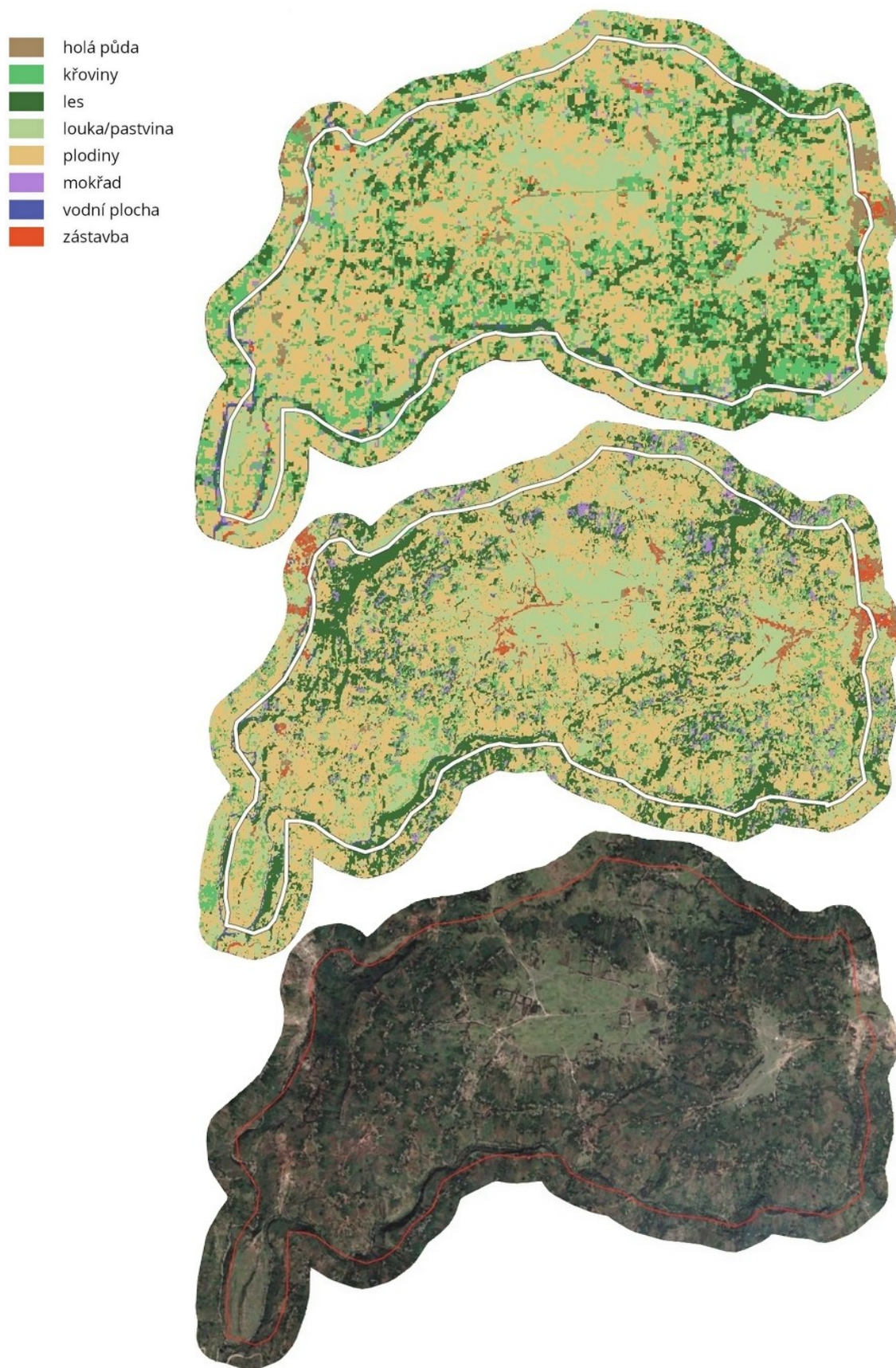
Obr. 32 – Porovnání zastoupení jednotlivých tříd land cover v kebelích Chancho a Dangora Morocho klasifikací nad daty Sentinel-2 a PlanetScope [%] (Zdroj: vlastní, vytvořeno v programu Excel)

K výsledným mapám klasifikace kebelí nad daty PlanetScope byly přidány výřezy těchto lokalit klasifikovány nad daty Sentinel-2 a snímky Maxar s velmi vysokým rozlišením pro představu o detailním krajinném pokryvu území kebelí. Na první sérii map (viz Obr. 33) lze pozorovat velké rozdíly klasifikačních tříd v západní oblasti kebele Chancho, kde se nachází rozlehlé území luk/pastvin a z velké části je podmáčené (= mokřad). Podle definice tříd by se tato oblast měla klasifikovat jako mokřad. Tento fenomén se podařilo lépe zachytit pomocí klasifikace Sentinel-2. Další výraznější rozdíl se nachází západně od zmíněné oblasti. Dle snímku Maxar a dalších pomocných zdrojů by se mělo jednat o holou půdu obklopenou mokřadem, která byla v případě dat Sentinel-2 klasifikována jako vodní plocha obklopená mokřadem a v případě dat PlanetScope jako holá půda se zástavbou. Shodně v obou klasifikacích se podařilo zachytit zpevněnou komunikaci vedoucí z východu na jih. Další rozdíly v klasifikaci lze pozorovat na jihovýchodě, kudy protéká řeka Gidabo. Na mapě s využitím dat Sentinel-2 je průběh řeky zřetelný, na mapě založené na snímcích PlanetScope není průběh řeky na mnohých místech zcela patrný. Klasifikace severovýchodní části kebele Chancho se jeví jako úspěšná u obou typů dat. Nachází se zde část vesnice Dale. Z tohoto hlediska se jeví správné mísení pixelů třídy zástavba a holá půda, neboť v malých vesnicích bývají cesty nezpevněné (= holá půda). Holou půdu se nepodařilo zcela správně určit na jihovýchodě území, kde se nacházejí degradované plochy, které byly v obou případech dat chybně klasifikovány jako zástavba.



Obr. 33 – Výsledné mapy krajinného pokryvu kebele Chancho s využitím dat Sentinel-2, PlanetScope ve srovnání se snímek velmi vysokého rozlišení (Zdroj satelitního snímku: Maxar 2021)

Na druhé sérii map (viz Obr. 34) lze pozorovat rozdíly tříd land cover v případě kebele Dangora Morocho. Velkého rozdílu v klasifikacích si lze všimnout v severozápadní části území, kde bylo velké území nad snímky PlanetScope klasifikováno jako les. Ve skutečnosti je tato oblast mozaikou křovin a plodin, což přesněji zachytila klasifikace nad daty Sentinel-2. Na jihozápadě kebele se stékají dvě významné řeky, Gidabo ze severu a Kola z východu. Pro klasifikace obou typů dat bylo zachycení jejich správného průběhu nesnadným úkolem a modrá barva, značící vodní tok, se objevuje přerušovaně. Na levém břehu řeky Kola, těsně předtím, než se vlévá do řeky Gidabo, se nachází prostorově významná mozaika křovin a plodin. Podobně jako v případě severozápadní části území klasifikace land cover se jeví kvalitnější nad daty Sentinel-2. Klasifikace dat PlanetScope vykazuje tuto oblast chybně jako les. Další problémová oblast je území na východě kebele, kde se nachází degradovaná plocha (= holá půda). Na rozdíl od klasifikace nad snímky Sentinel-2 se klasifikátoru s využitím snímků PlanetScope nepodařilo správně určit tuto třídu jako holá půda, ale toto území z velké části chybně detekoval jako zástavbu. Diskutabilním fenoménem je velké množství roztroušených mokřadů malých rozměrů klasifikovaných nad snímky PlanetScope. Správnost určení těchto ploch bohužel nelze ověřit ani s použitím snímků Maxar, pro ověření by byl nutný terénní výzkum.



Obr. 34 - Výsledné mapy krajinného pokryvu kebele Dangora Morocho s využitím dat Sentinel-2, PlanetScope ve srovnání se snímkem velmi vysokého rozlišení (Zdroj satelitního snímku: Maxar 2021)

5 Diskuze

Cílem diplomové práce bylo dosažení co nejpřesnější klasifikace krajinného pokryvu (min. 80 %) dle požadavků koncových uživatelů vybraných území Etiopie v roce 2020 nad snímky Sentinel-2 a PlanetScope v cloudovém prostředí GEE za pomoci metod strojového učení v různých měřítkových úrovních. Vypracovaný klasifikační algoritmus byl tak testován na dvou typech dat (Sentinel-2 a PlanetScope) s jiným prostorovým, spektrálním i časovým rozlišením na jiné měřítkové úrovni. Výsledné klasifikace byly podrobeny vyhodnocení přesnosti a vzájemně porovnány. Pro naplnění těchto cílů bylo nutno vyřešit několik zásadních výzkumně-metodických úkolů. V následujících odstavcích budou diskutovány jednotlivé zásadní aspekty zvolených řešení a dosažených výsledků. Pokud je to možné, zvolené metody a výsledky jsou porovnány s podobně orientovanými studii.

Prvotním metodickým úkolem byla definice legendy. Legenda byla zvolena na základě literární rešerše (viz kapitola 2.3), požadavků koncových uživatelů, a především na základě vlastního terénního průzkumu zájmového území. Bylo definováno 8 tříd land cover, a to zástavba, plodiny, louka/pastvina, les, křoviny, holá půda, mokřad a vodní plocha. Pečlivou definici si zasluhovala zástavba s ohledem na vysokou heterogenitu a charakter. Problematická se jevila kruhová obydlí, tzv. tukuly, a to z důvodu malých rozměrů a charakteru stavebního materiálu.

Dalším významným úkolem byla volba vhodných družicových snímků pro klasifikace a dále metody maskování oblačnosti, protože vybraný region Sidama se nachází v monzunové oblasti, kde se střídá období dešťů a sucha. Období vydatných dešťů trvá od června do poloviny září, od října do února panuje sucho, od března do května je období občasných dešťů. V období sucha není oblačnost problémem, ale vegetace je suchá, plodiny sklizené a není tak optimální klasifikovat krajinný pokryv. Mekuyie, Jordaan, Melka (2018) doporučují klasifikovat snímky z přelomového období sucha a dešťů. V přechodovém období byl realizován in-situ výzkum, a to konkrétně během října a listopadu, kdy vegetace byla stále vzrostlá a zelená po období dešťů a zároveň ještě nepanovalo silné sucho. Snímky Sentinel-2 a PlanetScope tak byly vybrány s ohledem na proběhlý in-situ výzkum. Vybrány byly snímky Sentinel-2 nasnímané dne 19. 11. 2020 a snímky PlanetScope ze dne 18. 11. 2020. Byla snaha vybrat snímky obou typů dat ze stejného dne. Společnost Planet Lab Inc. má v nabídce snímky PlanetScope z každého dne v roce 2020. Zájmové území dvou kebelí však byla

v měsíci listopadu snímána brzy ráno (v 5:42 hod) a v údolí řek byl stín. Pro klasifikaci byly tedy zvoleny snímky pořízené o den dříve než u dat Sentinel-2, aby byla zajištěna lepší kvalita snímků. Snímky pořízené 18. 11. byly snímány v 7:35 hod.

Důležitým úkolem byla volba a otestování vhodné metody maskování oblačnosti. Výhodou bylo to, že vybrané snímky PlanetScope byly zcela bezoblačné. Na vybraných snímcích Sentinel-2 byl však úzký pás oblačnosti. Pro masku oblačnosti byla využita pásma SCL a QA60 datasetu Sentinel-2. Maska byla tvořena hodnotou 0 pásma QA60 a hodnotami 1, 2, 4, 5, 6 a 7 pásma SCL. Tyto hodnoty ve snímku zůstaly, vše ostatní bylo odstraněno. Tak bylo docíleno odstranění oblačnosti z mozaiky snímků Sentinel-2. V průběhu klasifikace se však ukázalo, že některé jemné mraky a stíny nebyly zcela odstraněny a byly problematickým aspektem během klasifikace. Tato infikovaná místa byla chybně klasifikována (např. chybná klasifikace mokřadů či zástavby viz Obr. 27).

Pro vývoj algoritmu klasifikace krajinného pokryvu bylo využito cloudové prostředí GEE, které umožňuje rychle a účelně zpracovávat data dálkového průzkumu. Prostředí se ukázalo být velmi uživatelsky přívětivé, výpočty rychlé a proces testování různých nastavení efektivní, neboť bylo vysoce operativní měnit jednotlivé parametry a jiné vstupy klasifikace s rychle dostupným a viditelným výsledkem. Velkou výhodou platformy je bohatá nabídka datasetů, čímž je vyřešen jeden z hlavních limitů desktopového prostředí, tedy stahování a ukládání velkého objemu rastrových dat. Nevýhodou platformy je nutnost psaní skriptu, což může být pro některé uživatele překážkou. Na druhou stranu lze skripty či jejich části sdílet s dalšími uživateli GEE, což evokuje rychlý vývoj inovativních metod a způsobů zpracování dat v dálkovém průzkumu Země. Autorka diplomové práce před jejím zpracováním v Java Scriptu nepracovala a dá se říct, že se s tímto prostředím naučila efektivně pracovat v relativně krátkém čase. Jako výsledek byly sestaveny skripty pro klasifikaci krajinného pokryvu regionu Sidama, kebelí Chanco a Dangora Morocho a pro tvorbu grafů, které ukazují spektrální charakteristiky trénovacích dat vstupujících do klasifikace. Tyto skripty lze volně využít pro další výzkum. Odkazy jsou v příloze 2.

Jedním ze vstupních úkolů bylo zjistit, jaké možnosti klasifikátorů prostředí GEE nabízí a zároveň jaké klasifikátory pro tento typ výzkumu doporučuje odborná literatura. Literární rešerše ukázala na metodu RF (Ok, Akar, Gungor 2012, Noi, Kappas 2017) jako perspektivní pro klasifikaci krajinného pokryvu v pro nás exotické fragmentované krajině Etiopie. Noi, Kappas (2017) ve své studii upozorňují na nutnost testování více hodnot

parametrů metody RF, protože co funguje pro jedno území (Case study: Red River Delta ve Vietnamu), nemusí fungovat pro druhé. Noi, Kappas (2017) testovali nastavení 100, 200, 500 a 1000 stromů a 1 až 10 spektrálních pásem vstupujících do každého stromu, přičemž nejvyšší přesnosti bylo dosaženo s nastavením 500 stromů a stejně pro 2, 3 a 4 spektrální pásma. Ok, Akar, Gungor (2012) testovali 100, 200, 250 a 500 stromů a 1 až 4 spektrální pásma a nejvyšší přesnosti 86 % bylo dosaženo s počtem 200 stromů a 2 spektrální pásma. S inspirací metodických postupů těchto relevantních studií v této diplomové bylo testováno 10 kombinací 4 nejdůležitějších parametrů z 6, které prostředí GEE nabízí. Výslednou kombinací, která vedla k nejvyšší celkové přesnosti klasifikace, se ukázala být kombinace 5 s nastavením 100 stromů, 4 spektrální pásma, 1 pro minimální populaci listu a 40 % trénovacích dat vstupujících do každého stromu. Breiman (2001 cit. v Noi, Kappas 2017, s. 7) tvrdí, že „využití většího počtu stromů, než je požadováno, není nutné, ale kvalitu modelu to nezhorší“. Co se týče hlavních metodických poznatků a cenných metodických výsledků, pak na rozdíl od výše zmíněných studií tato diplomová práce dospěla k závěru, že vyšší počet stromů v lese nezvýšil celkovou přesnost klasifikace (viz Tab. 15). Pouze se tak zbytečně prodlužoval čas výpočtu a některé výpočty dokonce nemohly být provedeny, jelikož se proces přerušil. Nastavení 100 stromů se v případě této studie jeví jako ideální.

Dalším z dílčích úkolů bylo sesbírání kvalitních trénovacích a kontrolních množin. Eggen, Ozdogan, Zaitchik a Simane (2016) upozorňují právě na důležitost kvalitních referenčních dat pro klasifikaci většího fragmentovaného území. Terénní výzkum zájmové oblasti regionu Sidama proběhl pod záštitou projektu ČRA na podzim roku 2020. Přestože nebylo možné sesbírat data rovnoměrně po celém území, povedlo se sesbírat 582 bodů napříč všemi sledovanými třídami land cover a dále také vyzorovat strukturu krajiny, podobnosti, charakteristiky a prostorové a strukturální přechody jednotlivých tříd. Užitečná se ukázala být také aplikace ArcGIS Collector, která sloužila jako nástroj pro sbírání dat v terénu. Jednoduché uživatelské rozhraní, možnost přidávat poznámky k jednotlivým bodům přímo v terénu, zobrazení různých podkladů a efektivní uložení dat do formátu SHP přispělo k efektivitě terénního výzkumu. K větší přesnosti sběru trénovacích dat by mohla přispět možnost osobně vstoupit doprostřed zájmových plošek, které byly často za plotem. Z časových i bezpečnostních důvodů byla velká část dat sbírána tzv. z cesty a následně polohově a tematicky zpřesněna nad snímky s velmi vysokým rozlišením. Prázdná místa na mapě musela být doplněna trénovacími daty pomocí interpretace snímků Maxar a Airbus v

softwaru GEP ze stejného období, jako byla sbírána in-situ data. Kombinace těchto dvou typů dat mohla vést k určitým nepřesnostem, které ovšem nelze považovat za signifikantní.

Výsledný obraz klasifikace byl filtrován většinovým filtrem pro odstranění osamocených pixelů a dalších šumů. Tak byla vytvořena první varianta výstupu klasifikace. Druhá varianta byla vedena uživatelskými nároky koncových uživatelů a obsahovala navíc filtr pomocí segmentace, kdy byly zachovány prvky o rozloze 2500 m² pro klasifikaci krajinného pokryvu regionu Sidama nad snímky Sentinel-2. Díky segmentaci byly zachovány prvky, které mají úzký, ale dlouhý charakter, jako cesty nebo vodní toky. Prostředí GEE nabízí velké množství přístupů k filtraci obrazu a tato oblast by mohla být předmětem dalšího výzkumu a testování.

Sestavený algoritmus byl též aplikován a testován na datech PlanetScope, což jsou data s vyšším prostorovým rozlišením. Data PlanetScope a Sentinel-2 se liší v spektrálním, prostorovém i časovém rozlišení. Data Sentinel-2 Level-2A nabízí 13 spektrálních pásem, z nichž bylo využito 10 pro klasifikaci, PlanetScope nabízí 4 pásma (viz Tab. 3 a 4). Z využitých pásem má Sentinel-2 oproti PlanetScope navíc 4 pásma Red Edge a 2 pásma SWIR. Využitá spektrální pásma datasetu Sentinel-2 mají prostorové rozlišení 10 a 20 m/px, 4 pásma dat PlanetScope mají 3 m/px. Totožné místo je pomocí družice mise Sentinel-2 zaznamenáno minimálně jedenkrát za 5 dní, mise PlanetScope nabízí tyto snímky z každého dne. Nevýhodou dat PlanetScope je cena, protože narozdíl od dat Sentinel-2 nejsou dostupná zdarma. Cílem bylo zjistit, zdali lze využít snímky PlanetScope pro klasifikaci menších území kebelí v měřítku 1 : 10 000, jaké přesnosti klasifikace lze dosáhnout a jaké detaily a třídy land cover lze na snímcích zachytit.

V kontextu prostorového rozlišení Eggen, Ozdogan, Zaitchik a Simane (2016) použili snímky Landsat s prostorovým rozlišením 30 m/px pro klasifikaci krajinného pokryvu rozsáhlého území Etiopie s přesností 55 %. Nižší prostorové rozlišení dat Landsat označují autoři článku za jednu z překážek pro vyšší přesnost klasifikace tak fragmentované zemědělské krajiny, jako je etiopská vrchovina, kdy mnoho pixelů je tzv. mixed-pixels. Degife, Zabel, Mauser (2018) upozorňují na vyšší celkovou přesnost při využití snímků s vyšším prostorovým rozlišením (data Sentinel-2 oproti Landsat 5, 94,4 oproti 88,4 %). Tato práce prokázala užitečnost využití dat Sentinel-2, kdy nejvyšší přesnost klasifikace dat Sentinel-2 překročila požadovanou hranici 80 % a dosáhla hodnoty 84,1 %. Na druhou stranu klasifikace dat velmi vysokého prostorového rozlišení PlanetScope nepřinesla vyšší přesnost a

dosáhla maximální přesnosti 73,74 %. Nižší přesnost byla v tomto případě částečně způsobena nízkým spektrálním rozlišením. Data PlanetScope obsahují 4 pásma, z nichž pásmo NIR by mělo být užitečné především v detekci vlhkosti, tedy pro třídy vodní plocha a mokřad. Jak bylo popsáno v kapitole 4.3, tyto třídy se však ukázaly být velmi problematické. Vodní toky na území kebelí byly zaneseny materiálem, což vedlo k záměně s jinými třídami land cover, protože vykazovaly podobné spektrální vlastnosti. Třída mokřad byla spektrálně velmi podobná třídě louka/pastvina, což vedlo k jejich záměně. Totéž platilo pro třídy zástavba a holá půda, které byly také zaměňovány. V tomto případě byla záměna způsobena charakterem střech budov, které byly převážně plechové a jevíly se spektrálně podobné jako skalní útvary (= holá půda). Dalším negativním aspektem dat PlanetScope bylo, že území kebelí byla snímána brzy ráno a tím vznikaly v údolí stíny, čímž byla přesnost klasifikace také negativně ovlivněna. Nejvíce se to projevilo u třídy vodní plocha, neboť řeky na území kebelí tekly územími, která byla ovlivněna zastíněním. Časové rozlišení dat PlanetScope je inzerováno jako 1denní, avšak na příkladu této studie se prokázala nižší kvalita a užitečnost těchto dat. Přestože byl v oblasti kebelí proveden detailní sběr trénovacích dat, snímky PlanetScope mají vysoké prostorové rozlišení a bylo využito vegetačních indexů, klasifikace nad těmito snímky vyšla s nižší přesností než klasifikace snímků Sentinel-2. Na druhou stranu by si klasifikace dat PlanetScope zasloužila detailnější testování. Jelikož se klasifikace týkala nepřilíh rozsáhlého území zájmu, šlo by klasifikaci realizovat v desktopovém prostředí a výsledky následně porovnat s vyvinutým algoritmem v GEE. Dále by bylo možné zkoumat jiná území zájmu a výsledky porovnat.

Co se týče jednotlivých land cover tříd, problémovou třídou klasifikace regionu Sidama nad snímky Sentinel-2 byly mokřady, které byly hodnoceny s nejnižší uživatelskou i zpracovatelskou přesností, a to 37,50 a 42,86 %. Jedná se o velmi sezónně proměnlivý typ pokryvu, klasifikace této třídy je tedy problematická. Další problémovou třídou byly vodní plochy, které zahrnovaly vodní plochy a toky s čistou i znečištěnou vodou. Právě znečištěné vodní toky, jež s sebou nesly bohaté sedimenty písku, prachu a sedimenty erodované půdy, způsobovaly chyby v klasifikaci. Silně znečištěné vodní toky byly často klasifikovány jako holá půda, přestože existovala in-situ měření a fotografie z těchto míst a jednalo se skutečně o vodní toky. Tento problém byl částečně vyřešen přidáním dalších trénovacích bodů v místech znečištěných řek. Problémy také vznikaly v rozpoznání tříd plodiny a louka/pastvina, a to především v případě nízkých zeleně zbarvených plodin, např. obilí teff. Určitým řešením se jeví zjištění, že na snímcích s velmi vysokým rozlišením lze louku a pole

s teffem rozpoznat pomocí krajinné struktury. Pole bývají pravidelně rozdělená, kdežto třída louka/pastvina mívá nepravidelný tvar a větší rozměry. Pro rozlišení těchto dvou fenoménů by mohla být vhodná objektová klasifikace, která hledí nejen na spektrální příznaky, ale také na tvar a velikost zájmových plošek.

Z hlediska porovnání klasifikace území kebelí nad daty Sentinel-2 a PlanetScope lze diskutovat vzniklé rozdíly v zastoupení tříd land cover. Zásadní je rozdíl v podílu třídy mokřad. V případě kebele Chancho byl podíl třídy mokřad v klasifikaci nad daty PlanetScope poloviční oproti klasifikaci nad daty Sentinel-2. V případě kebele Dangora Morocho tomu bylo přesně naopak. Na území kebele Chancho se nachází velké území luk/pastvin, které je v klasifikovaném období (listopad) podmáčeno, tedy spadá do třídy mokřad. Tento fenomén nebyl nad daty PlanetScope zaznamenán. Na území kebele Dangora Morocho se takto prostorově výrazný mokřad nenacházel, rozdíl v zastoupení byl způsoben diskutabilní klasifikací malých roztroušených ploch mokřadů nad snímky PlanetScope. Velký procentuální rozdíl zastoupení byl v případě třídy křoviny, kde obě území zaznamenala více než 10% rozdíl, co se týče klasifikací nad snímky PlanetScope a Sentinel-2. Na území obou kebelí byl podíl třídy křoviny větší v případě klasifikace nad snímky Sentinel-2. Jako křoviny byla klasifikována četná území plodin, která měla podobné spektrální příznaky jako křoviny, např. uschlá kukuřice.

V zadání projektu ČRA měla třída holá půda, a zároveň tak degradované plochy, nejvyšší stupeň důležitosti. Tuto třídu se však nepodařilo na úrovni kebelí kvalitně klasifikovat. Řešením z hlediska lokálních studií území kebelí (8 a 10 km²) by přesnější možností klasifikace land cover mohla být vektorizace. Jejím kvalitnímu zpracování by však muselo předcházet detailní seznámení s krajinnou strukturou a s vizuální interpretací každé třídy land cover. K tomu by mohla posloužit nasbíraná in-situ data a pořízené fotografie z terénu. Druhou možností by mohla být objektově-orientovaná klasifikace, která ke spektrálním příznakům přidává tvarové a velikostní. Nadstavbou pro tuto práci by mohlo být vyvinutí algoritmu či cyklu, který by vypočítal celkovou přesnost klasifikace pro všechny kombinace v určitém rozsahu hodnot zvolených parametrů a automaticky by určil, která kombinace je nejlepší, a to bez manuálního přepisování vstupních hodnot. Tím by se proces testování urychlil a bylo by ověřeno větší množství možností vstupních parametrů.

6 Závěr

Hlavním cílem diplomové práce byla klasifikaci krajinného pokryvu exotické lokality regionu Sidama v Etiopii v roce 2020, a to s využitím družicových dat Sentinel-2 a PlanetScope. Klasifikace proběhla v cloudovém prostředí GEE za pomoci metod strojového učení.

Kromě volby a testování klasifikační metody hlavními geoinformačními úkoly k řešení byla volba přesné metody maskování oblačnosti a mozaikování. Dalším netriviálním úkolem byla definice legendy a minimální mapovací jednotky pro klasifikaci s využitím terénního průzkumu. Celý metodický proces je k dispozici ve formě algoritmu ve volně dostupné platformě GEE a tímto způsobem jsou vytvořené metody k dispozici široké uživatelské skupině. Výsledky klasifikace odpovídají požadavkům koncových uživatelů, tzn. institucím regionální a místní správy. Kód nebo jeho části lze využít pro další výzkum v oblasti klasifikace a dálkového průzkumu Země.

Přes všechny překážky a vzniklé problémy bylo docíleno více než 80% přesnosti klasifikace krajinného pokryvu regionu Sidama v roce 2020 nad snímky Sentinel-2, což odpovídá cílům a zadání práce. Pomocí dostupných nástrojů byla navržena metoda maskování oblačnosti a mozaikování. Problematika maskování oblačnosti dat Sentinel-2 je stále ne zcela vyřešenou problematikou, což se ukázalo i v případě této práce, kdy problematickou se jeví detekce nevýrazné oblačnosti, která má však vliv na přesnost klasifikace. V průběhu klasifikace byly testovány kombinace parametrů klasifikátoru RF a ukázalo se, že nastavení velkého počtu stromů nepřináší vyšší přesnost klasifikace. Nejvyšší celkovou přesnost přinesla kombinace parametrů 100 stromů, 4 spektrální pásma, hodnota 1 pro populaci listu a 40 % trénovacích dat. Klíčové se ukázalo být testování více možností. Přestože trénovací data nebyla zcela rovnoměrně rozmístěna v zájmovém území, celková přesnost klasifikace vyšla nad 80 %, tedy lze metodu považovat za perspektivní pro klasifikaci land cover ve fragmentované a vysoce heterogenní krajině etiopské vrchoviny, kde není snadné získat kvalitní referenční data.

Co se týče datových vstupů klasifikace, nepotvrdila se hypotéza, že by vyšší prostorové rozlišení dat PlanetScope mělo přinést lepší výsledky klasifikace na lokální úrovni menších území (tzv. kebelí). V případě této práce a použitých metod nepřinesla data PlanetScope přesné výsledky klasifikace území v etiopské vrchovině. Jelikož je tato

diplomová práce pilotní studií pro projekt ČRA, zjištění ohledně dat PlanetScope je velmi přínosné. Problematiku využití dat velmi vysokého rozlišení by tak bylo vhodné detailněji prozkoumat v navazujícím výzkumu s otestováním jiných území, případně i jiných metod klasifikace, např. objektově-orientovaných. Další oblastí výzkumu může být testování využití dat PlanetScope na území Česka a následné porovnání s touto studií věnující se pro nás exotické fragmentované krajině v Etiopii. Velkou výhodou by ve výzkumu na území Česka měla být možnost vstoupit do zájmových plošek a trénovací data tak naměřit velice přesně.

Závěrem lze shrnout, že diplomová práce přináší metodický návod pro klasifikaci výrazně heterogenního a fragmentovaného území etiopské vrchoviny a pokouší se expertně porovnat využitelnost dat Sentinel-2 a PlanetScope pro účely klasifikace land cover na odlišných měřítkových úrovních. Pro tyto účely práce využila moderních metody zpracování dat ve volně otevřené výzkumné cloudové platformě GEE. Dosažené výsledky lze brát za dílčí posuny v řešené problematice. Otevřenost dosažené metodiky a vytvořených skriptů vybízí v pokračování řešení této netriviální problematiky ideálně v rámci mezinárodního týmu tvořeného jak experty DPZ, tak i zástupci expertů jiných disciplín a místních obyvatel.

7 Použitá literatura a data

ADRA (2014): Humanitární organizace ADRA. Pomoc v zahraničí. Dostupné z: <https://www.adra.cz/pomoc-v-zahranici> (cit. 27. 5. 2020).

ALC (2018): Africa Land Cover demo. Databáze krajinného pokryvu afrického kontinentu v roce 2018, program Copernicus. Dostupné z: <https://africa.lcvviewer.vito.be/2018> (cit. 7. 6. 2020).

BRIGGS, P. (2012): Ethiopia. The Glove Press Inc, Guilford, USA.

BROWNLEE, J. (2016): Bagging and Random Forest Ensemble Algorithms for Machine Learning. Machine Learning Mastery. Dostupné z: <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/> (cit. 22. 6. 2021).

CCI (2016): CCI Land Cover – S2 Prototype Land Cover 20 m map of Africa. Dostupné z: <http://2016africalandcover20m.esrin.esa.int> (cit. 7. 6. 2020).

ČERVENÁ, L. (2016a): Neřízená klasifikace. Výukový list. Dostupné z: <https://www.natur.cuni.cz/geografie/geoinformatika-kartografie/ke-stazeni/projekty/moderni-geoinformacni-metody-ve-vyuce-gis-kartografie-a-dpz/nerizena-klasifikace/view> (cit. 17. 6. 2020).

ČERVENÁ, L. (2016b): Řízená klasifikace. Výukový list. Dostupné z: <https://www.natur.cuni.cz/geografie/geoinformatika-kartografie/ke-stazeni/projekty/moderni-geoinformacni-metody-ve-vyuce-gis-kartografie-a-dpz/rizena-klasifikace/view> (cit. 17. 6. 2020).

ČERVENÁ, L. (2018): Hodnocení přesnosti klasifikace. Výukový list. Dostupné z: <https://www.natur.cuni.cz/geografie/geoinformatika-kartografie/ke-stazeni/projekty/moderni-geoinformacni-metody-ve-vyuce-gis-kartografie-a-dpz/hodnoceni-presnosti-klasifikace/view> (cit. 25. 7. 2021).

Člověk v tísni (2020): Etiopie – Člověk v tísni. Dostupné z: <https://www.clovekv tisni.cz/co-delame/humanitarni-a-rozvojova-pomoc/etiopie> (cit. 27. 5. 2020).

ČRA (2020): Česká rozvojová agentura. Co děláme – v Etiopii. Dostupné z: http://www.czechaid.cz/vysledky-filtrovani/?wpv-zeme=etiopie&wpv-temata=0&aktivni%5B%5D=&wpv_filter_submit=filtrovat (cit. 27. 5. 2020).

- DEGIFE, A. W., ZABEL, F., MAUSER, W. (2018): Assessing land use and land cover changes and agricultural farmland expansions in Gambella Region, Ethiopia, using Landsat 5 and Sentinel 2a multispectral data. *Heliyon*, 4, 11. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S2405844018347601?via%3Dihub> (cit. 10. 6. 2020).
- EGGEN, M., OZDOGAN, M. ZAITCHIK, F., SIMANE, B. (2016): Land Cover Classification in Complex and Fragmented Agricultural Landscapes of the Ethiopian Highlands. *Remote Sensing*, 8, 12. Dostupné z: <https://www.mdpi.com/2072-4292/8/12/1020> (cit. 26. 6. 2020).
- ETH (2017): ETH Land Cover EOS4 SD. Online databáze krajinného pokryvu části afrického kontinentu v rámci projektu EOS4 SD. Dostupné z: [https://eo4sd.lizard.net/en/map/satellite,raster\\$7661173/point/127728/@9.8227,39.1141,9/-1Days19Hours+0Days8Hours](https://eo4sd.lizard.net/en/map/satellite,raster$7661173/point/127728/@9.8227,39.1141,9/-1Days19Hours+0Days8Hours) (cit. 7. 6. 2020).
- FETENE, A. HILKER, T., YESHITELA, K., PRASSE, R., COHEN, W., YANG, Z. (2016): Detecting Trends in Landuse and Landcover Change of Nech Sar National Park, Ethiopia. *Environmental Management*, 57, 137–147.
- GEE (2020): Google Earth Engine. FAQ. Dostupné z: <https://earthengine.google.com/faq> (cit. 21. 5. 2020).
- HISHE, S., BEWKET, W., NYSSSEN, J., LYIMO, J. (2019): Analysing past land use land cover change and CA-Markov-based future modelling in the Middle Suluh Valley, Northern Ethiopia. *Geocarto International*, 35, 3. Dostupné z: <https://doi.org/10.1080/10106049.2018.1516241> (cit. 27. 5. 2020).
- CHAI, W. (2020): Definition of cloud computing. Dostupné z: <https://searchcloudcomputing.techtarget.com/definition/cloud-computing> (cit. 28. 6. 2021).
- KOLÁŘ, J. (1990): Dálkový průzkum Země. SNTL – Státní nakladatelství technické literatury, Praha.
- LÓPEZ, B. R. (2018): What is Google Earth Engine and how you can use it? *Gis Online Courses*. Dostupné z: <https://www.giscourse.com/what-is-google-earth-engine-and-how-you-can-use-it> (cit. 21. 5. 2020).
- MEKUYIE, M., JORDAAN, A., MELKA, Y. (2018): Land-use and land-cover changes and their drivers in rangeland-dependent pastoral communities in the southern Afar Region of

- Ethiopia. African Journal of Range and Forage Science, 35, 1. Dostupné z: <https://doi.org/10.2989/10220119.2018.1442366> (cit. 27. 5. 2020).
- MENDELU (2020): Kapitola 4 – Úvod do hodnocení krajiny. Dostupné z: https://is.mendelu.cz/eknihovna/opory/zobraz_cast.pl?cast=71682 (cit. 26. 6. 2020).
- NAGI, R. (2010): On map scale and raster resolution. Dostupné z: https://www.esri.com/arcgis-blog/products/product/imagery/on-map-scale-and-raster-resolution/?fbclid=IwAR2Z3GP1b-pAkn9TpCmRN_FwxNJdgki4RfWOLqusdlECZ5QrvDSNO9w9U7E (cit. 12. 7. 2021).
- NOI, P. T., KAPPAS, M. (2017): Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, 18, 1. Dostupné z: <https://www.mdpi.com/1424-8220/18/1/18> (cit. 26. 6. 2020).
- OK, A. O., AKAR, O., GUNGOR, O. (2012): Evaluation of random forest method for agricultural crop classification. European Journal of Remote Sensing, 45, 1. Dostupné z: <https://www.tandfonline.com/doi/abs/10.5721/EuJRS20124535> (cit. 28. 6. 2020).
- ORACLE (2020): Oracle Artificial Intelligence (AI) - co je strojové učení? Dostupné z: <https://www.oracle.com/cz/artificial-intelligence/what-is-machine-learning.html> (cit. 18. 6. 2020).
- PLANET (2021): Planet Imagery Product Specifications. Dostupné z: https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf (cit. 15. 6. 2021).
- POLÁČEK, P. (2012): Etiopie, kolébka lidstva, království ptactva a opic. Pavel Poláček, Praha.
- POTŮČKOVÁ, M. (2020): Získávání informace z dat DPZ – Klasifikace obrazu. Prezentace k přednášce.
- SENTINEL-HUB (2021a): Normalized Difference Vegetation Index. Custom Scripts. A repository of custom scripts that can be used with Sentinel-Hub services. Dostupné z: <https://custom-scripts.sentinel-hub.com/sentinel-2/ndvi/> (cit. 29. 7. 2021).
- SENTINEL-HUB (2021b): Normalized Difference Water Index. Custom Scripts. A repository of custom scripts that can be used with Sentinel-Hub services. Dostupné z: <https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndwi/> (cit. 29. 7. 2021).

SRIVASTAVA, T. (2015): Ladění parametrů vašeho modelu Random Forest. Analytics Vidhya. Dostupné z: <https://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/> (cit. 22. 6. 2021).

ŠTYCH, P., JEŘÁBKOVÁ, B., LAŠTOVIČKA, J., RIEDL, M., PALUBA, D. (2019): A Comparison of WorldView-2 and Landsat 8 Images for the Classification of Forests Affected by Bark Beetle Outbreaks Using a Support Vector Machine and a Neural Network: A Case Study in the Sumava Mountains. Geosciences, 9, 9. Dostupné z: <https://www.mdpi.com/2076-3263/9/9/396> (cit. 26. 6. 2020).

WIKIPEDIE (2020): Etiopie (článek s 160 zdroji, k nalezení v databázi Národní knihovny ČR). Otevřená encyklopedie. Dostupné z: <https://cs.wikipedia.org/wiki/Etiopie> (cit. 26. 6. 2020).

WIKIPEDIE (2021): Sidama. Otevřená encyklopedie. Dostupné z: <https://cs.wikipedia.org/wiki/Sidama> (cit. 16. 6. 2021).

YIU, T. (2019): Understanding Random Forest. Dostupné z: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2> (cit. 18. 6. 2020).

ŽUBRIENTOVSKÝ, L., ŠVÁBOVÁ, L., ŠVÁB, O. (2017): Sentinel-2 – Datová specifikace. Uživatelská příručka. Dostupné z: <https://collgs.czechspaceportal.cz/sentinel-2-datove-specifikace> (cit. 21. 5. 2020).

Použitá data

AIRBUS (2021): One Atlas Satellite Imagery. Dostupné z: <https://oneatlas.airbus.com/> (cit. 13. 7. 2021).

GOOGLE EARTH (2021): Google virtuální glóbus. Dostupné z: <https://earth.google.com/web/> (cit. 26. 4. 2021).

MAXAR (2021): Maxar technologies Satellite Imagery. Dostupné z: <https://www.maxar.com/> (cit. 13. 7. 2021).

OPEN AFRICA (2020): Ethiopia shapefiles. Repozitář open dat pro africký kontinent. Dostupné z: <https://open.africa/dataset/ethiopia-shapefiles> (cit. 17. 6. 2020).

PLANET (2021b): Planet Explorer. Dostupné z: <https://www.planet.com/explorer/#/zoom/2.48> (cit. 27. 6. 2021).

SENTINEL-2 MSI (2021): MultiSpectral Instrument, Level-2A. Earth Engine Data Catalog. Dostupné z: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR (cit. 27. 6. 2021).

USGS (2020): The United States Geological Survey Explorer. Dostupné z: <https://earthexplorer.usgs.gov/> (cit. 28. 6. 2021).

Seznam příloh

Příloha 1 Fotodokumentace klasifikačních tříd land cover	86
Příloha 2 Skripty	88
Příloha 3 Mapy krajinného pokryvu	88
Příloha 4 Geodatabáze výstupů	89

Příloha 1 Fotodokumentace klasifikačních tříd land cover

Snímky s velmi vysokým rozlišením (Maxar 2021, Airbus 2021) a fotografie (vlastní) příkladů jednotlivých klasifikačních tříd land cover.



Třída zástavba (settlement)



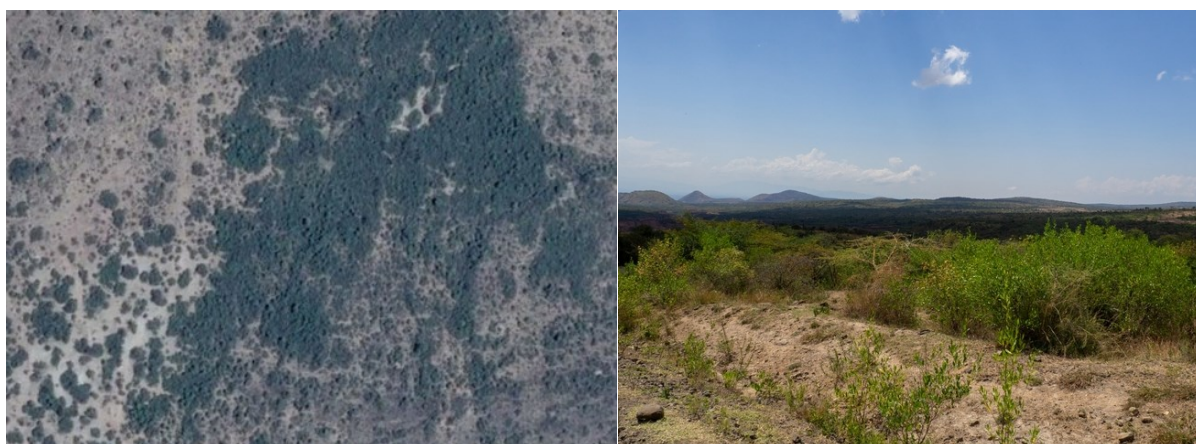
Třída plodiny (crops)



Třída louka/pastvina (grassland/pasture)



Třída les (forest)



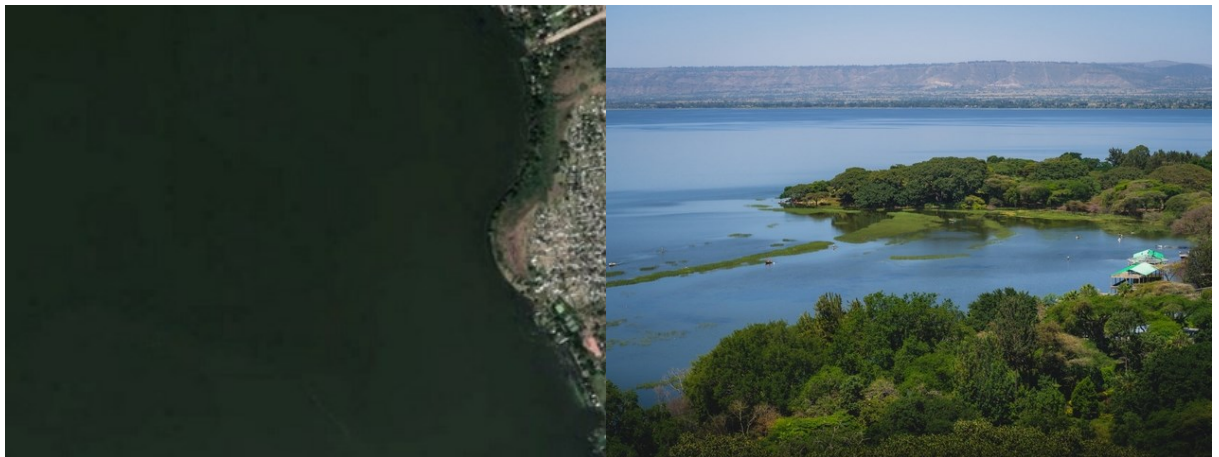
Třída křoviny (shrubland)



Třída holá půda (bareland)



Třída mokřad (wetland)



Třída vodní plocha (water body)

Příloha 2 Skripty

Tvorba grafů spektrálních příznaků trénovacích dat

Odkaz do GEE: <https://code.earthengine.google.com/0601cd6367ea7b18c95d3676171da1cc>

Textový soubor: Skript_grafy_spektrálních_příznaků.txt

Klasifikace krajinného pokryvu regionu Sidama

Odkaz do GEE: <https://code.earthengine.google.com/c6727b3bff6df2c99e74c495bcaba5ab>

Textový soubor: Skript_klasifikace_Sidama.txt

Klasifikace krajinného pokryvu kebelí Chancho a Dangora Morocho

Odkaz do GEE: <https://code.earthengine.google.com/89ccae6b993c856fb3ee411aa87dbff6>

Textový soubor: Skript_klasifikace_kebele.txt

Příloha 3 Mapy krajinného pokryvu

Mapa krajinného pokryvu regionu Sidama v roce 2020

Výstup v PDF: Mapa_LC_Sidama_100tis.pdf

Mapa krajinného pokryvu kebele Chancho v roce 2020

Výstup v PDF: Mapa_LC_Chancho_10tis.pdf

Mapa krajinného pokryvu kebele Dangora Morocho v roce 2020

Výstup v PDF: Mapa_LC_DangoraMorocho_10tis.pdf

Příloha 4 Geodatabáze výstupů

Krajinný pokryv regionu Sidama v roce 2020 (10 m/px)

Soubor v GeoTIFF: lc_sidama.tif

Krajinný pokryv regionu Sidama v roce 2020 (pro mapu v 1 : 100 000)

Soubor v GeoTIFF: lc_sidama_100tis.tif

Krajinný pokryv kebele Chancho v roce 2020 (3 m/px)

Soubor v GeoTIFF: lc_chancho.tif

Krajinný pokryv kebele Chancho v roce 2020 (pro mapu v 1 : 10 000)

Soubor v GeoTIFF: lc_chancho_10tis.tif

Krajinný pokryv kebele Dangora Morocho v roce 2020 (3 m/px)

Soubor v GeoTIFF: lc_dangoramorocho.tif

Krajinný pokryv kebele Dangora Morocho v roce 2020 (pro mapu v 1 : 10 000)

Soubor v GeoTIFF: lc_dangoramorocho_10tis.tif