



VYUŽITÍ SOUBORU DAT CORINE LAND COVER PRO PIXELOVOU KLASIFIKACI SCÉN SENTINEL-2

Ing. Tomáš Bouček

1. České vysoké učení technické v Praze, Fakulta stavební, Katedra Geomatiky, Praha, Thákurova 7/2077, tomas.boucek@fsv.cvut.cz

ABSTRAKT

Práce se zabývá zkoumáním možností využití souboru vektorových dat CORINE Land Cover jako trénovacích ploch pro následnou řízenou klasifikaci nad scénami z družic Sentinel-2. Klasifikace probíhá na třech úrovních tematické podrobnosti, která je dána nomenklaturou CORINE. Pro testování byla vybrána tři evropská geograficky a klimaticky rozdílná území. Pomocí úprav dat CORINE je snahou docílit co největší přesnosti klasifikace. Z výsledků je následně analyzováno, jaké třídy z nomenklatury CORINE lze takto klasifikovat, a které je naopak lepší z klasifikace odstranit. Ke klasifikování je použit statistický klasifikátor maximum likelihood. V závěru je pak využito i klasifikátoru random forest, který patří do skupiny klasifikátorů spadající pod strojové učení. Výsledky obou klasifikátorů jsou poté porovnány. Výsledná klasifikace bude vyhodnocena s ohledem na víceúrovňovou soustavu CORINE tříd.

KLÍČOVÁ SLOVA

CORINE, land cover, Sentinel-2, řízená klasifikace, trénovací data, maximum likelihood, random forest

ÚVOD

Tato práce byla zpracovávána v rámci projektu „Geo-harmonizer: EU-wide automated mapping system for harmonization of Open Data based on FOSS4G and Machine Learning“, jehož cílem je vyvinout původní, webově orientovaný, modulární systém („Geo-harmonizer“) určený pro uchování a sdílení geografických dat s tematickým obsahem (land cover, kvalita životního prostředí apod.) pokrývajících území kontinentální Evropy. Na projektu spolupracuje více zemí Evropské unie. Českou republiku v projektu zastupuje ČVUT, které je i koordinátorem celého projektu 0.

Práce se zabývá řízenou klasifikací obrazových dat pořízených metodou dálkového průzkumu Země. Konkrétně se jedná o družicová multispektrální data z družic Sentinel-2 (produkty úrovně Level-2A) získaná na principu odrazu elektromagnetického záření. Klasifikace byla provedena pro různé úrovně tematické podrobnosti, kde byla sledována přesnost, na kterou se lze v jednotlivých úrovních dostat.

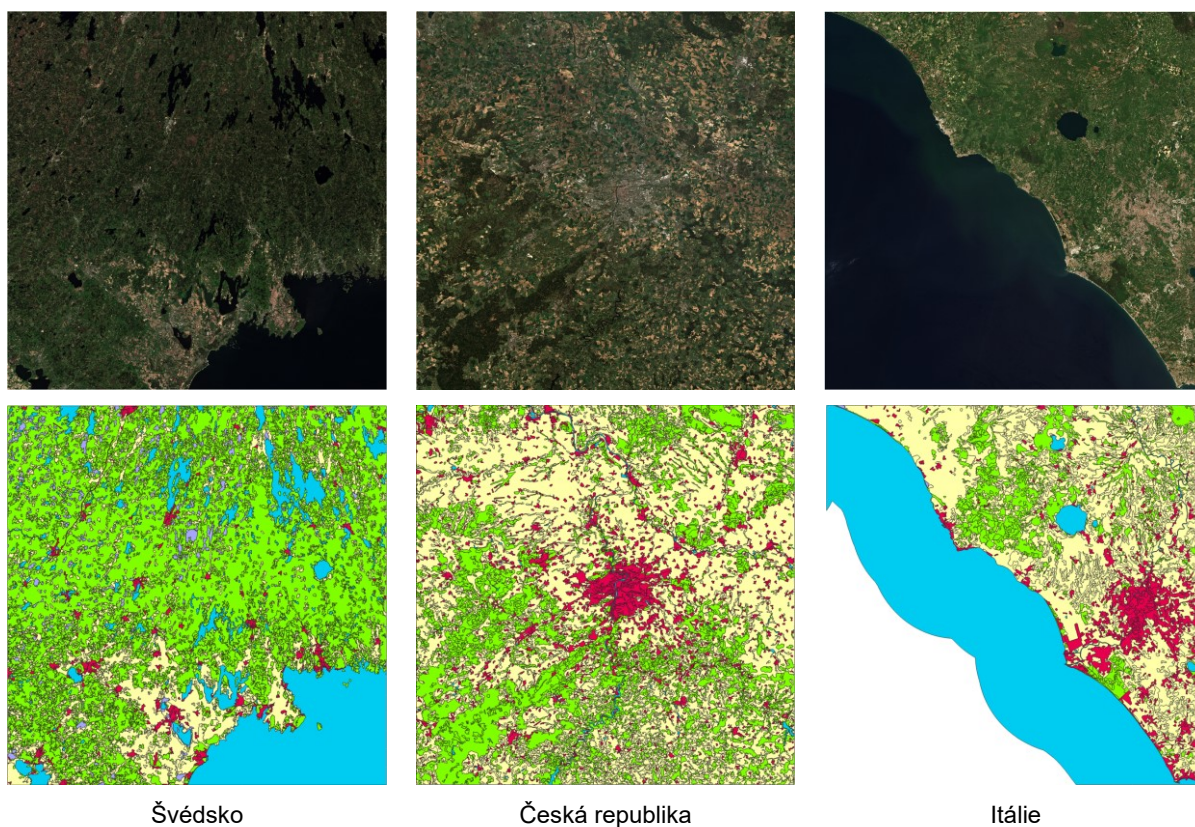
Při práci byly použity dva způsoby klasifikace. Jedním z nich je způsob využívající metodu *maximum likelihood*, druhým způsobem je metoda random forest založená na tzv. strojovém učení (*machine learning*). Samotné klasifikaci předchází příprava a úprava trénovacích ploch. Pro trénovací plochy bylo využito volně dostupných dat CORINE Land Cover (CLC) poskytujících informace o druhu pokrytí Země (tzv. *land cover*). Družicové scény spolu s daty CLC byly vzaty z roku 2018, jelikož tento rok je zatím posledním referenčním rokem, pro který byla vrstva CLC vytvořena.



Klasifikována byla tři různá evropská území reprezentující rozdílné klimatické podmínky v Evropě. Jedná se o území ve Švédsku, v České republice a v Itálii, která nejlépe vystihují různorodost evropského kontinentu (viz Obr. 1 a Obr. 2).



Obr. 1: Poloha zpracovávaných lokalit



Obr. 2: Vybrané dlaždice (RGB kompozit + 1. úroveň CLC)



Cílem této práce je tedy klasifikovat tyto tři vybraná území, upravit data CORINE tak, aby mohla sloužit jako trénovací plochy a zjistit, které třídy z nomenklatury CORINE je možné na základě spektrálních příznaků dostatečně přesně klasifikovat. Data CORINE jsou rozdělena do tří úrovní. První úroveň tvoří pět základních tříd, druhou úroveň patnáct tříd a třetí úroveň celkem 44 tříd. Přehled jednotlivých tříd znázorňuje Obr. 3.



Obr. 3: Třídy CLC [2]



METODIKA

Pro přípravu trénovacích ploch byl použit software QGIS. Následná řízená klasifikace byla spouštěna v softwaru GRASS. Oba softwary jsou vyvíjeny jako open source pod všeobecnou licencí GNU GPL.

V práci byly použity pro každé z vybraných území dvě totožné scény, ale z jiného časového období. Jedná se o jarní a letní scénu z roku 2018 a použita byla všechna pásma s rozlišením 20 m. Takových pásem je pouze šest, nicméně stažená data obsahují i převzorkovaná desetimetrová data (kromě pásma B8, které bylo převzorkováno ručně), takže celkový počet pásem použitých ke klasifikaci je deset pro jednu scénu (B2, B3, B4, B5, B6, B7, B8, B8a, B11 a B12) a tedy dvacet pro jedno území. Scény byly vybírány tak, aby byly pokud možno bezoblačné, případně s maximální hodnotou oblačnosti 5 %.

Hodnocení klasifikace bylo prováděno pomocí tzv. kontingenčních tabulek, které ukazují, jak se pixely z trénovacích ploch zařadily do jednotlivých tříd. Pokud se pixely trénovacích ploch jedné třídy zařadily ve větším počtu do třídy jiné, indikuje to, že trénovací množiny nebyly vhodně zvoleny a jednotlivé třídy se vzájemně překrývají [3].

Klasifikace byla dále posuzována pomocí celkové přesnosti, která udává poměr všech správně klasifikovaných pixelů a všech pixelů z trénovacích ploch.

$$CP = \frac{1}{n} \sum_{k=1}^N a_{kk} \quad , \quad (1)$$

kde

n je celkový počet pixelů použitých pro trénovací plochy,

N je celkový počet klasifikačních tříd,

a_{kk} jsou diagonální prvky kontingenční tabulky udávající počet správně zařazených pixelů z trénovacích a poté i testovacích ploch do jednotlivých tříd.

Kromě celkové přesnosti byl použit i tzv. *kappa index*. Ten přesnost klasifikace zjišťuje pro dvě nezávislé klasifikace se stejnou spolehlivostí. V potaz bere náhodnou shodu, což je pravděpodobnost, že se klasifikace s referenčními daty shoduje pouze náhodně. Index kappa lze zapsat následovně:

$$\kappa = \frac{P_0 - P_c}{1 - P_c} \quad , \quad (2)$$

kde

P_0 vyjadřuje shodu mezi referenčními daty a klasifikací,

P_c vyjadřuje náhodnou shodu.

Kappa index nabývá hodnot v rozmezí $<0; 1>$, kdy hodnota 1 by znamenala, že se klasifikace vyhnula 100 % chyb vznikajících při náhodném procesu [3].



Příprava trénovacích dat a klasifikace

Klasifikovat se začalo nejprve na první úrovni CLC. Ta obsahuje následující třídy:

- Zástavba (1)
- Zemědělské oblasti (2)
- Lesní a polopřírodní oblasti (3)
- Mokřady (4)
- Vodní plochy (5)

A jako první krok bylo vyzkoušeno, jak dopadne klasifikace bez jakýchkoliv úprav dat CLC pro území ve Švédsku.

Tab. 1: Kontingenční tabulka (Švédsko) bez úprav CLC

Třída	1	2	3	4	5	Počet pixelů
1	67,54	10,52	20,30	1,30	0,33	730883
2	13,59	65,27	17,68	3,24	0,22	5046981
3	4,31	3,96	77,59	13,48	0,65	18462414
4	0,96	2,57	11,42	83,48	1,57	710945
5	1,11	0,30	3,09	1,73	93,77	5188877

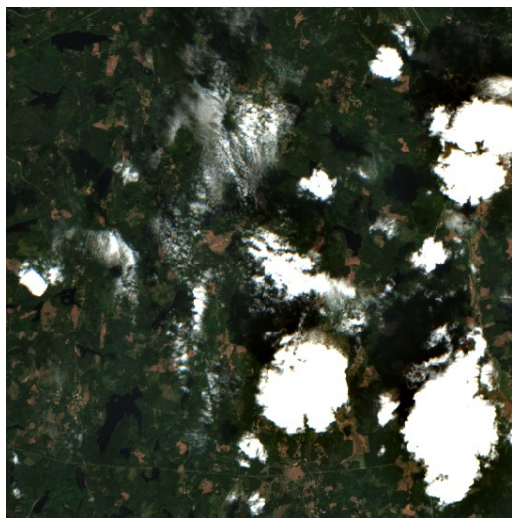
Poznámka: Údaje v tabulce (i ve všech následujících kontingenčních tabulkách), kromě informace o počtu pixelů, jsou v procentech.

Celková přesnost	78,21 %
Kappa index	0,65

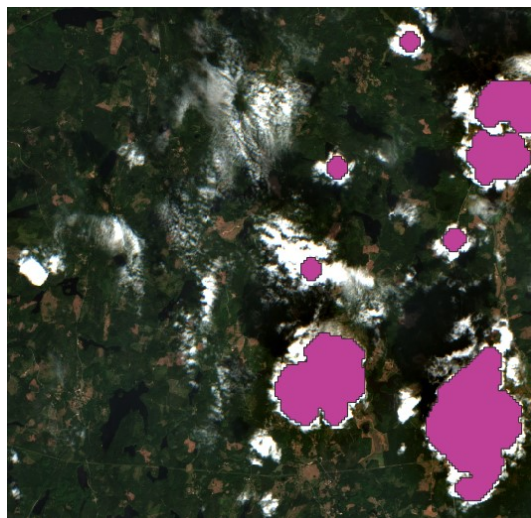
Jak je z kontingenční tabulky vidět, tak kromě vodních ploch výsledky nedosahují nijak závratných přesností. Především u zástavby a zemědělských oblastí je přesnost poměrně nízká a je vidět vzájemná záměna mezi těmito třídami. Obě třídy pak značnou částí zasahují i do třídy lesů a polopřírodních oblastí. Z výsledků tedy vyplývá, že je potřeba trénovací plochy jednotlivých tříd upravit.

Úprava dat CLC

Scény byly vybírány tak, aby byly bezoblačné. Takové scény ale nejsou vždy k dispozici a stejně to bylo i v tomto případě, kdy bylo pár míst pokrytých mraky tak, jako je tomu na Obr. 4. Nejedná se o velké plochy, ale i tak mohou výsledek klasifikace negativně ovlivnit. V souborovém adresáři obsahující jednotlivá pásma ze Sentinel-2 se nachází také soubor s názvem *MSK_CLOUDS_B00.gml*. Jedná se o vektorovou masku pokrývající oblačnost ve scéně. Na Obr. 5 je vidět, že maska ne vždy pokryje celou oblačnost. Proto je vhodné ji pomocí nástroje *Buffer* dostatečně rozšířit, aby oblačnost byla zakrytá co nejvíce. Následně byly tyto plochy z trénovacích dat odstraněny.



Obr. 4: Oblačnost ve scéně



Obr. 5: Maska oblačnosti

Dále bylo vhodné se podívat na jednotlivé podtřídy tvořící pět základních tříd CLC. Z výčtu podtříd tvořící základní třídu „Zástavba“ je patrné, že ne všechny podtřídy jsou vhodné jako trénovací plochy pro tuto třídu. Např. podtřída 1.4 „*Artificial, non-agricultural vegetated areas*“ se jistě jako trénovací množina bude zaměřovat s třídou lesů či zemědělských oblastí. Jako trénovací množiny pro třídu „Zástavba“ byly tedy použity pouze následující podtřídy: 1.1.1, 1.2.1, 1.2.2, 1.2.3 a 1.2.4. Zbylé podtřídy byly z trénovacích ploch odstraněny.

Stejně se postupovalo i u dalších tříd první úrovně. Pro trénovací množiny reprezentující „Zemědělské oblasti“ byly vybrány pouze podtřídy 2.1.1 a 2.3.1. U třídy „Lesy a polopřírodní oblasti“ bylo rozhodnuto, že pro klasifikaci první úrovně se tato třída zúží pouze na „Lesy“, tzn. že byly vybrány pouze podtřídy „Listnatý les“ (3.1.1) a „Jehličnatý les“ (3.1.2). Dalšími třídami jsou třídy „Mokřady“ a „Vodní plochy“. Ty se také dělí do různých podtříd, nicméně jejich různorodost není tak velká, a proto byly tyto třídy ponechány se všemi svými podtřídami.

Po výběru vhodných podtříd reprezentující jednotlivé základní třídy proběhla ještě jedna menší úprava. Tou byl tzv. vnitřní buffer, kterým byly odstraněny přechody mezi jednotlivými polygony různých tříd. Byla tedy použita funkce *Buffer* a nastavena hodnota -20 m, což odpovídá jednomu dvacetimetrovému pixelu. Poté byla klasifikace spuštěna znovu s následujícím výsledkem.

Tab. 2: Kontingenční tabulka (Švédsko) po základních úpravách

Třída	1	2	3	4	5	Počet pixelů
1	71,90	9,54	16,27	2,05	0,24	100386
2	7,62	78,77	11,95	1,61	0,05	3529294
3	2,90	3,45	86,31	6,82	0,51	15020855
4	0,97	2,20	15,50	79,89	1,44	634840
5	1,08	0,14	2,36	1,12	95,30	5033153

Celková přesnost	86,85 %
Kappa index	0,78

Z výsledků je vidět zlepšení přesnosti klasifikace (kromě mokřadů, které se poněkud překvapivě zhoršily), nicméně stále dochází k velké záměně mezi třídami. Bylo tedy zapotřebí podívat se na jednotlivé třídy hlouběji. Nejprve bylo potřeba zjistit, proč dochází k záměně mezi třídami. Na základě šetření byly zjištěny následující problémy a použita následující řešení.

Tab. 3: Úpravy trénovacích ploch

Třída	Problém	Řešení	Pravidlo
1. Zástavba	Vegetace v zástavbě	Odstranění všech pixelů vegetace	NDVI > 0,2
	Vodní plochy v zástavbě	Odstranění všech pixelů vody	NDVI < 0
2. Zemědělské oblasti	Holá půda v zemědělských oblastech	Odstranění všech pixelů holé půdy	NDVI < 0,2
	Možný výskyt zástavby a vodních ploch v zemědělských oblastech	Odstranění pixelů zástavby a vodních ploch (vyřešeno předchozím řešením)	NDVI < 0,2
	Stromová seskupení v zemědělských oblastech	Odstranění všech pixelů stromových seskupení	Smazání trénovacích ploch zemědělských oblastí klasifikovaných jako třída lesů
3. Lesy	Holá půda či komunikace v lesích	Odstranění všech pixelů holé půdy a komunikací	NDVI < 0,2
	Vodní plochy v lesích	Odstranění všech pixelů vodních ploch (vyřešeno předchozím řešením)	NDVI < 0
	Vykácená místa či paseky v lesích	Odstranění těchto nelesních pixelů	Smazání trénovacích ploch lesů klasifikovaných jako zemědělské oblasti
4. Mokřady	Záměna s třídou lesů	Odstranění pixelů lesa	Smazání trénovacích ploch mokřadů klasifikovaných jako třída lesů
5. Vodní plochy	Ostrovy menší než MMU	Odstranění pixelů neobsahující vodu	NDVI > 0 (> 0,1 pro oblast v Itálii)

Jak je z tabulky vidět, úpravy jsou založené převážně na tzv. NDVI neboli na normovaném rozdílovém vegetačním indexu. Jedná se o poměrový vegetační index a mezi vegetačními indexy je nejpoužívanější. Vyjadřuje nárůst hodnoty odrazivosti mezi vlnovou délkou



červeného a blízkého infračerveného záření, která je nejvyšší u zdravé zelené vegetace
Chyba! Nenalezen zdroj odkazů.. Nabývá hodnot <-1; 1> a vypočte se podle vzorce

$$NDVI = \frac{NIR - R}{NIR + R}, \quad (3)$$

kde

NIR jsou hodnoty blízkého infračerveného záření,

R jsou hodnoty červeného záření.

Hodnoty NDVI pro různé druhy povrchů jsou uvedeny v následující tabulce.

Tab. 4: Hodnoty NDVI [4]

Druh povrchu	NDVI
Hustá vegetace	0,500 – 1
Střední vegetace	0,140 – 0,500
Řídká vegetace	0,090 – 0,140
Holá půda	0,025 – 0,140
Oblačnost, led a sníh	-0,046 – 0,025
Voda	-1 – -0,046

Poznámka: Hodnoty NDVI se z různých zdrojů pro různé povrchy mírně liší a je potřeba je brát orientačně. Hodnoty NDVI mohou být ovlivněny např. množstvím srážek a z toho plynoucím výskytem vodních kapek na vegetaci, čímž se hodnota NDVI snižuje. V práci bylo vycházeno z hodnot udávaných ve zdrojích 0 a 0.

Výsledek klasifikace po úpravách zmíněných v Tab. 3 dopadl následovně.

Tab. 5: Kontingenční tabulka (Švédsko) po finálních úpravách

Třída	1	2	3	4	5	Počet pixelů
1	99,26	0,70	0,02	0,01	0,01	36618
2	0,84	97,16	0,09	1,90	0,00	2724293
3	0,32	2,38	95,91	1,34	0,05	13350020
4	0,61	3,32	0,19	95,25	0,63	484226
5	0,91	0,00	0,07	0,00	99,02	4009643

Celková přesnost	96,67 %
Kappa index	0,94



Z výsledků vyplývá, že díky provedeným úpravám bylo dosaženo vyšší přesnosti klasifikace. Klasifikace pro tuto scénu na první úrovni CLC mohla být tedy považována za dokončenou. Pro zbylé dvě oblasti byla klasifikace provedena stejným způsobem a budou zde ukázány výsledky před a po úpravách trénovacích ploch.

Tab. 6: Kontingenční tabulka (Česká republika) bez úprav CLC

Třída	1	2	3	5	Počet pixelů
1	65,60	14,59	12,85	6,96	2768028
2	8,33	79,99	10,60	1,08	18697470
3	3,09	3,63	92,22	1,07	8478948
5	6,31	1,92	5,71	86,06	193863

Celková přesnost	82,15 %
Kappa index	0,69

Tab. 7: Kontingenční tabulka (Česká republika) po úpravách CLC

Třída	1	2	3	5	Počet pixelů
1	99,91	0,01	0,00	0,08	135348
2	0,77	95,31	3,92	0,00	11589343
3	0,53	3,75	95,72	0,00	4835052
5	2,01	0,00	0,00	97,99	70718

Celková přesnost	95,48 %
Kappa index	0,90

Při klasifikování scény z České republiky stačilo v úpravách trénovacích ploch zahrnout pouze úpravy na základě NDVI. Úpravy typu „Smazání ploch z trénovacích množin klasifikovaných jako jiná třída“ nebylo nutné zavádět, protože nedocházelo k tak významné záměně tříd, jako tomu bylo ve scéně ze Švédska.

Tab. 8: Kontingenční tabulka (Itálie) bez úprav CLC

Třída	1	2	3	5	Počet pixelů
1	71,17	17,26	11,56	0,01	1813797
2	10,66	61,12	28,22	0,00	9944164
3	2,35	5,92	91,71	0,01	2482870
5	0,81	0,07	0,06	99,06	9081939

Celková přesnost	79,93 %
Kappa index	0,71



Tab. 9: Kontingenční tabulka (Itálie) po úpravách CLC

Třída	1	2	3	5	Počet pixelů
1	99,99	0,01	0,00	0,00	217757
2	1,23	98,72	0,04	0,00	3825558
3	0,15	2,30	97,55	0,00	1661716
5	0,54	0,00	0,00	99,46	8960214

Celková přesnost	99,06 %
Kappa index	0,98

Stejný postup byl následně použit i při klasifikování 2. a 3. úrovně CLC. Výsledky spolu s podrobnějším postupem jsou zveřejněny v autorově diplomové práci [6].

Strojové učení – klasifikátor Random Forest

Doposud byla pro všechny klasifikace použita metoda maximum likelihood. V současné době se však do popředí dostává tzv. strojové učení, a proto jsem i já vyzkoušel klasifikaci s využitím této metody, konkrétně s klasifikátorem random forest. V tomto případě byla ale klasifikace prováděna pouze na již upravených trénovacích datech. Ke klasifikaci pomocí strojového učení byl znovu použit program GRASS GIS.

Po prvním spuštění se ukázalo, že na mém počítači nejsem schopen klasifikaci spočítat při použití všech trénovacích ploch jako u maximum likelihood (výpočet skončil chybou, že se nástroj pokoušel alokovat příliš mnoho paměti). Bylo tedy potřeba z trénovacích ploch vybrat jen určitý vzorek, který by byl následně použit pro trénování. Jako celkový počet pixelů určených pro trénovací data klasifikace pomocí klasifikátoru random forest byla zvolena hodnota 2 500 000.

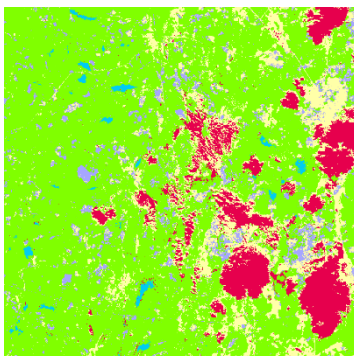
Při klasifikaci pomocí klasifikátoru random forest se téměř u všech scén a všech úrovní CLC klasifikovalo 100 % trénovacích pixelů do správných tříd. S největší pravděpodobností zde tak došlo k přetrénování, což znamená, že klasifikátor ztratil schopnost generalizovat a funguje tedy dobře pouze na vybraných trénovacích plochách, případně na daném území, na kterém byl natrénován. Otázkou zůstává, jak moc jsou tyto výsledky směrodatné a jak by se výsledné přesnosti změnily, kdyby byly použity všechny trénovací plochy a nejen jejich část. Nicméně i tak tato klasifikace poskytla zajímavé výsledky.

Porovnání výsledků

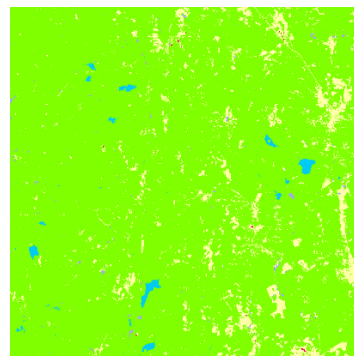
Výsledky obou klasifikátorů byly porovnány pomocí vizuální kontroly. Klasifikátor random forest si dokázal lépe poradit s oblačností ve scéně, a i přes její výskyt dokázal dané území pokryté oblačností správně klasifikovat (výhoda použití dvou scén z různého časového období). Oproti tomu klasifikátor maximum likelihood klasifikoval tyto oblasti jako zástavbu.



RGB scéna



Maximum likelihood

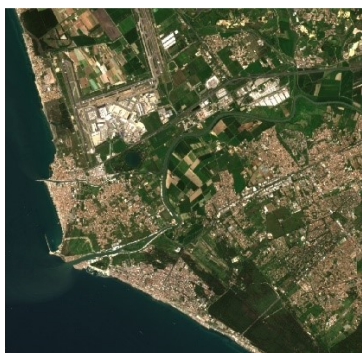


Random forest

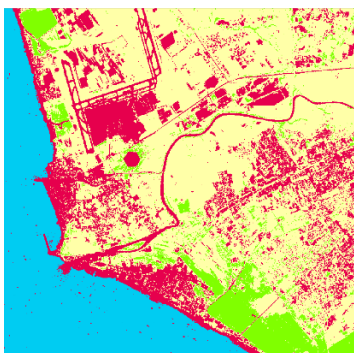
červená – zástavba, světle žlutá – zemědělské oblasti,
zelená – lesy, modrošedá – mokřady, modrá – vodní plochy

Obr. 6: Porovnání klasifikace s oblačností (Švédsko, 1. úroveň CLC)

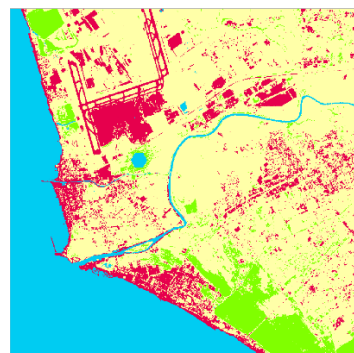
Lepších výsledků dosáhl klasifikátor random forest i při klasifikaci u řeky Tibery, kterou se oproti maximum likelihood podařilo na první úrovni CLC správně klasifikovat jako vodní plochu.



RGB scéna



Maximum likelihood

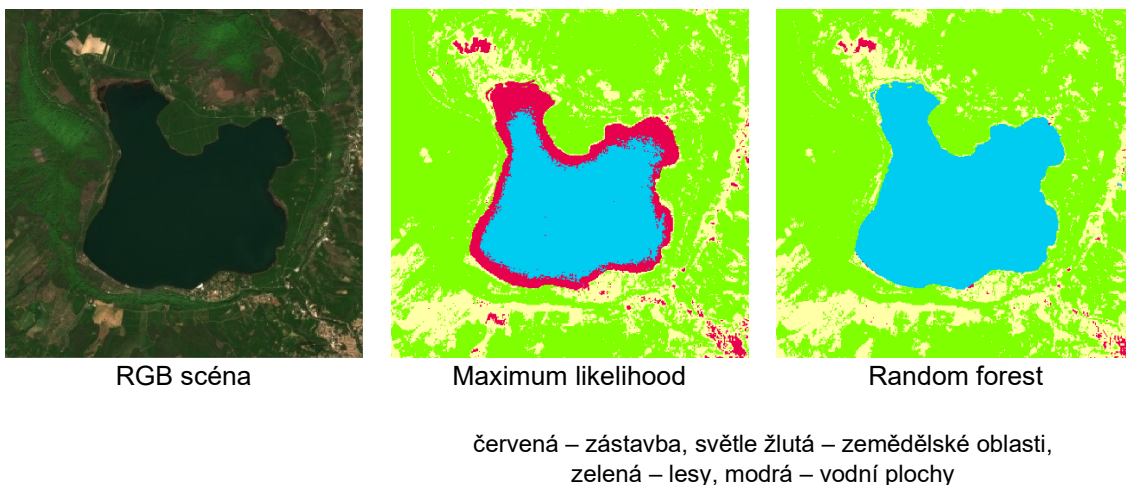


Random forest

červená – zástavba, světle žlutá – zemědělské oblasti,
zelená – lesy, modrá – vodní plochy

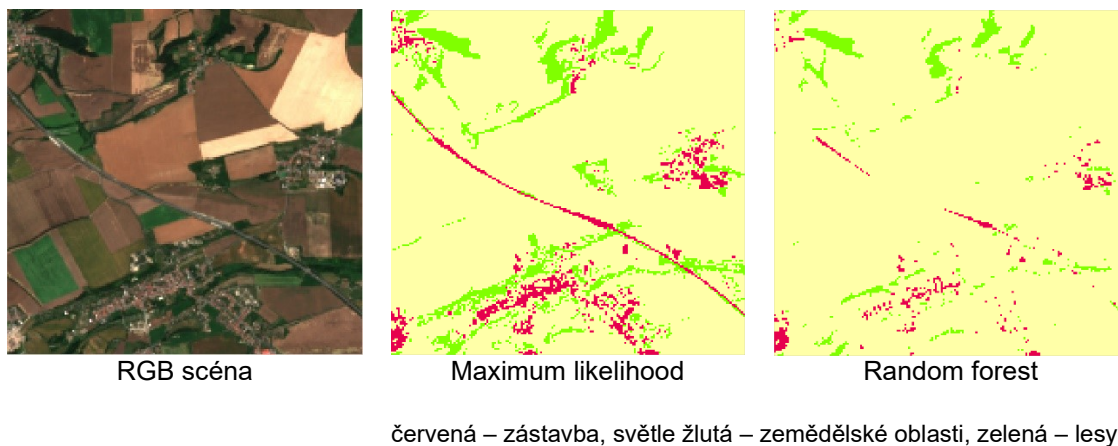
Obr. 7: Porovnání klasifikace Tibery při první úrovni CLC

Lépe si random forest poradil i při klasifikování jezer na první úrovni CLC, u kterých docházelo k nějakému znečištění, zákalu vody nebo většímu výskytu vodních rostlin či řas. V tomto případě se ale nejpravděpodobněji jedná o místa s mělkou vodou, kde naměřené hodnoty nejsou pouze výsledkem odrazivosti vody, ale i dna.



Obr. 8: Porovnání klasifikace jezer (Itálie, 1. úroveň CLC)

Na druhé úrovni CLC již k takovýmto chybám nedochází a řeky i jezera jsou správně klasifikovány i metodou maximum likelihood. Kde maximum likelihood poskytovalo lepší výsledky, jsou dopravní komunikace a zástavba. To bude ale nejspíš dáno počtem trénovacích ploch, který byl u těchto tříd pro random forest poměrně nízký, a nebyl tak dostatečně reprezentující.



Obr. 9: Porovnání klasifikace dopravní komunikace (ČR, 1. úroveň CLC)

ZÁVĚR

Cílem práce bylo klasifikovat tři vybraná evropská území s použitím dat CORINE jako trénovací množiny, upravit tyto data tak, aby je bylo možno pro trénovací množiny použít a zjistit, které třídy z nomenklatury CORINE lze úspěšně klasifikovat na základě spektrálních příznaků. Všechny tyto cíle byly úspěšně splněny. Klasifikace pro všechny lokality byla provedena na třech úrovních tematické podrobnosti, kdy první úroveň obsahovala 4–5 tříd a třetí úroveň až 28. Pro klasifikaci byla použita metoda maximum likelihood.



Všechny třídy první úrovně byly klasifikovány s přesností vyšší než 95 %. Bylo tomu i u třídy 4 „Mokřady“, u které se předpokládaly s klasifikací problémy. Úspěšná byla i klasifikace druhé úrovně, ze které byly nejprve odebrány dvě nehomogenní třídy. Jednalo se o třídy 1.4 „Umělé nezemědělské oblasti s vegetací“ a 2.4 „Různorodé zemědělské oblasti“. Ze zbylých tříd se následně hůře klasifikovala pouze třída 1.2 „Průmyslové, obchodní a dopravní jednotky“. U oblasti z Itálie se s nižší přesností klasifikovaly i třídy 4.1 „Vnitrozemské mokřady“ a 4.2 „Přímořské mokřady“. Těchto tříd se ale v dané scéně nachází pouze minimum.

Nejnáročnější částí byla úprava dat třetí úrovně CLC a následná klasifikace daných oblastí na této úrovni. Ukázalo se, že i na třetí úrovni CLC je možné většinu tříd klasifikovat s přesností vyšší než 85 %. Třídy s nižší přesností se pro různé oblasti liší. Pro oblast ve Švédsku to jsou třídy 4.2.3 „Přílivové oblasti“, kterou ale tvoří jediný polygon, a 3.1.3 „Smíšený les“, u které byla přesnost klasifikace nižší než 50 %, a nebyla tedy snaha ji následně klasifikovat ve zbylých dvou scénách. Nižších přesností dosáhli také třídy 1.2.1 „Průmyslové nebo obchodní jednotky“ a 1.2.3 „Přístavy“. Třída 1.2.1 byla klasifikována s nižší přesností na všech scénách. Na oblasti z České republiky dosáhly kromě třídy 1.2.1 nižší přesnosti třídy 1.2.2 „Silniční a železniční sítě a přidružené pozemky“, 3.2.4 Přechnodné lesnaté či keřovité oblasti a 4.1.2 „Rašelinisté“. Oblast z Itálie dosáhla nižší přesnosti pouze u třídy 1.2.1.

Kromě metody maximum likelihood byla vyzkoušena metoda random forest založená na strojovém učení. Klasifikaci bohužel nebylo možno spustit na celé množině trénovacích dat, a proto bylo náhodně vybráno celkem 2 500 000 pixelů. Jednotlivé třídy v nich byly úměrně zastoupeny tak, aby jejich zastoupení odpovídalo rozdělení tříd v původních trénovacích datech. Tímto způsobem byly následně obdrženy zajímavé výsledky a to takové, že prakticky u všech tříd bylo dosaženo 100% přesnosti klasifikace. Výsledky by bylo ale nejspíš vhodné dále zkoumat, případně pro klasifikaci použít větší množství trénovacích dat a sledovat, zda se budou výsledky klasifikace měnit, protože v tomto případě zde došlo pravděpodobně k přetrérování.

Dalším krokem, který bylo zamýšleno udělat, byla validace dosažených výsledků na testovacích množinách. Pro validaci se chtěly použít data OSM (OpenStreetMap). Ukázalo se ale, že k tomu, aby bylo možné tyto data k validaci použít, je potřeba data OSM dosti upravit (přeložení nomenklatury OSM do CORINE, buffer liniových prvků, výběr vhodných prvků pro validování rastrových dat o velikosti pixelu 20 m × 20 m a další), což bohužel nebylo v silách v čas stihnout.

Klasifikována byla celkem tři evropská území s rozdílnými klimatickými podmínkami a biotopy. Evropu tvoří ale více takovýchto oblastí a pro úspěšné klasifikování celé Evropy bude nutno zpracovat další scény. Pro úspěšnou klasifikaci je potřeba převzaté trénovací množiny z dat CORINE náležitě upravit, jelikož data CLC jsou omezeny minimální mapovací jednotkou 25 ha, čímž dochází k významné agregaci spektrálně odlišných prvků. Při zavedení úprav na základě NDVI a odstranění záměny mezi třídami tak přesnost klasifikace narostla o desítky procent.

PODĚKOVÁNÍ

Děkuji paní prof. Ing. Leně Halounové, CSc., za odborné vedení této práce. Dále děkuji panu Ing. Lukáši Brodskému, Ph.D., za užitečné rady a konzultace, díky nimž mohla tato práce vzniknout a panu Ing. Martinu Landovi, Ph.D., za pomoc při řešení technických problémů v programech QGIS a GRASS.



REFERENCE

- [1] Iniciační fond fakulty stavební ČVUT - Aktuality | FSV ČVUT [online]. [cit. 2020-04-13]. Dostupné z: <https://web.fsv.cvut.cz/aktuality/490/>
- [2] Corine Land Cover Classes. In: [cit. 2020-05-01]. Dostupné z: https://land.copernicus.eu/Corinelandcoverclasses.eps.75dpi.png/image_view_fullscreen
- [3] HALOUNOVÁ, L.: Zpracování obrazových dat. ČVUT v Praze, 2008. ISBN: 978-80-01-04253-3
- [4] Kafira, Vasiliki & Albanakis, K. & Oikonomidis, Dimitrios. (2015). Flood Susceptibility Assessment using G.I.S. An example from Kassandra Peninsula, Halkidiki, Greece.
- [5] HASHIM, H., Z. ABD LATIF a N. A. ADNAN. Urban vegetation classification with NDVI threshold value method with very high resolution (vhr) pleiades imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* [online]. 2019, roč. XLII-4/W16, s. 237-240. ISSN 2194-9034. DOI: 10.5194/isprs-archives-XLII-4-W16-237-2019
- [6] BOUČEK, Tomáš. *Testování způsobů klasifikace pokrytí území vybraných evropských oblastí*. Praha, 2020. Diplomová práce. ČVUT v Praze. Vedoucí práce prof. Ing. Lena Halounová, CSc.