ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA PROSTOROVÝCH VĚD

Modelování výšky lesa z kombinace družicových dat GEDI a Sentinel

DIPLOMOVÁ PRÁCE

Vedoucí práce: Jiří Prošek
Diplomant: Eliška Šárovcová

2022

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Eliška Šárovcová

Inženýrská ekologie
Ochrana přírody

Název práce

Modelování výšky lesa z kombinace družicových dat GEDI a Sentinel

Název anglicky

Modeling of forest canopy height using integration of GEDI and Sentinel satellite data

---

Cíle práce

Cílem diplomové práce je modelování výšky lesních porostů pomocí statistických modelů založených na datech Dálkového průzkumu Země. Primárním zdrojem proměnných pro modelování budou zejména multispektrální data z družicové platformy Sentinel – 2. Pro trénování a validaci modelů budou využita LiDAR data z družicové platformy GEDI.

Specifikace dílčích cílů DP je úkolem studentky na základě průběžných konzultací.

Metodika

1. Rešeršní část práce se zaměří především na rozbor použitých datových vstupů (družicových platforem), jejich intepretaci a možné použití. Dále bude shrnuta problematika statistického modelování v prostorových vědách.

2. Praktická část práce bude spočívat v modelování výšky lesních porostů pomocí statistických modelů. Součástí praktické části DP bude zpracování a příprava datových vstupů, modelování výšky lesních porostů a validace modelů na zvolených studijních oblastech.

Doporučený rozsah práce

30 až 50 stran

Klíčová slova

DPZ, CHM, LiDAR

Doporučené zdroje informací

Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., & Chen, H. (2020). High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. International Journal of Applied Earth Observation and Geoinformation, 92, 102163.

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. Journal of Geophysical Research: Biogeosciences, 116(G4).

Wang, Y., Li, G., Ding, J., Guo, Z., Tang, S., Wang, C., ... & Chen, J. M. (2016). A combined GLAS and MODIS estimation of the global distribution of mean forest canopy height. Remote Sensing of environment, 174, 24-43.

Předběžný termín obhajoby

2021/22 LS – FZP

Vedoucí práce

Ing. Jiří Prošek

Garantující pracoviště

Katedra prostorových věd

Elektronicky schváleno dne 21. 3. 2022

doc. Ing. Petra Šímová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 22. 3. 2022

prof. RNDr. Vladimír Bejček, CSc.

Děkan

V Praze dne 28. 03. 2022

Čestné prohlášení

Prohlašuji, že jsem diplomovou práci na téma: Modelování výšky lesa z kombinace družicových dat GEDI a Sentinel vypracovala samostatně a citovala jsem všechny informační zdroje, které jsem v práci použila a které jsem rovněž uvedla na konci práce v seznamu použitých informačních zdrojů.

Jsem si vědoma, že na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů, především ustanovení § 35 odst. 3 tohoto zákona, tj. o užití tohoto díla.

Jsem si vědoma, že odevzdáním diplomové práce souhlasím s jejím zveřejněním podle zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů, ve znění pozdějších předpisů, a to i bez ohledu na výsledek její obhajoby.

Svým podpisem rovněž prohlašuji, že elektronická verze práce je totožná s verzí tištěnou a že s údaji uvedenými v práci bylo nakládáno v souvislosti s GDPR.

V Praze dne ..……

...……...

podpis autorky práce

Poděkování

Děkuji vedoucímu práce za odborné vedení a věcné připomínky.

Abstrakt

Tato práce se zabývá tvorbou modelu pro extrapolaci výšky vegetace naměřenou satelitem GEDI na základě radarových a multispektrálních dat pořízených družicemi Sentinel-1, respektive Sentinel-2. Zájmovou oblastí jsou tři pohoří táhnoucí se po obvodu České republiky – Šumava, německá strana Krušných hor a česká a polská strana Krkonoš. Pro trénování modelu byla použita data výšky vegetace v podobě datového produktu GEDI – L2A verze 002. Jako vysvětlující proměnné byla použita data Sentinel stažená přes platformu Google Earth Engine. V práci bylo s daty Sentinel nakládáno v podobě buněk 100x100 metrů. Pokud do jedné buňky spadalo více měření GEDI, byla v rámci ní tato měření zprůměrována. Mimo základních spektrálních pásem jako NIR, SWIR, RGB a Red edge byly také vypočítány vegetační indexy NDVI, NDMI a EVI a jako další potenciální vysvětlující proměnné byla brána radarová měření (VV a VH). Modely byly trénovány na 70 % dat (52 544 měření) a ve vybraném nejlepším modelu byly vysvětlujícími proměnnými Red edge, NDMI, směrodatná odchylka z letního měření VV, letní medián měření VV a EVI. Tento model celkem vysvětlil 52 % variability a následně jím byly predikovány hodnoty pro zbylou (testovací) část dat, výsledná validace byla provedena pomocí Pearsonova korelačního koeficientu (0,7) a metrik chyb RMSE (7,60 m), ME (-0,06 m) a jejich relativních alternativ %RMSE (40,66 %) a %ME (-0,33 %.). RMSE dosažená v této práci je oproti jiným obdobným pracím přibližně jednou tak vysoká, ME je naopak významně nižší. Může to být dáno různým prostředím v jednotlivých zájmových územích, nezanesením informace o topografii a typu vegetace do modelu, měřítkem či zvolenou vysvětlovanou proměnnou (kvantil 98 % výšky vegetace). Je na dalším výzkumu, aby bylo objasněno, proč zde vznikly chyby tak odlišné od srovnatelných studií.

Klíčová slova: LiDAR, Multispektrální snímání, Vegetační indexy, Boosted Regression Trees

Abstract

This thesis deals with the validation of the model for extrapolation of the canopy height measured by the GEDI satellite based on radar and multispectral data acquired by satellites Sentinel-1 and Sentinel-2, respectively. The study area consists of three mountainous regions around the borders of the Czech republic – Bohemian Forest, the German side of the Ore Mountains, and both Czech and Polish sides of the Giant Mountains. For the training of the model data of the GEDI data product L2A version 002 were used. Sentinel data were processed in a resolution of 100x100 meters. In case of multiple GEDI measurements in one grid cell, the value was calculated by the average of those values. Besides basic main spectral bands such as NIR, SWIR, RGB, and Red edge several vegetation indices were calculated such as NDVI, NDMI, and EVI. Radar measurements (VV and VH) were considered as potential explanatory variables as well. The models were trained with 70 % of the data (i.e. 52 544 measurements) and in the best-fitted model Red edge, NDMI, a standard deviation of summer VV measurements, a median of the summer VH measurements, and EVI were its explanatory variables. The chosen model explained 52 % of the variability and then it was used to predict values for the testing part of the dataset. Final validation was performed with Pearson correlation coefficient (0,7) and standard error metrics as RMSE (7,60 m), ME (-0,06 m), and their alternatives %RMSE (40,66%) and %ME (-0,33%). The value of RMSE achieved in this thesis is approximately twice as big as in similarly focused papers, ME is on the contrary significantly smaller. It may have been caused by the variable environment in individual regions of interest, by not including topography and type of vegetation in the model, scale, or chosen explained variable (98 % quantile of canopy height). It is on the subsequent works to explain why these errors were so distant from the values in compared studies.

Keywords: LiDAR, Multispectral imaging, Vegetation indices, Boosted Regression Trees

Obsah

1. Úvod a cíle práce 1
2. Literární rešerše 3
2.1 GEDI 3
2.2 Sentinel-1 4
2.3 Sentinel-2 6
2.4 Vegetační indexy 6
2.5 Strojové učení v prostorových vědách 9
3. Metodika 11
3.1 Zájmová území 11
3.1.1 Šumava 12
3.1.2 Krušné hory 13
3.1.3 Krkonoše 14
3.2 Sběr a příprava dat 15
3.3 Zpracování dat 17
3.4 Hodnocení přesnosti 20
4. Výsledky 23
4.1 Model 23
4.1.1 Vysvětlující proměnné 23
4.2 Korelace 26
4.3 RMSE a ME 27
5. Diskuze 30
6. Závěr 37
7. Reference 38

1. Úvod a cíle práce

V posledních dvou dekádách dochází k výraznému rozmachu jak dálkového průzkumu Země a s ním spojeného výzkumu souvisejících metod (LiDAR, RADAR apod.), tak umělé inteligence. Současně je realizováno stále více projektů, kde jsou tato dvě odvětví spojována a společně využívána za účelem vytěžení více informací, než je možné pouhým pozorováním – ať už kvůli efektivnějšímu využívání financí či jen snaze vývoj technologií posunout. Propojení těchto oborů je zároveň vcelku logickým krokem a vzchází z něj mnoho nových možností. Aktuálně je možné satelitním snímáním získat mnoho dat o povrchu, terénu, vegetaci a dalších atributech Země. Avšak přinejmenším snaha o satelitní „skenování" LiDARem je stále výrazně limitována frekvencí snímání a neposkytuje tak dokonale souvislá data. Zároveň se i technologie snímání stále posouvají, a tak jsou do vesmíru průběžně vypouštěny výkonnější a dokonalejší instrumenty. Možnost získání rozsáhlejších (spojitých) dat o Zemi z vesmíru by přispělo mnoha vědeckým odvětvím od batymetrie přes zkoumání lesních ekosystémů k mapování terénu a povrchu v jinak nedostupných (např. horských) oblastech. Jako příklad „užitečných" dat, která zatím nejsou ve velké míře dostupná lze uvést výšku lesní vegetace. Ta je perspektivním ukazatelem, který se dá v dosti přesné podobě získat z leteckého laserového skenování, avšak vzhledem k rozsahu některých porostů by to bylo značně nákladné. Plošně se tedy ve velké míře až na družicové snímání běžně neměří, nicméně se z této veličiny dá odvodit mnoho dalších faktorů. Výška vegetace je například podstatným indikátorem množství uhlíku uloženého v lese, dále z ní lze odvozovat produktivitu a v zásadě i biodiverzitu stanoviště (Li et al., 2020).

Modely jsou jednou z možností k zprostředkování spojitých dat a je s jejich pomocí možné predikovat mnoho proměnných z různých oborů. Jelikož při vytváření modelů jde vlastně o zjednodušení vybraného fenoménu, snahu o napodobení jeho průběhu (kvantifikaci komplexního systému) a jeho převedení na jednodušší a předvídatelnější reprezentaci, nelze přesnost těchto predikcí nijak garantovat. Součástí procesu tvorby modelu tak například může být odstranění nadbytečných komponentů přidávajících systému variabilitu a zpravidla jsou do něj ve výsledku zaneseny jen nejvýznamnější proměnné, tedy ty, které jsou pro zachycení predikovaného jevu nejpodstatnější a nejvýznamnější (Imam, 2019).

Tato práce se zabývá tvorbou modelu, s jehož pomocí by bylo možné extrapolovat měření LiDARu GEDI, který byl v roce 2018 umístěn na ISS (International Space Station). Jako vysvětlující proměnné, pomocí kterých extrapolace probíhala byla použita data získaná satelity Sentinel-1 a Sentinel-2, které byly postupně vypouštěny mezi lety 2014 a 2017. Cílem je predikce výšky vegetace v zalesněných oblastech Šumavy, německé strany Krušných hor a české a polské strany Krkonoš. V rámci práce byl vybrán nejlepší model s dostupnými proměnnými a následně budou jeho predikce validovány metrikami ME (Mean Error), RMSE (Root Mean Square Error) a pomocí výpočtu korelačního koeficientu s měřením GEDI. Z dosavadních vědeckých prací se jich mnoho nezabývalo extrapolací výšky vegetace pomocí předpovědních modelů, což nejspíše lze vnímat jako nevytěžený potenciál.

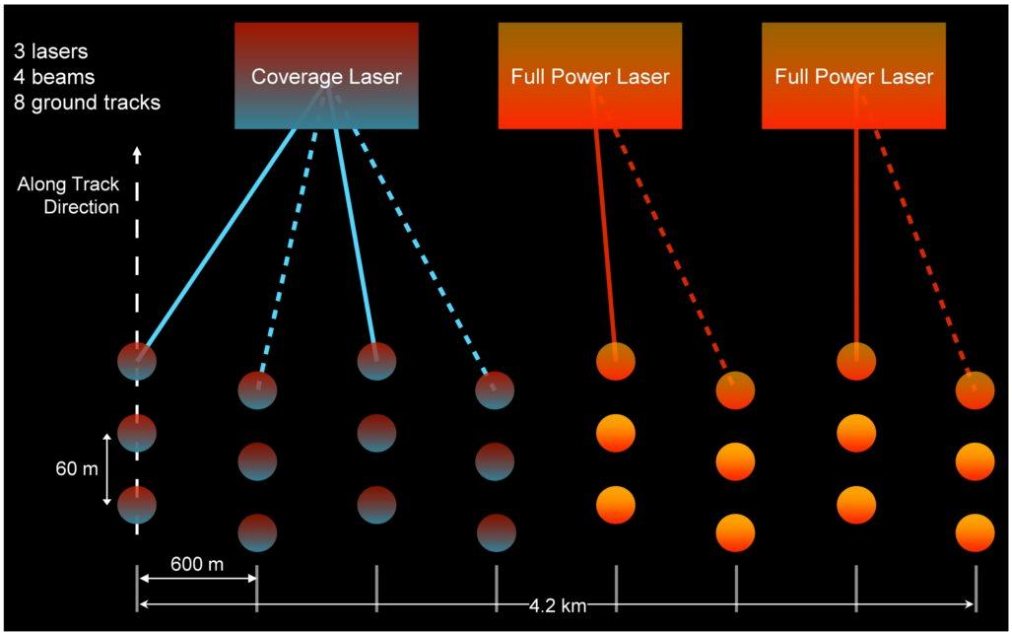
Diplomová práce se zabývá tvorbou modelu predikujícího výšku vegetace z měření satelitu GEDI a následnou evaluací vytvořených předpovědí. K přípravě modelu byla jako vysvětlující proměnné použita data Sentinel-1 a Sentinel-2. Byly stanoveny následující dílčí cíle: výběr nejvýznamnějších spektrálních pásem jako vysvětlujících proměnných a vytvoření modelu s co nejvyšším podílem vysvětlené variability (1) a validace a vyhodnocení přesnosti modelu pomocí chybových metrik a korelace s nezávislými daty naměřenými GEDI (2).

2. Literární rešerše

2.1 GEDI

Projekt GEDI (The Global Ecosystem Dynamics Investigation) začal v listopadu roku 2014. Jeho laserový systém je původně návrhem Marylandské univerzity (Cole et Northon, 2014), která v červenci téhož roku vyhrála soutěž NASA (National Aeronautics and Space Administration) o nástroj k mapování vlivu klimatických změn na vegetaci, který měl být následně umístěn na Mezinárodní vesmírnou stanici (ISS) a z ní několik let operovat. Do počátku roku 2018 byl sestavován a dále 5. 12. 2018 umístěn na ISS. Přibližně po čtyřech měsících skončilo testovací období a od té doby konstantně sbírá data (NASA, 2021a).

GEDI je prvním nástrojem mapujícím výšku povrchu, výšku vegetace a její vertikální strukturu ve vysokém rozlišení (NASA, 2021a). To poskytuje lepší možnosti pro studium biodiverzity, habitatů a procesů v rámci koloběhu uhlíku na Zemi. Mimo to je možné data využít pro monitorování ledovců, v lesním hospodářství a při předpovědi počasí. Data získaná GEDI lze kombinovat s datovými sadami z jiných senzorů (NISAR, ICESat-2, TanDEM-X) a rozšířit tak možnosti pozorování Země (NASA, 2021a).



Obrázek 1: Schéma stopy měření GEDI na povrchu Země (NASA, 2021e).

Charakteristika laserů

GEDI je LiDAR, jedná se tedy o aktivní snímání povrchu, jsou při něm emitovány fotony a na základě času jejich návratu je vypočítána vzdálenost od odrazového bodu,

z té jsou následně odvozeny datové produkty. Z distribuce laserové energie nad zemí lze odvodit výšku a hustotu objektů v rámci stopy (footprint) satelitu na zemi (NASA, 2021b). GEDI má tři lasery, které dohromady emitují čtyři paprsky, které jsou dále rozdělené a vytváří tak osm pozemních drah (ground track spot beams), čtyři silné a čtyři „slabé“ (NASA, 2021c). Jejich vlnová délka je 1064 nm, čímž spadají do části spektra infračerveného světla (Leonardo technology, 2015). Velikost stopy na Zemském povrchu je 25 m. Jejich příčná vzdálenost je 600 metrů a podélná 60 metrů, celá šíře dráhy LiDARu na Zemském povrchu tak zabere 4,2 km na šířku (viz schéma na obr. 1).

Datové produkty

NASA z dat získaných GEDI vytváří datové produkty, které jsou volně dostupné[1]. Nabízí několik úrovní předzpracování dat (Level 1 až Level 4). Datová sada první úrovně zpracování obsahuje surová data (waveforms) opatřená prostorovým umístěním (geolokací). Na druhé úrovni (L2) jsou výstupem výšky vegetace a metriky profilu, z těchto dat se například dá již dobře interpretovat vertikální struktura vegetace. Třetí úroveň (L3) zpracování obsahuje v mřížce odhady z L2 dat vytvořené prostorovou interpolací. Nejvyšší úrovní zpracování jsou data L4, která jsou vytvořena modelem, z L2 produktů jsou vytvořeny odhady hustoty nadzemní biomasy a z ní odvozené další metriky (průměr, chyba) (NASA, 2021d).

2.2 Sentinel-1

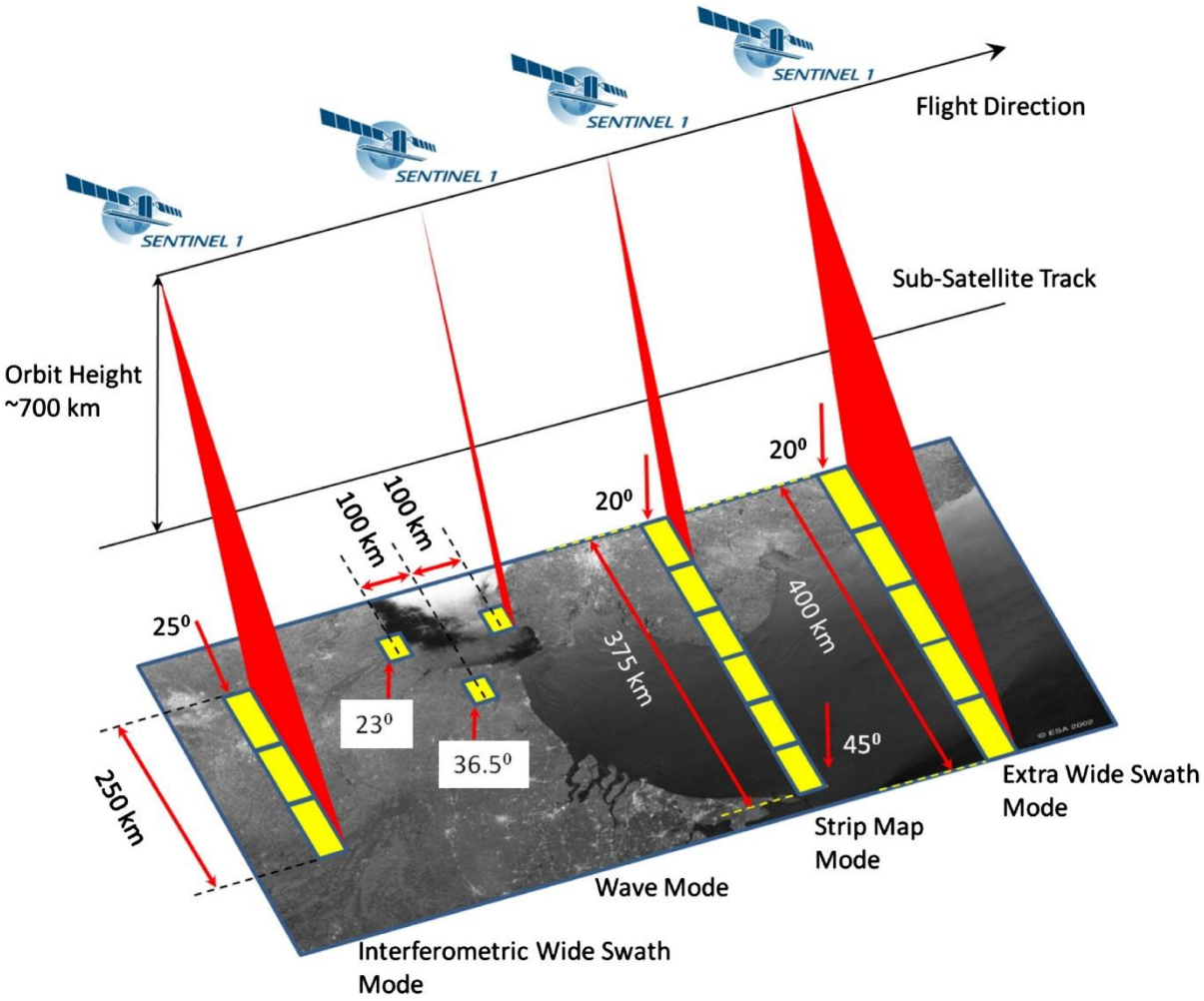
Mise Sentinel vytvořila Evropská vesmírná agentura (ESA – European Space Agency) pro program Copernicus (pozorování Země a jejího prostředí v zájmu obecného užitku) (Copernicus, 2021). Satelit Sentinel-1A byl vypuštěn třetího dubna roku 2014, o dva roky později ho následoval Sentinel-1B. Mise Sentinel-1 tedy sestává ze dvou satelitů na polární orbitě vybavených radarem. Získaná data je možné uplatnit v mnoha odvětvích – například při monitorování ledovců, mořského prostředí (detekce ropných havárií) a mapování nakládání s lesy či půdou (Sentinel Online, 2021).

Charakteristika snímání

Sentinel-1 nese radar se syntetickou aperturou – SAR (C-band Synthetic Aperture Radar), který vysílá mikrovlny (Herndon et al., 2020). Má několik módů snímání

[1] https://search.earthdata.nasa.gov/

(každý má jiný účel, liší se především úhlem vůči zemi a šířkou pozemní stopy paprsků – „řádky“; viz obr. 2) (Sentinel Online, 2021). Wave mód je určen pro snímání vln v oceánu (cílem je ve spojení s globálními modely oceánských vln určit jejich směr, délku a výšku), měření se skládá ze tří snímků o rozměrech 20x20 kilometrů, které se pořizují každých 100 kilometrů „cesty“. Mód Stripmap je určený pro kontinuální monitorování, v úzkém záběru (řádku) o šířce 80 kilometrů snímá v rozlišení 5x5 metrů. Interferometric Wide Swath mód se obecně používá pro pozorování Země, má výsledné rozlišení 5x20 metrů a snímá tři dílčí řádky o šíři 250 kilometrů. Extra Wide Swath mód je určen především pro oblasti, kde je třeba velké pokrytí a krátká doba opakování (přímořské, polární atd.), jeho rozlišení je 20x40 metrů. Datové produkty z této mise jsou také zpracovávané a následně dostupné dle zmíněných módů snímání. Opakovací cyklus za využití obou družic je 6 dní (Sentinel Online, 2021).



Obrázek 2: Schéma snímacích módů družice Sentinel-1 (ESA, 2021).

2.3 Sentinel-2

Obdobně jako Sentinel-1 i tato mise sestává ze dvou polárně orbitujících družic. Sentinel-2A byl vypuštěn 23. června 2015 a Sentinel-2B následně 7. března 2017, oba z kosmodromu v Kourou ve Francouzské Guayaně (Sentinel Online, 2021). Hlavním záměrem mise je především pozorovat proměnlivost podmínek na zemském povrchu, konkrétní oblasti využití dat jsou například bezpečnost, klimatická změna a podmínky na moři (Sentinel Online, 2021). Družice Sentinel-2 poskytují multispektrální data s vysokým rozlišením a krátkou dobou opakování a zároveň i data pro další úroveň produktů jako jsou mapy land cover, mapy detekce změny využití půdy a geofyzikálních proměnných (Sentinel Online, 2021).

Charakteristika snímání

Družice z mise Sentinel-2 nesou multispektrální instrument (MultiSpectral Instrument – MSI), který snímá ve třinácti spektrálních pásmech. V rozlišení 10 metrů jsou snímána pásma viditelného světla (červená, zelená a modrá) a viditelného a blízkého infračerveného světla (VNIR, 842 nm), tato pásma (bands) jsou využitelná pro klasifikaci využití ploch (land cover). Dále v rozlišení 20 metrů opět viditelné a blízké infračervené světlo (VNIR, 865 nm, 783 nm, 740 nm a 705 nm), krátkovlnné infračervené (SWIR, 1610 nm a 2190 nm), tyto bandy slouží především k monitorování stavu vegetace. Data získávaná v rozlišení 60 metrů jsou mimo jiné určena k zachycení prostorové variability atmosférických geofyzikálních parametrů, je snímáno pásmo krátkovlnné infračervené (SWIR, 1375 nm a 940 nm), vodní pára a pobřežní aerosol (Ultra blue, 443 nm) (Doron, 2019).

Šířka orbitu je 290 kilometrů a doba, za kterou jeden ze satelitů obletí orbit a vrátí na původní místo, je 10 dnů (nevstoupí-li do snímání oblačnost, je pomocí obou satelitů stejné místo snímáno jednou za pět dnů) (Sentinel Online, 2021). Z mise Sentinel-2 jsou dva typy datových produktů – úroveň 1C (vrchol atmosféry, TOA) a 2A (spodek atmosféry, BOA), jsou volně dostupné po „dlaždicích“ mřížky, do kterých je rozdělen Zemský povrch.

2.4 Vegetační indexy

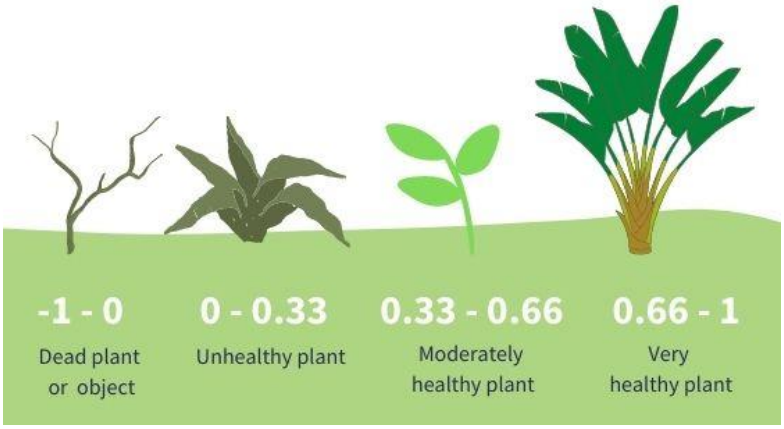
Multispektrální data přináší možnost hlubší analýzy snímků a s tím i nové pohledy na povrch. Nad jednotlivými snímky různých vlnových délek lze provést různé (i zcela jednoduché) matematické operace, jejichž výstupem je například zdůraznění

některých prvků samostatně neklasifikovatelných z vrstev jednoho typu snímání. Tímto způsobem jsou hojně využívány vegetační indexy. Výsledkem jsou nové vrstvy, které poskytují další informace o vegetaci (zdravotní stav, vlhkost, obsah dusíku atd.). Uplatnění tedy nachází především v environmentálním monitorování, zemědělství, lesnictví a ochraně biodiverzity. V rámci snímání vegetace jsou v rámci elektromagnetického klíčové oblasti ultrafialového záření, viditelného světla a blízké a střední infračervené světlo. Je známo, že míra odrazivosti světla od listu se liší s jeho obsahem vody a při zahrnutí infračerveného spektra je navíc možné hodnotit obsah cukrů, pigmentů proteinů a dalších složek v rostlině (Batten, 1998). Proces validace vegetačních indexů zpravidla sestává z vyhodnocení korelace mezi vegetačním indexem a stejnými charakteristikami měřenými na konkrétní lokalitě (vegetační pokryv, biomasa atd.). Vegetační indexy tedy obecně zvýrazňují množství vegetace, dále by měly redukovat efekty atmosféry a minimalizovat efekty půdy na pozadí(jež mohou výrazně zkreslit výsledky v oblastech, kde je vegetace málo). Před volbou vegetačního indexu je vždy nutné nejprve zvážit jeho výhody a omezení, protože každý index je vhodný k jiným účelům. Zároveň i různá prostředí mají své vlastní charakteristiky a proměnné, které je třeba zhodnotit z hlediska vhodnosti vybraného vegetačního indexu v konkrétním místě (Xue et Su, 2017). Takto lze jejich použití mnohdy přizpůsobit konkrétním aplikacím. Vegetační indexy lze rozdělit do několika kategorií dle zájmu pozorování – Širokopásmová zeleň (nejjednodušší měření množství a vitality porostu), Úzkopásmová zeleň (obdobný princip jako Širokopásmová zeleň, jsou určeny k použití se spektrometry), Efektivita využití světla (souvisí s vegetativním růstem a frakční absorpcí fotosynteticky aktivního záření, jejich pomocí se tedy odhaduje produkce a rychlost růstu), Dusík v porostu (dusík je významnou částí chlorofylu a ve vysoké koncentraci je přítomen ve vegetaci, která rychle roste), Uhlík v suché nebo stárnoucí vegetaci (poskytují odhad množství uhlíku, ten se hojně nachází ve dřevě a ne tak aktivních jedincích – například stárnoucích, spících atp.), Listové pigmenty (měření stres indikujících pigmentů a tím určení oslabené vegetace), Obsah vody v listech porostu (indikace zdraví vegetace a případně odolnost vůči požárům) (L3Harris, 2020).

DVI (Difference Vegetation Index) je jedním z nejjednodušších vegetačních indexů a s jeho pomocí se určuje přítomnost vegetace (rozliší vegetaci a půdu), zároveň je také citlivý na množství vegetace. Funguje na principu zvýraznění rozdílu mezi odrazivostí červeného a infračerveného světla. Obdobně funguje například i RVI

(Ratio Vegetation Index), který je také založen na poměru zmíněných oblastí elektromagnetického záření, nese vysoké hodnoty pro vegetaci a nízké pro všechny ostatní plochy (voda, půda, led atd.). Jednoduchým indexem, jehož hodnoty se běžně vyhodnocují i mimo dálkový průzkum Země (DPZ) je LAI (leaf area index), ten vyjadřuje plochu povrchu listů a dají se z něj odvozovat biochemické procesy jako fotosyntéza, a respirace rostlin. V DPZ se odvozuje z jiných vegetačních indexů (například NDVI, SR) nebo z odrazivosti povrchu (Liang et Wang, 2020).

Pro účely vytváření modelu v této práci byly jako relevantní vybrány tři vegetační indexy – NDVI, NDMI a EVI. Jedná se o indexy normalizované a nabývají hodnot od -1 do 1. Jedním z obecně nejznámějších a nejpoužívanějších vegetačních indexů je NDVI (Normalized Difference Vegetation Index), který lze vypočítat jako normalizovaný poměr červeného a infračerveného pásma záření. Je citlivý k množství chlorofylu a používá se tedy k určení růstu vegetace či vitality porostů. Konkrétní příklady hodnot NDVI v kontextu stavu vegetace obsahuje schéma na obrázku 3.

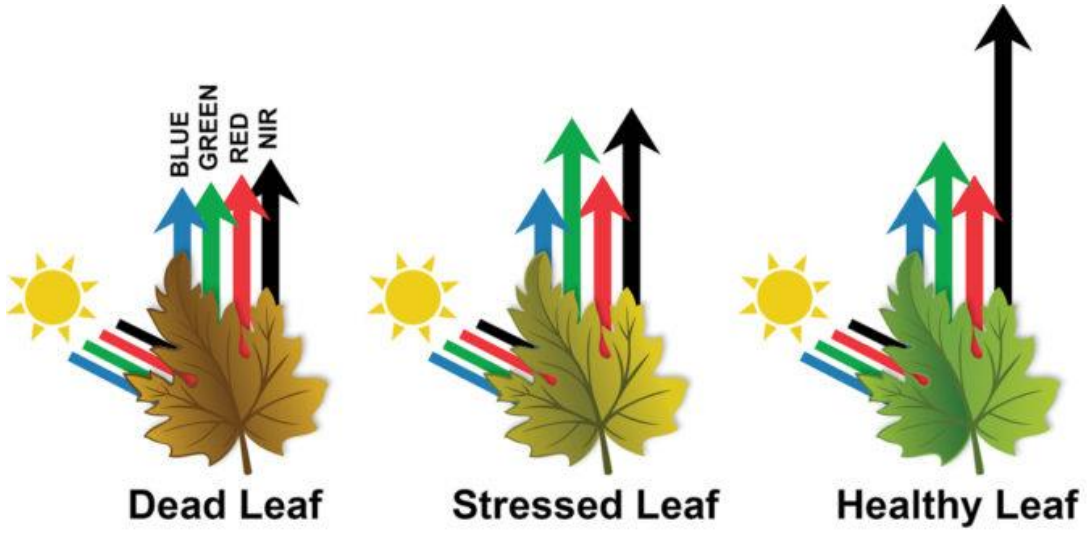


Obrázek 3: Schéma vztahu vitality vegetace a hodnot indexu NDVI (Kraetzig, 2020).

NDMI (Normalized Difference Moisture Index) je index vlhkosti. Vypočítává se pomocí oblasti blízkého infračerveného světla (NIR), jehož odražené množství je ovlivněno množstvím sušiny v listu, a oblasti krátkovlnného infračerveného záření (SWIR), které je naopak citlivé na změny v obsahu vody a struktuře mezofylu v porostu (ESRI, 2020b). Index EVI (Enhanced Vegetation Index) je citlivý ke změnám struktury vegetace a byl vyvinut pro zvýšení citlivosti v oblastech husté vegetace oddělením „šumu“ pozadí a snížením vlivů atmosféry. Vypočítává se z oblastí modrého, červeného a blízce infračerveného světla (NIR) (Huete et al., 2002).

Obrázek 4 znázorňuje schéma poměru odražených světelných pásem různě vitální vegetací od mrtvé (vlevo) přes stresovanou po zdravou (vpravo). Mrtvá

vegetace odráží všechna pásma stejnou měrou, u stresované vegetace se hodnoty blízkého infračerveného záření a červeného již vzájemně blíží a zdravá vegetace odráží až 50 % blízkého infračerveného záření a k tomu 8 % červeného.



Obrázek 4:Schéma pásem odrážených různě vitální vegetací (Palatine, 2017).

2.5 Strojové učení v prostorových vědách

Strojové učení je odnoží umělé inteligence a definovat ho lze jako algoritmy, které se učí ze strojově čitelných dat. Obecně se sem řadí skupiny algoritmů, které dokáží provést vícerozměrnou, nelineární, neparametrickou regresi nebo klasifikaci. Konkrétně do něj například spadají metody jako rozhodovací stromy (decision trees), neuronové sítě (neural networks), náhodné lesy (random forests) a další (Lary et al., 2016). Pro své široké možnosti využití jsou již téměř dvacet let hojně využívané v různých oblastech vědy. Co se týká prostorových věd, je strojové učení čím dál tím více využíváno k interpretaci snímků z dálkového průzkumu Země – jako příklady lze uvést odhad srážek tajfunu nad oceánem (Chen et al., 2011), analýzu distribuce vlhkosti v půdě (Makkeasorn et al., 2006) a monitorování kvality vody v rezervoáru (Chen, 2010).

Modely strojového učení lze rozdělit do několika kategorií – řízené, neřízené, posilované (reinforcement) a hluboké (deep learning). Řízené modely strojového učení jsou nejjednodušší na implementaci a pracují na principu cvičení algoritmu na testovacím souboru dat od uživatele. Takto se algoritmus vycvičí na jeden typ úlohy, který následně zvládne kvalitně zpracovávat i nad neznámými daty. Konkrétní modely spadající do této kategorie jsou například rozhodovací stromy, naivní Bayes a logistická regrese (Kumar, 2020). Princip neřízeného učení funguje obráceně.

Uživatel algoritmu zadá velké množství netříděných dat a algoritmus v nich sám najde systém a roztřídí a seskupí je do podoby srozumitelné pro uživatele. Tyto modely mají ve světě pro svou schopnost najít v datech systém široké využití, jako příklad lze uvést k-means clustering (Garbade, 2018). Reinforcement učení vychází ze signalizace chyb a správných rozhodnutí. Algoritmus se tedy v procesu vlastních chyb učí (na začátku jich dělá nejvíce), hlavní využití pro to je ve videohrách a simulacích (Heidenreich, 2018).

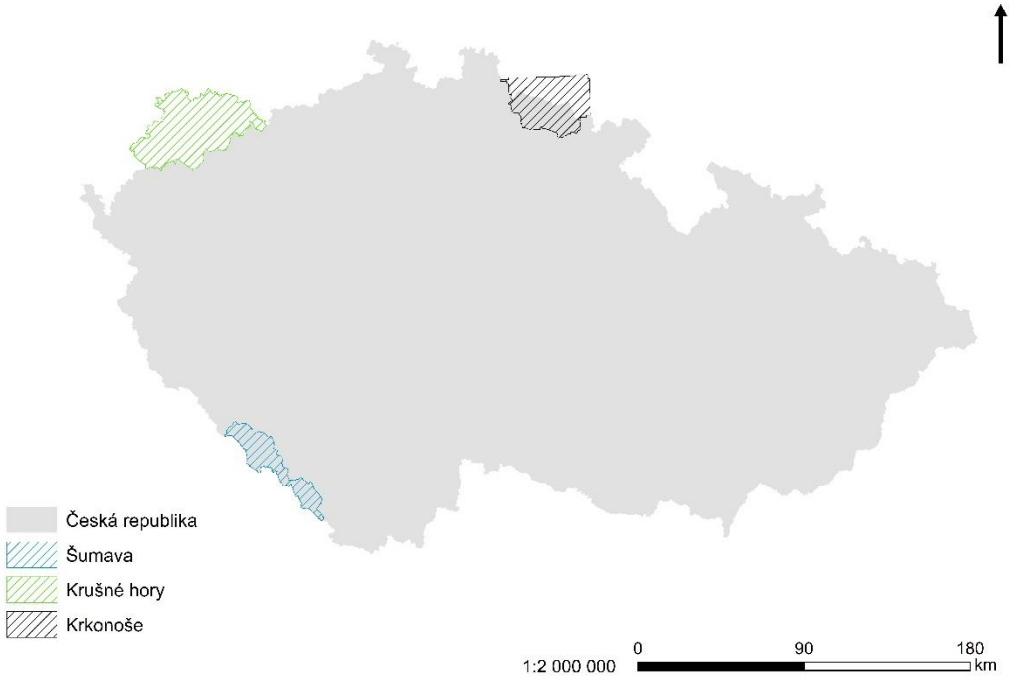
V současné době na popularitě hodně nabírají deep-learning modely (modely hlubokého učení), které sestávají ze čtyř a více vrstev konvolučních neuronových sítí. Tyto modely bývají kvůli své komplexnosti náročnější na vytrénování a pomalejší na učení, často se děje, že se vrstvy učí různou rychlostí, což celý proces částečně limituje (Nielsen, 2019). Oproti výše zmíněným metodám zpracovávají větší množství dat a potřebují větší výpočetní kapacitu, nicméně navzdory delší přípravné fázi se jedná o velmi efektivní nástroje. Jejich hlavní využití je v oblastech, kde jsou vstupem analogová data (určení ručně psaného textu, audio soubory) (Brownlee, 2019).

3. Metodika

3.1 Zájmová území

V roce 2020 tvořily lesní pozemky v ČR celkem 2 614 614 hektarů. Z toho většinu rozlohy zabíraly lesy hospodářské (74,2 %), jako druhé byly nejrozsáhlejší lesy zvláštního určení (23,8 %) a nejmenší podíl tvořily lesy ochranné (2 %) (Riedl et al., 2021). Co se druhového složení lesů celé ČR týče, jasně převládá smrk ztepilý (48,8 %; dle doporučené skladby lesů by to bylo 36,5 %), po něm borovice (16,1 %; doporučeně 16,8 %) a buk (9,0 %; doporučeně 18 %). Jehličnaté stromy tedy jasně dominují a celkem zabírají 70,4 % porostní půdy (Riedl et al., 2021).

Jako zájmové území byla vybrána tři pohoří u hranic ČR – německá strana Krušných hor, polská i česká strana Krkonoš a česká strana Šumavy, všechna jsou zachycena na mapě na obrázku 5. Pro všechna zmíněná pohoří jsou snad nejvýznamnějším prvkem jejich lesy, přičemž ještě většinu z nich tvoří lesy horské. Vývoj lesů je v největší míře určen klimatem, dále půdami a jejich stavem a vzájemnou konkurencí dřevin. Každý les tak má při vzniku jiné podmínky a jejich kombinací vznikají rozmanité lesní porosty. K tomu lze navíc započítat vývoj společenstev v čase a možnosti současného stavu různých lesů jsou nepřeberné (Hubený et Čížková, 2016). Protože se práce zabývá modelováním výšky lesa, budou výsledky úzce souviset s dřevinným zastoupením vybraného území, distribucí porostů a druhem lesů.



Obrázek 5: Vybraná zájmová území v rámci ČR. (Zdroj dat: ČÚZK).

3.1.1 Šumava

Studované území na Šumavě se rozkládá na ploše 668,32 km^2. Na český národní park Šumava (založený v roce 1991) v Německu navazuje národní park Bavorský les (založen v roce 1970). Dále je pohoří součástí Biosférické rezervace Šumava (vyhlášení v roce 1990) a chráněné krajinné oblasti – ta byla založena v roce 1963 a zmíněný národní park byl vyhlášen později v jejím jádře a CHKO tak dnes tvoří jeho „obalovou zónu“. V tomto území se dále nachází i evropsky významná lokalita (Šumava), dvě ptačí oblasti (Šumava a Boletice), čtyři národní přírodní rezervace (Boubínský prales, Černé a Čertovo jezero, Bílá Strž a Velká Niva) a mnoho přírodních rezervací (např. Čertova stráň a Zelenský luh) a přírodních památek (např. Rašeliniště Kyselov, Pasecká slať). Průměrná nadmořská výška Šumavy je 930 metrů, nejvyšším bodem je na německé straně Velký Javor (1 456 metrů nad mořem), na české straně to je Plechý (1 378 metrů nad mořem). Co se stupňovitosti týče, lze na Šumavě nalézt šest vegetačních stupňů od jedlobukového po alpínský.

Na základě pylové analýzy je známo, že historicky na Šumavě dominoval smrk (přibližně před 7 000 lety) a v průběhu času se k němu přidávala jedle a buk. Na přelomu letopočtu se situace změnila a na většině míst začal převažovat buk s výjimkou ostrůvkovitých smrčin na hřbetech hor a v podmáčených lokalitách (Hubený et Čížková, 2016). Za posledních tisíc let se pak stavy smrku opět zvýšily a dnes opět dominuje, je to dáno nejen činností člověka (například sklárny), ale nejspíše především klimatickými podmínkami (malá doba ledová), které hojnost jeho zastoupení v lesích ovlivňují (Hubený et Čížková, 2016). Smrky na Šumavě dnes tedy rostou až k horní hranici lesa (1 200 metrů nad mořem) a tvoří přibližně 75 % porostu (dohromady živé stromy i souše) (Hubený et Čížková, 2016). Po smrku je nejpočetněji zastoupený buk lesní (9,1 %), břízy (5,6 %), borovice lesní (4,8 %), borovice blatka (1,9 %) a olše šedá, jeřáb ptačí, jedle bělokorá, javor klen, vrby a další. Zbytky pralesů dnes tvoří přibližně 5 % rozlohy NP a CHKO Šumava, 35 % tvoří vysázené smrčiny či bory a zbylých 60 % smrkové porosty vzniklé přirozenou obnovou (s případným ovlivněním člověkem) (Hubený et Čížková, 2016). Faktory, které nejvíce ohrožují lesní porosty zde jsou kůrovec (podíl 63 % na odumírání smrků) a vichřice (37% podíl).

3.1.2 Krušné hory

Území Krušných hor zabírá 1821,67 km^2. Na české straně krušných hor se nachází dvě evropsky významné lokality (Východní Krušnohoří a Tisovec) a ptačí oblast (Východní Krušné hory), dále se zde nachází několik národních přírodních rezervací (např. Velký močál a Jezerka), přírodních památek (např. Kokrháč a Sfingy) a přírodních parků (např. Východní Krušné hory a Přebuz). Krušné hory zatím nejsou součástí žádného velkoplošně chráněného území, ale vyskytují se snahy o prosazení vyhlášení CHKO (AOPK, 2021). Nejvyšším bodem Krušných hor je na české straně Klínovec (1 244 metrů nad mořem) a na německé straně Fichtelberg (1 214 metrů nad mořem) (David et Soukup, 2020). Celé pohoří se táhne v délce přibližně 130 kilometrů podél hranic a má průměrnou šířku 40 kilometrů.

Ještě na přelomu letopočtu byly Krušné hory zalesněné souvislým pralesem a až s pozdější kolonizací začalo být území ve větší míře odlesňováno. Během první kolonizační vlny byly zakládány rudné doly, rýžoviště a sklárna, ve vlně druhé se přidala těžba kovů a vznik hornických měst. V 16. století byly Krušné hory nejobydlenějšími horami Evropy (Krušnohorské lesy, 2018). V té době docházelo k významnému odlesňování náhorních partií hor kvůli stavebnímu a palivovému dříví. Zároveň na území docházelo k pastvě dobytka, která stavu porostů také nepomáhala, následně nastal nedostatek dřeva a bylo vyhlášeno mnoho nařízení o hospodaření s lesy (Historie krušnohorských lesů, 2019). Během 18. století v horských částech Krušných hor převažoval smrk s příměsí buku a jedle, v nižších partiích navíc například duby, javory a jasany. Stejně jako Šumava byly i porosty Krušných hor během svého vývoje výrazně ovlivněny klimatem, vichřicemi a případnými kůrovcovými kalamitami. Ve 20. století byly vysazovány především rychle rostoucí dřeviny a vznikaly tak smrkové monokultury. Dále byly porosty znovu výrazně omezeny, a to emisemi především z energetického a průmyslu, v té době docházelo k velkoplošnému odumírání lesů – nejprve jedle na vrcholových partiích hor. Celkově kvůli imisní zátěži odumřelo přibližně 80 % smrkových porostů (Historie krušnohorských lesů, 2019). V druhé polovině 20. století byly uměle vysazovány nepůvodní náhradní dřeviny, které byly schopné odolat imisní zátěži. Po zlepšení imisní situace odsířením elektráren na konci 20. století jsou nyní porosty opět obnovovány pomocí původních dřevin především z hercynské směsi (smrk, buk a jedle) (Krušnohorské lesy, 2018).

3.1.3 Krkonoše

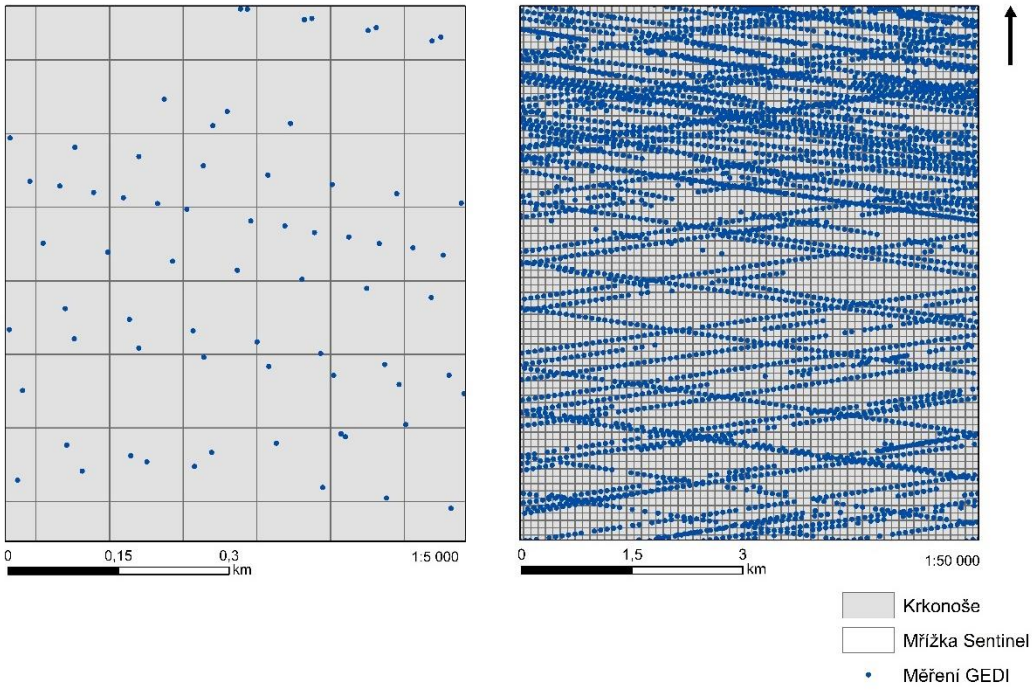
Vybrané území Krkonoš má rozlohu 1187,82 km^2. Na polské i české straně Krkonoš je biosférická rezervace, jako první přeshraniční území tohoto typu byla vyhlášena v roce 1992 (Štursa, 2011). Dále do území spadá polský Krkonošský národní park (založený roku 1959) a český Krkonošský národní park (vyhlášen roku 1963), dohromady tyto parky tvoří území o rozloze 418 km^2. Co se týká lokalit Natura 2000, spadá do oblasti Krkonoš ptačí oblast i evropsky významná lokalita Krkonoše (KRNAP, 2021a). Dále se v Krkonoších nachází několik přírodních památek (např. Anenské údolí a Slunečná stráň) (KRNAP, 2021b). Nejvyšším vrcholem Krkonoš je Sněžka (1 603 metrů nad mořem) a vyskytují se zde čtyři vegetační stupně od submontánního po alpínský.

V období přibližně 7 000 let před naším letopočtem v Krkonoších tvořily lesy borovice, dub, jilm, lípa a líska; smrk a borovice kleč se objevili až o 1 500 let později ve starším atlantiku. Následně byla borovice kleč vytlačena do nejvyšších poloh a nahradil ji buk. Po přelomu letopočtu byly přibližně do 14. století Krkonošské lesy tvořeny pralesy (opět hercynská směs – smrk, jedle a buk) a nedocházelo k významným změnám v jejich druhovém složení (Kajzarová, 2012). Nad horní hranicí lesa (1 200 metrů nad mořem) se vyskytovala pouze borovice kleč a zakrslé formy smrku. Obdobně jako tomu bylo v Krušných horách, i zde začaly být lesy hojně vytěžovány důsledkem rozvoje hutnictví, sklářství, pastevectví a zemědělství. Toto intenzivní využívání lesa se projevilo mezi 16. a 18. století. Vzniklé holiny byly následně osety smrkem pro jeho rychlý růst, následkem toho se po letech společenstva stala nestabilní, protože je ve velké míře tvořil hustý stejnověký smrkový porost (Kajzarová, 2012). Tímto se okyselily půdy což zapříčinilo vyplavování živin a došlo k oslabení zdravotního stavu lesů, takže následně špatně odolávaly klimatickým vlivům a škůdcům. Později se k tomu přidala zhoršená imisní situace, která měla na Krkonoše nejhorší vliv na konci 20. století opět důsledkem provozu tepelných elektráren. Navíc v té době vyvrcholila kůrovcová kalamita, takže byly provedeny velké těžby a porosty nebyly schopné rychlé regenerace (Kajzarová, 2012). Bylo tedy investováno do umělé obnovy porostů a přijata opatření, která dlouhodobě povedou k původní podobě ekosystému.

3.2 Sběr a příprava dat

GEDI

Data GEDI jsou volně dostupná na stránkách searchdata.nasa.gov. V této práci byla použita výšková data L2A verze 002. Jak je již zmíněno výše, úroveň zpracování 2 obsahuje již roztřízené vlnění. Pro každou ze tří oblastí byl stažen obdélník obsahující jednotlivá bodová měření. Tyto datové produkty jsou dostupné ve formátu h5, jako první je tedy bylo třeba převést do formátu, s kterým je možné pracovat v prostředí ArcMap (csv). To bylo provedeno v programu R Studio a při té příležitosti byly z dat jednotlivých přeletů vybrány jen informace potřebné pro zbytek práce, čímž se velikost souborů výrazně zmenšila. Vybrány byly atributy DELTA TIME (čas uběhlý od prvního ledna 2018 v sekundách), výšku nejvyššího a nejnižšího návratu vztaženou k elipsoidu, relativní výšky RH98 a RH100, QUALITY (výběr „nejužitečnějších“ dat), SENSITIVITY (maximální množství vegetace, které paprsek pronikne), DEM (výšku naměřenou TanDEM-X v bodě měření GEDI), LAT a LON (souřadnice zeměpisné výšky a šířky), BEAM (identifikaci paprsku), FLY (datum přeletu nad měřeným bodem) a SHOT NUMBER. Dále byly v prostředí R Studia vybrány jen „kvalitnější“ záznamy, které měly v atributu QUALITY hodnotu TRUE. Po převedení do formátu csv a načtení dat do ArcMap byla data přeletů zobrazena přiřazením X a Y souřadnic (display XY data). Tato vrstva byla uložena (export data), a před další prací ještě přímo ořezána jednotlivými polygony zájmových území (funkce CLIP). Ukázka hustoty bodových měření je na mapě na obrázku 6. Data GEDI jsou ke stažení v koordinačním systému WGS 84 a v něm s nimi bylo i pracováno.



Obrázek 6: Zobrazení hustoty měření GEDI na mřížce 100x100 metrů.

Sentinel

K přípravě a stažení dat Sentinel-1 a Sentinel-2 byl použit Google Earth Engine (GEE), nepřekvapivě se jedná o platformu Google pro geoprostorové analýzy, ve které je možné připravovat a zpracovávat velké objemy dat (např. satelitní snímky) v prostředí cloudu a mimo jiné je tam následně vizualizovat atp. Zpracování dat Sentinel proběhlo v nativním rozlišení a při exportu byla data převedena na rozlišení 100 x 100 metrů, ve kterém s nimi bylo v rámci zbytku práce dále pracováno. Data byla stažena v prostorovém rozsahu obdélníku kolem celé ČR pomocí skriptu, který byl poskytnut vedoucím práce. Mimo stažení dat a nastavení zájmové oblasti zahrnoval rovnou také výpočet vegetačních indexů NDVI, NDMI a EVI, takže výsledná data obsahovala 36 pásem. Z toho jich 28 pocházelo z měření Sentinel-2 a osm z měření Sentinel-1. Z dat Sentinel-2 byly mimo vegetační indexy získány také hodnoty modrého, zeleného a červeného pásma, čtyři různé vlnové délky Red edge světla (RE), blízké infračervené světlo (NIR), krátkovlnné infračervené záření (SWIR), kvalita snímků (QA60, oblačnost). Ze zmíněných bylo jedno pásmo vždy pro medián a jedno pro směrodatnou odchylku.

Data Sentinel-1 se různí především polarizací, polovina obsahuje VV (vertikální polarizace při emisi i příjmu) a polovina VH (vertikální polarizace při emisi a horizontální při příjmu), dále jsou čtyři z osmi pásem pouze z letního období

a od každé intenzity a časového úseku (VV_all, VH_all; VV_summer, VH_summer) byl brán opět jak medián, tak směrodatná odchylka. Výsledkem byl tedy soubor ve formátu tif obsahující mnoho vrstev, který byl z GEE exportovaný na Google Disk, odkud byl již jen stažen. I tato data byla rovnou připravena v koordinačním systému WGS 84 a bylo s nimi v něm i dále pracováno.

Fishnet

V prostředí ArcMap byla vytvořena mřížka (FISHNET) odpovídající buňkám rastru Sentinelu a takto bylo každé buňce přiřazeno vlastní ID (field calculator). Kvůli vysokému počtu záznamů bylo nutné ID field type nastavit na double, protože šlo o sedmimístné cifry a běžný long integer nestačil. Následně byl kvůli úspoře času nad výpočty fishnet původně vytvořený přes celou ČR rozdělen na tři menší území okolo vybraných zájmových oblastí. Toho bylo dosaženo vytvořením nových polygonů, pomocí kterých byla původní mřížka nástrojem CLIP rozdělena.

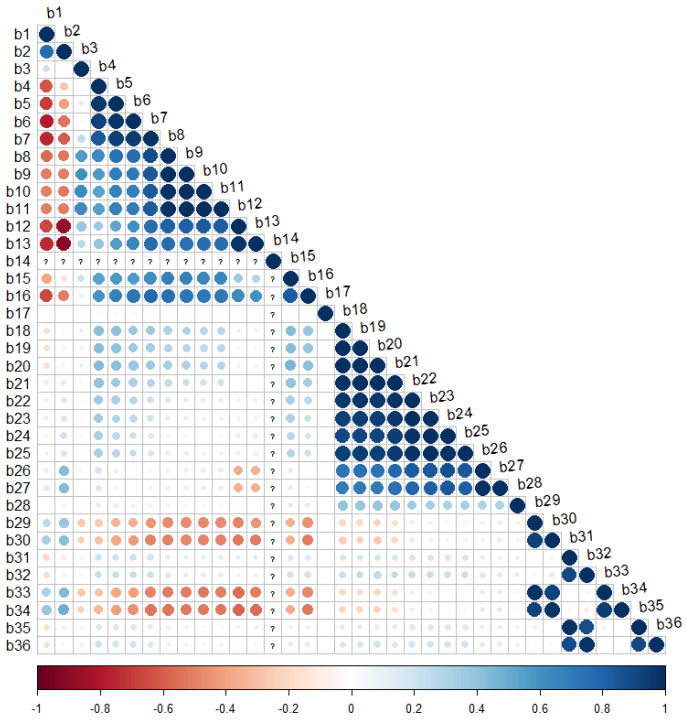
3.3 Zpracování dat

Zpracování vstupních dat probíhalo opět v prostředí ArcMap a kontrola korelace vysvětlujících proměnných a tvorba modelu následně v R Studiu. Jako první bylo třeba spojit předpřipravená měření obou satelitů do jedné tabulky a k záznamům o bodech (GEDI) přidat informaci, ve které buňce leží (Sentinel a fishnet). Do shapefile souboru s daty GEDI byly přidány prostorově odpovídající hodnoty naměřené družicí Sentinel pomocí funkce EXTRACT MULTI VALUES TO POINTS. Dále byly do tabulky přidány hodnoty ID jednotlivých buněk Sentinelu – funkcí INTERSECT. Tam, kde do jedné buňky Sentinel spadalo více než jedno měření GEDI, byly v rámci dané buňky hodnoty výšky vegetace zprůměrovány. To bylo provedeno, protože při přípravě modelu je třeba celý dataset rozdělit na trénovací a testovací data. Na trénovacích datech se obecně vybírá a „nacvičuje“ model, který se poté pustí na datech testovacích a na základě kvality předpovědí je pak validován. Jelikož v datech této práce spadalo někdy více měření GEDI do jedné buňky Sentinel, pravděpodobně by při náhodném rozdělení datasetu na onu testovací a trénovací část vždy nějaké hodnoty z jedné buňky spadly do různých částí a celé rozdělování by bylo bezpředmětné, kdyby se model ve výsledku cvičil na stejných datech jako testoval. Toto zprůměrování bylo provedeno v prostředí MS Excel funkcí AVERAGEIF. Na závěr byly ještě z datasetu odstraněny

„no data“ řádky, které pro Sentinel reprezentovala hodnota -9999. Výsledkem byla tabulka o 75 064 řádcích.

Korelace

Když byla data z jednotlivých zájmových území připravena v tabulkách (pro každé bodové měření GEDI jeden řádek), byly tyto tabulky opět načteny do prostředí R Studia, kde bylo nejprve třeba provést kontrolu vzájemné korelace vysvětlujících proměnných (dat Sentinel-1 a Sentinel-2). Z jednotlivých sloupců s pásmy byly vytvořeny datové rámce, ze kterých byly funkcí cor() vypočítány korelační koeficienty. Tyto byly zobrazeny do korelační matice (viz obr. 7) funkcí corrplot() a z ní byly hodnoty ručně porovnány. V jednom modelu nesměly být použity proměnné, mezi nimiž byl korelační koeficient (záporný či kladný) roven nebo vyšší 0,7. Obecně kladně korelovaly mediány pásem červené, zelené a modré, Red edge, SWIR, směrodatné odchylky téměř všech pásem a vzájemně směrodatné odchylky a mediány měření Sentinel-1. V tabulce 1 je seznam všech zmíněných pásem a k nim přiřazeným číslům.



Obrázek 7: Korelační matice uvažovaných vysvětlujících proměnných.

Pásma			
Název	Číslo	Název	Číslo
NDVI Medián	1	Zelené Směrodatná odchylka	19
NDMI Medián	2	Červené Směrodatná odchylka	20
EVI Medián	3	Red Edge 1 Směrodatná odchylka	21
Modré Medián	4	Red Edge 2 Směrodatná odchylka	22
Zelené Medián	5	Red Edge 3 Směrodatná odchylka	23
Červené Medián	6	Red Edge 4 Směrodatná odchylka	24
Red Edge 1 Medián	7	Blízké Infračervené Světlo Směrodatná odchylka	25
Red Edge 2 Medián	8	Krátkovlnné infračervené záření Směrodatná odchylka	26
Red Edge 3 Medián	9	Krátkovlnné infračervené záření 2 Směrodatná odchylka	27
Red Edge 4 Medián	10	Kvalita snímku Směrodatná odchylka	28
Blízké Infračervené Světlo Medián	11	VV Medián Celoroční	29
Krátkovlnné infračervené záření Medián	12	VH Medián Celoroční	30
Krátkovlnné infračervené záření 2 Medián	13	VV Směrodatná odchylka Celoroční	31
Kvalita snímku Medián	14	VH Směrodatná odchylka Celoroční	32
NDVI Směrodatná odchylka	15	VV Medián Letní	33
NDMI Směrodatná odchylka	16	VH Medián Letní	34
EVI Směrodatná odchylka	17	VV Směrodatná odchylka Letní	35
Modré Směrodatná odchylka	18	VH Směrodatná odchylka Letní	36

Tabulka 1: Čísla a názvy pásem.

Model

Nejprve byla data rozdělena na trénovací a testovací v poměru 7:3. Trénovací část dat byla následně použita k přípravě a „ladění“ modelu a testovací část byla uložena stranou pro výslednou validaci. Ze všech 75 064 záznamů jich tedy 52 544 bylo odebráno na trénování modelu a na zbylých 22 520 záznamech byl později spuštěn až vybraný model. K tvorbě modelu pro následné předpovídání naměřených hodnot byla použita metoda „posílených regresních stromů“ (boosted regression trees, BRT).

Boosted Regression Trees (BRT), známé také jako Gradient Boosted Machine (GBM) nebo Stochastic Gradient Boosting (SGB), jsou neparametrické flexibilní regresní techniky, které kombinují regresní strom s boostovacím algoritmem (Friedman, 2001). Toto rozšíření klasického regresního stromu umožňuje větší flexibilitu a prediktivní výkon při modelování dat (Colin et al., 2017). Jak již bylo zmíněno, kombinují dva algoritmy – regresní stromy a boosting. Klasifikační metoda náhodných stromů je soubor jednotlivých rozhodovacích stromů, v němž je každý strom generován z různých vzorků a podmnožin trénovacích dat. Tato metoda se nazývá náhodné stromy, protože ve skutečnosti klasifikuje datovou sadu několikrát na základě náhodného podvýběru trénovacích dat, čímž vzniká mnoho rozhodovacích stromů (ESRI, 2020a). Boosting je iterační postup, který snižuje odchylku započtením reziduí každého předchozího stromu přizpůsobením dalšího stromu. Každý jednotlivý strom tak informuje následující stromy, a tím i konečný model (Stephanie, 2016). Výhod BRT je mnoho – například to, že nejlepší fit je algoritmem automaticky

detekován, výsledný model představuje účinek každého prediktoru po zohlednění účinků ostatních prediktorů, a především je to metoda robustní a odolná vůči vlivu chybějících a odlehlých hodnot (BCCVL, 2021). Zároveň pro jejich použití nejsou žádné formální předpoklady distribuce dat, boostované regresní stromy jsou neparametrické, a proto mohou zpracovávat zkreslená a multimodální data, stejně tak jako data kategoriální, která jsou ordinální nebo neordinální.

V R Studiu byly k této metodě použity balíčky dismo a gbm pro grafické zobrazení. Jako funkce pro vytvoření modelu byla použita funkce gbm.step(), jako argumenty bylo třeba nastavit datový rámec, vysvětlovanou (průměr výšek naměřených GEDI v jedné Sentinel buňce) a vysvětlující (multispektrální data Sentinel) proměnné, maximální počet vytvořených stromů (max.trees), komplexitu stromů (tree.complexity), míru učení (learning.rate) a proporci pozorování při výběru proměnných (bag.fraction). Vždy byly vybrány nekorelující vysvětlující proměnné, se kterými byl spuštěn model, následně bylo vyhodnoceno, kolik v daném modelu, která vysvětlující proměnná vysvětlila variability a na základě toho probíhaly úpravy následujícího modelu. Variabilita vysvětlená modelem byla vypočítána dle následujícího vzorce.

$$Variabilita\ vysvětlená\ modelem = \frac{Celková\ variabilita\ -\ Křížově\ validovaná\ reziduální\ variabilita}{Celková\ variabilita}$$

3.4 Hodnocení přesnosti

Po zvolení nejlepšího modelu byla na testovací části datasetu funkcí predict.gbm() spuštěna predikce. Argumenty byly vybraný model a část testovacích dat obsahující zvolené vysvětlující proměnné. Funkce vrátila předpovědi naměřené výšky vegetace, které byly následně porovnávány s místně odpovídajícími měřeními GEDI. Přesnost modelu byla posuzována několika způsoby. Nejprve byla vypočítána korelace testovacích dat a hodnot předpovězených modelem. K tomu byla použita funkce cor(), kde byl jako argument zvolen Pearsonův korelační koeficient (značený jako r, vzorec pro výpočet níže). Obecně se korelační koeficienty používají dva – Pearsonův a Spearmanův. Určují míru korelace náhodných veličin a mohou nabývat hodnot od -1 do 1, přičemž 0 znamená nezávislost vstupních dat, kladné hodnoty přímou úměru v datech a záporné hodnoty naopak úměru nepřímou. Čím je hodnota dále od nuly, tím je závislost silnější. Jako silně korelované jsou veličiny obecně považovány při

dosažení hodnoty 0,7 (ať už kladně či záporně) (Clark, 2018). Pearsonův korelační koeficient se používá pro kvantifikaci lineárního vztahu vybraných veličin, oproti němu je Spearmanův korelační koeficient, který je neparametrický a používá se tak i k vyhodnocení nelineární formy závislosti, v zásadě popisuje, jak se vztah dvou veličin podobá monotónní funkci (Holčík et Komenda, 2015).

$$r = \frac{\sum(x - \hat{x}) * (y - \hat{y})}{\sqrt{\sum(x - \hat{x})^2 * \sum(y - \hat{y})^2}}$$

Následně byly na základě rozdílů předpovězených hodnot a hodnot výšky vegetace z testovací části datasetu vypočítány dvě metriky chyb – odmocněná střední kvadratická odchylka (Root Mean Squared Error, RMSE) a průměrná chyba (Mean Error, ME). Průměrná chyba byla vypočtena jako suma rozdílů vydělená celkovým počtem záznamů. Obecně se z této metriky dá odvodit, jestli má model tendence spíše nadhodnocovat nebo podhodnocovat reálná (testovací) data. Čím blíže je výsledná hodnota nule, tím jsou nadhodnocení a podhodnocení vyrovnanější, anebo se tím spíše nedějí. Vyjde-li výsledná hodnota ME kladná, model spíše nadhodnocuje a vyjde-li záporná, má model naopak tendence predikovat nižší hodnoty, než byla na místě reálně naměřena. U této metriky je sporná její užitečnost, právě protože může nabývat od záporných po kladné hodnoty, může být velmi zavádějící. Jak tomu u průměrů bývá, ve výsledku se dílčí záporné a kladné chyby vyruší a celková průměrná chyba tak může vyjít velmi nízká, přestože jsou v datech i hojně zastoupeny chyby významné. Proto je vhodné metriku průměrné chyby kombinovat s dalšími, mnohdy je například nahrazována výpočtem průměrné absolutní chyby (Mean Absolute Error, MAE), která počítá s rozdíly vstupních hodnot v absolutní hodnotě.

RMSE byla vypočtena jako suma kvadratických rozdílů předpovězených hodnot a hodnot naměřených LiDARem podělená počtem měření a na závěr odmocněna. Tato metrika indikuje míru rozptýlení dat kolem hodnot předpovídaných modelem (Gupta, 2021) a lze ji chápat jako směrodatnou odchylku predikce chyb. V zásadě tedy udává průměrnou vzdálenost mezi predikcí a testovacími daty. Kvůli odmocnění nabývá jen kladných hodnot a čím větší jsou, tím je chyba mezi předpovědí modelu a naměřenými daty větší. Je stejně jako ME v jednotkách sledované proměnné. I na tuto metriku je třeba nahlížet s perspektivou zprůměrovaných hodnot, kde i jedna vysoká hodnota může výsledek významně změnit.

$$ME = \sum \frac{x - y}{n}$$

$$RMSE = \sqrt{\frac{\sum (x - y)^2}{n}}$$

Posledními metrikami chyby předpovědí byly %RMSE a %ME, tedy odmocněná střední kvadratická odchylka a průměrná chyba vyjádřená procentuálně. Tyto metriky byly vypočteny vydělením již vypočítané chyby (RMSE či ME) průměrem referenčních hodnot z měření GEDI ($\hat{y}$,) a vynásobením stem. Protože je hodnota chyby daná pouze v metrech relativní, je tento „standardizovanější“ způsob nahlížení na míru chybovosti čitelnější a intuitivnější.

$$\%ME = \frac{ME}{\hat{y}} * 100$$

$$\%RMSE = \frac{RMSE}{\hat{y}} * 100$$

4. Výsledky

4.1 Model

Jako nejlepší byl vybrán model s vysvětlujícími proměnnými Red edge 1, NDMI, VV Směrodatná odchylka Letní, VV medián Letní a EVI, který celkem vysvětlil 51,63 % variability. Jako ideální počet stromů bylo softwarem určeno 12 000.

Celý zápis výsledného modelu vypadal následovně:

gbm.step(data = Train, gbm.x = c("b2_S1S2_Su", "b3_S1S2_Su", "b7_S1S2_Su", "b35_S1S2_S", "b33_S1S2_S"), gbm.y = 39, family = "gaussian", max.trees = 15000, tree.complexity = 1, learning.rate = 0.001, bag.fraction = 0.5)

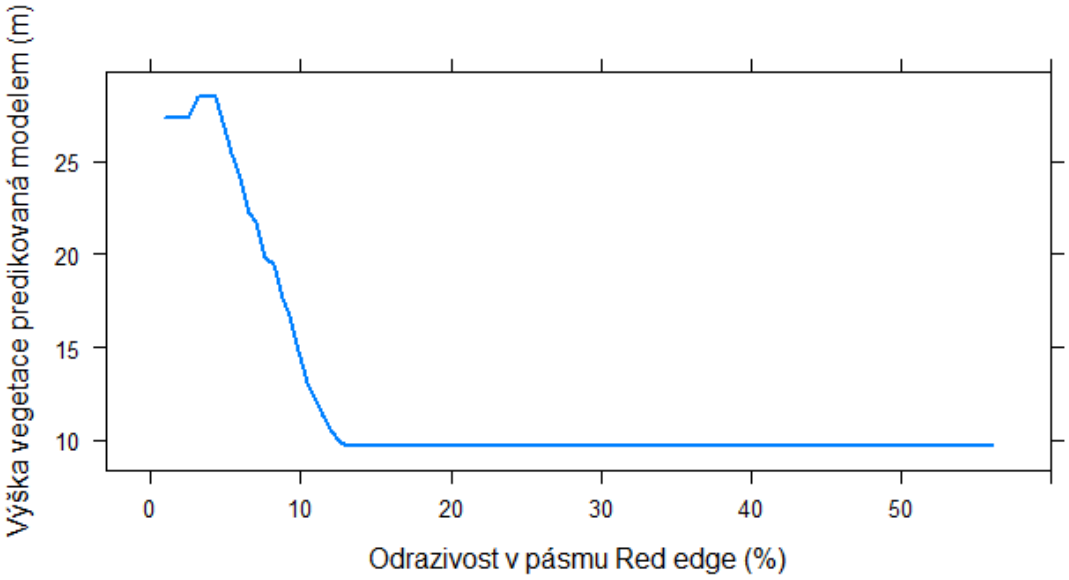
V tabulce 2 je seznam použitých vysvětlujících proměnných navíc doplněný o míru, nakolik se na vysvětlené variabilitě podílely. Pásmo Red edge 1 vysvětlilo 87,7 %, vegetační index NDVI a VV Směrodatná odchylka Letní oba vysvětlily necelých 5 %, VV medián Letní přibližně 2 % a vegetační index EVI necelé jedno procento.

Vysvětlující proměnné	
Pásmo	Vysvětlená variabilita (%)
Red edge 1	87,74
NDMI	4,73
VV Směrodatná odchylka Letní	4,62
VV medián Letní	2,18
EVI	0,72

Tabulka 2: Vybrané vysvětlující proměnné a jimi vysvětlená variabilita v procentech.

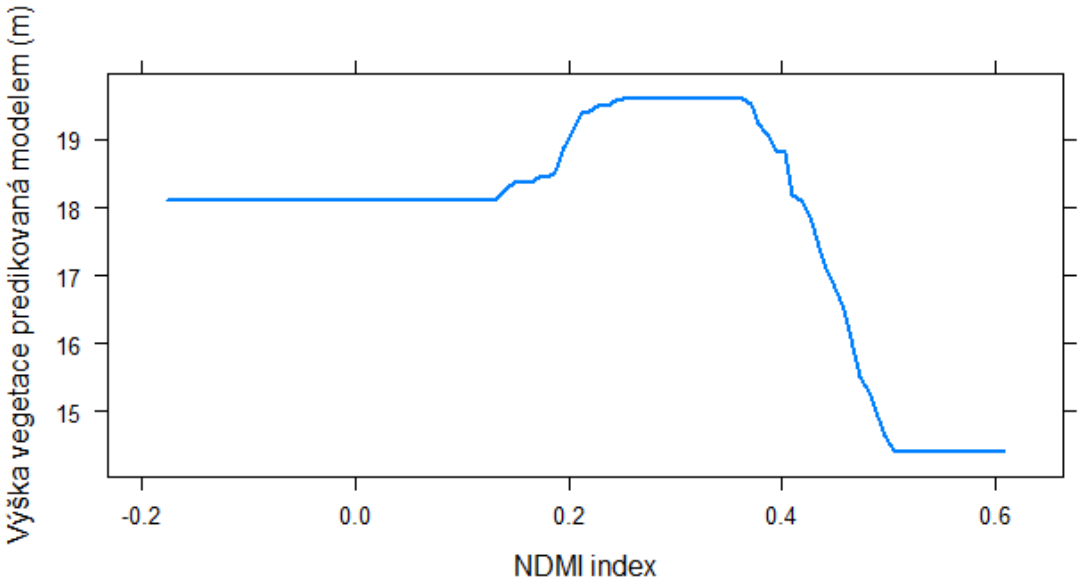
4.1.1 Vysvětlující proměnné

Níže je několik grafů částečné závislosti (partial dependence plots), které zobrazují, jaký vliv mají jednotlivé proměnné na predikci modelu a jaká hodnota vysvětlující proměnné modelu indikuje jakou předpovídanou hodnotu výšky vegetace. Na ose y je vždy jako závislá proměnná predikce výšky stromů v metrech a na ose x vysvětlující proměnné z výsledného modelu. Nejmenší odrazy pásma Red edge v modelu odpovídaly nejvyšší vegetaci a následně se s klesajícím vzrůstem vegetace zvyšovaly (obr. 8). Nejvyšší hodnoty odraženého záření byly modelem očekávány pro vegetaci řádově vysokou 10 metrů.



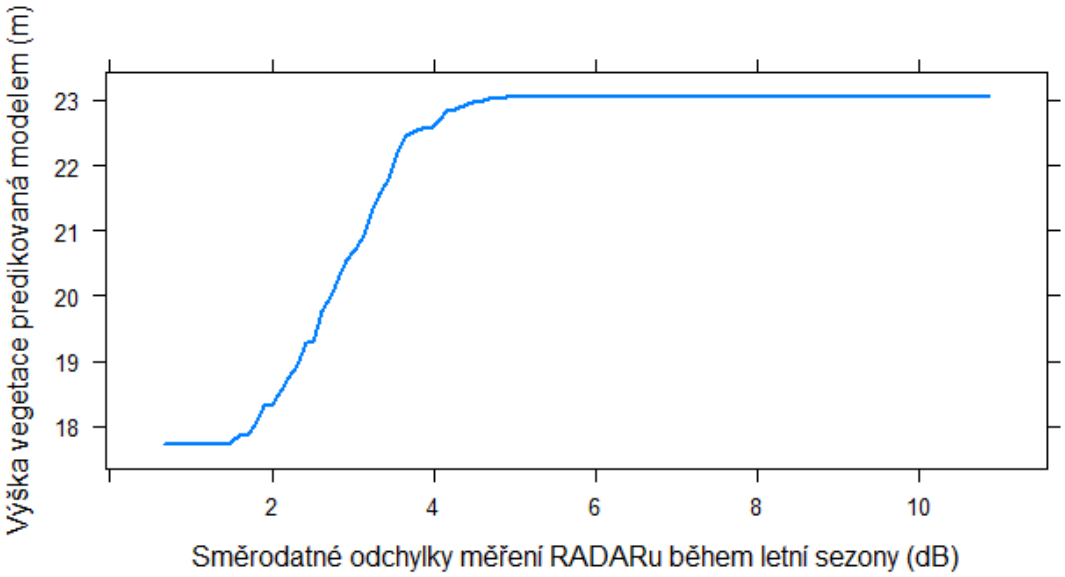
Obrázek 8: Graf částečné závislosti predikované výšky vegetace na odrazivosti Red edge. Na ose x odrazivost pásma Red edge v procentech. Toto pásmo reaguje na množství chlorofylu v listech, platí, že čím vyšší je jeho hodnota, tím je vyšší odrazivost (tedy tím více chlorofylu ve vegetaci).

Přibližně od hodnoty NDMI 0,4 předpovídaná výška stromů klesala (obr. 9). Počátek poklesu se zde shoduje s pomezím hodnot NDMI pro vysoké vegetační pokrytí zcela bez stresu (0,4-0,6) a středně vysokého vegetačního pokryvu s vysokým stresem nebo nižšího pokryvu rostlinami jen málo stresovanými nedostatkem vody (0,2-0,4). Nejvyšší predikované stromy (20 metrů) byly na pomezí zdravého stavu, ale zároveň model očekával i další vzrostlé dřeviny, které neměly ideální přísun vody a projev stresu se na nich kvůli tomuto nedostatku začínal projevovat. Dle křivky by se tedy v zájmových územích jako jediná vegetace nestresovaná nedostatkem vody daly očekávat dřeviny nižší 15 metrů.

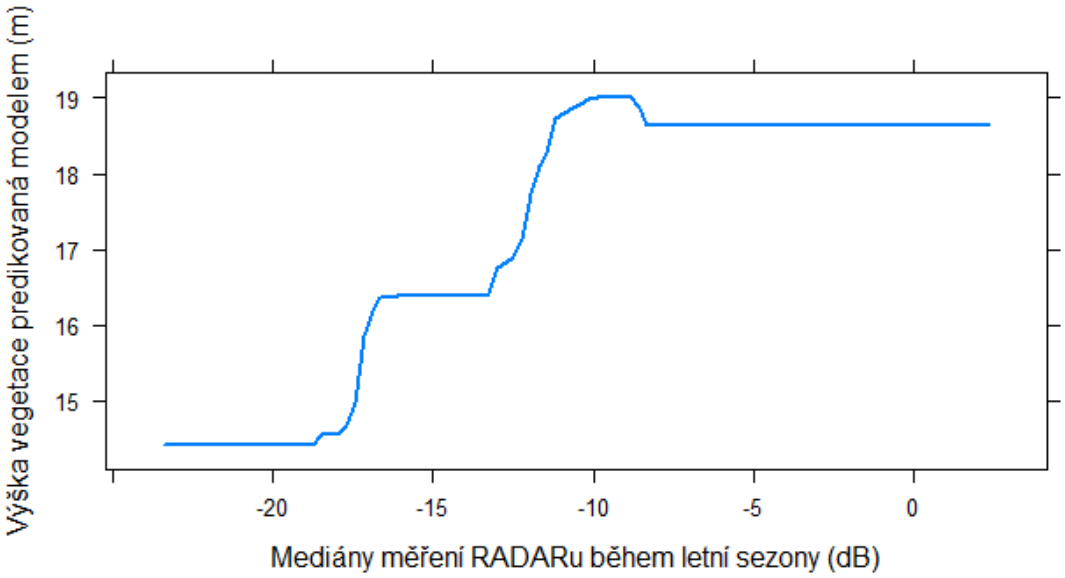


Obrázek 9: Graf částečné závislosti predikované výšky vegetace na indexu NDMI. Na ose x hodnoty indexu NDMI Hodnota -1 značí území bez vegetačního pokryvu, naopak hodnota 1 znamená hustou vegetaci bez stresu z absence vody.

Z grafu zobrazujícího směrodatnou odchylku měření radaru (obr. 10) je zřetelný pozitivní vztah obou proměnných – čím větší byla směrodatná odchylka, tím vyšší vegetaci model předpovídal. Nejmenší odchylka (tím i nejméně navrátivšího se signálu) byla predikována u vegetace nižší 18 metrů. Křivka nejstrměji stoupá mezi hodnotami odchylky 2 a 4. Od predikované výšky 23 metrů se hodnoty výšky vegetace nezvyšovaly a rostla pouze směrodatná odchylka. V průběhu křivky závislosti predikované výšky vegetace na mediánu měření radaru lze na první pohled rozlišit tři různé výšky vegetace zastoupené širším intervalem síly odrazu (obr. 11). Nízkou odrazivost model očekával u nižší vegetace (pod 15 metrů) a vysokou naopak u vzrostlých stromů (19 metrů).

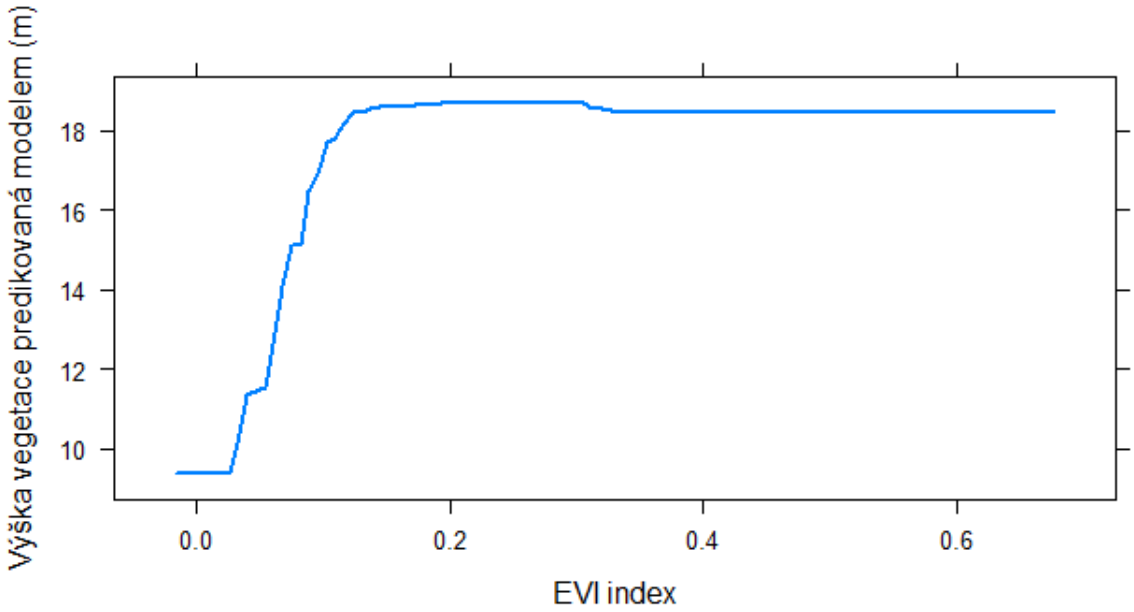


Obrázek 10: Graf částečné závislosti predikované výšky vegetace na směrodatné odchylce radarových měření. Na ose x hodnoty odchylky odrazu v decibelech. Čím vyšší vegetace, tím více odražené energie a tím vyšší směrodatná odchylka.



Obrázek 11: Graf částečné závislosti predikované výšky vegetace na mediánu radarových měření. Na ose x hodnoty síly odrazu v decibelech – záporná čísla značí útlum a čím více se hodnoty blíží nule, tím je zpětný odraz silnější a vegetace hustší.

Z grafu vztahu předpovědí modelu a hodnot EVI indexu spočtených z měření družic Sentinel je zjevné, že proměnné jsou v pozitivním vztahu – s predikovanou výškou vegetace rostou i hodnoty EVI indexu (obr. 12). Jelikož hodnoty indexu 0,2-0,8 odpovídají zdravé vegetaci co se množství chlorofylu týče, vzrostlé stromy (řádově 19 metrů) model predikuje tam, kde je chlorofylu vyšší množství. Zlom v průběhu linie se v hodnotách indexu nachází přibližně mezi 0 a 0,1. Lze tedy soudit, že jako zdravou vegetaci model očekává spíše tu vysokou, nikoli nižší 10 metrů.

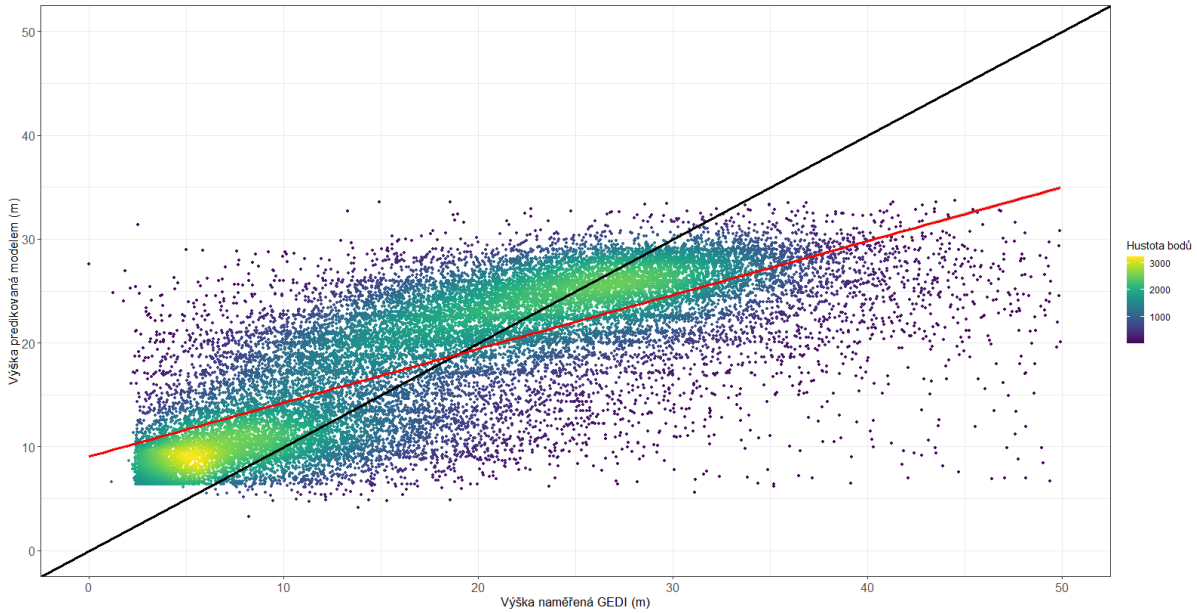


Obrázek 12: Graf částečné závislosti predikované výšky vegetace na EVI indexu. Na ose x hodnoty EVI – mohou nabývat od -1 do 1. Zdravou vegetaci ukazují hodnoty 0,2-0,8.

4.2 Korelace

Vzájemná korelace mezi výslednými hodnotami predikovanými modelem vyšla 0,70. Jelikož je hodnota korelačního koeficientu kladná, jedná se o pozitivní vztah – přímou úměru. Tedy s růstem hodnot předpovězených modelem by měly narůstat i hodnoty měřené satelitem. Pro další srovnání predikovaných a naměřených hodnot je v grafu na obrázku 13 černě vykreslena linie identity (identity line), která má standardně sklon 1 a počátek v bodě 0 (Sterner et Elser, 2003). Jelikož se obecně tato linie využívá při porovnávání dvou datasetů v 2D prostoru, byly-li by zmíněné hodnoty identické, spadaly by pod linku a kopírovaly tak její průběh. Vzhledem k tomu, že se hodnoty bodů zvyšují zleva doprava a mají částečně (přestože ne velmi výrazně) podobu křivky, i zde se dá soudit, že proměnné mají pozitivní lineární vztah, což také odpovídá hodnotě korelačního koeficientu. Světlejší barva zde indikuje větší hustotu (množství)

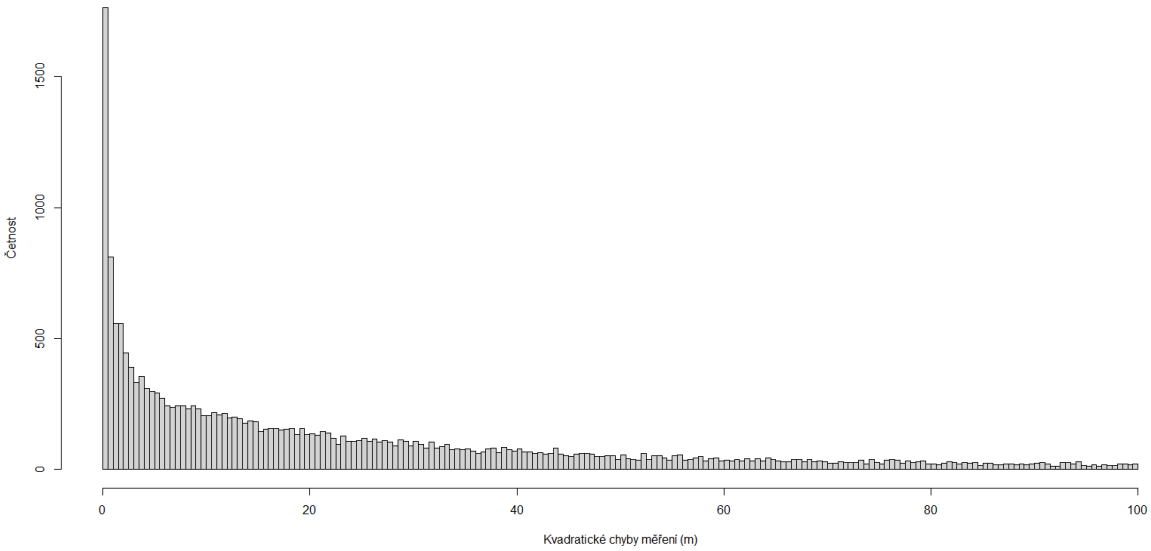
bodů v oblasti a tmavší barva naopak představuje méně bodů. Rozložení bodů kolem identity line naznačuje, že model má spíše tendence nadhodnocovat.



Obrázek 13: Zobrazení vztahu obou proměnných. Červenou barvou je vykreslena regresní křivka závislosti predikcí (osa y) na naměřených datech (osa x), černě identity line.

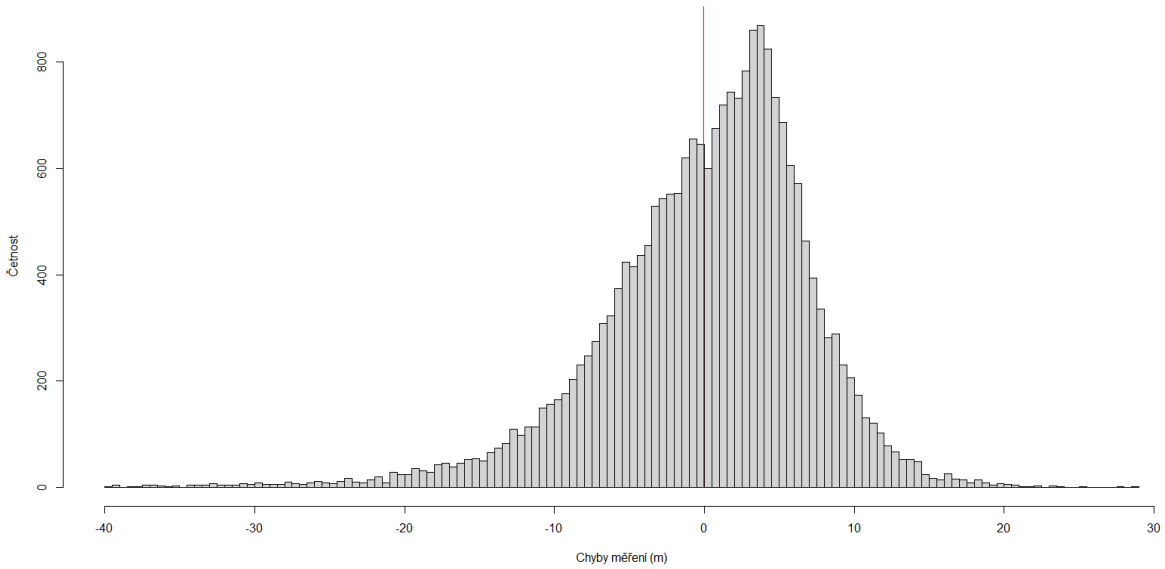
4.3 RMSE a ME

Na celém souboru testovacích dat vyšla RMSE 7,60 metrů a ME -0,06 metru. Hodnota procentuálně vyjádřené RMSE byla 40,66 % a procentuální ME -0,33 %. Z histogramu umocněných chyb předpovězených dat (obr. 14) je patrné, že nejvíce chyb bylo do jednoho metru a následně se četnost snižovala se zvyšující se chybou, což je běžný jev. Pro účely zobrazení bylo odebráno 2 786 odlehlých hodnot.



Obrázek 14: Histogram kvadratických chyb predikcí.

Histogram na obrázku 15 zobrazuje četnosti jednotlivých chyb (rozdílů hodnot předpovězených modelem a naměřených LiDARem GEDI) predikce modelu. Červená linka označuje průměrnou chybu -0,06 metru. Z výsledné záporné hodnoty by bylo možné soudit, že model spíše podhodnocuje, přestože jde o jednotky centimetrů. Z histogramu je patrné, že absolutně bylo více chyb kladných, avšak i v tomto případě byly pro účely zobrazení odebrány některé odlehlé extrémní hodnoty. Tyto byly pouze ze záporných hodnot a výsledný histogram tedy zobrazuje o 64 záznamů méně, než bylo původně všech predikcí. Minimum všech hodnot bylo -82,20 metrů a maximum 28,87 metrů. Co se rozložení hodnot do kvartilů týká, 50 % chyb se pohybovalo v rozmezí -3,69 a 4,60 metrů.



Obrázek 15: Histogram chyb predikcí. Červená linka znázorňuje jejich průměr.

Do zájmové oblasti Šumava spadalo 3 688 měření z testovací části datasetu. Průměrná chyba zde byla -0,09 metru a RMSE 8,33 metrů. Obě tedy vyšší než z celkového počtu dat. V histogramu na obrázku 16 jsou hodnoty více distribuované kolem nuly bez tak výrazného píku, jaký byl v zobrazení histogramu chyb ze všech území najednou (obr. 15). Při vytváření grafu bylo odebráno 19 záporných odlehlých hodnot, minimum této části datasetu bylo v bodě -82,20 a zároveň představovalo nejnižší hodnotu z chyb všech testovacích dat. Maximum bylo 25,07 a bylo o necelé čtyři metry menší než nejvyšší hodnota chyby celého testovacího datasetu. Stejně tak rozdělení dat do kvartilů přibližně následovalo trend všech testovacích dat. Procentuální chyba RMSE zde vyšla 43,03 % a ME -0,50 %.

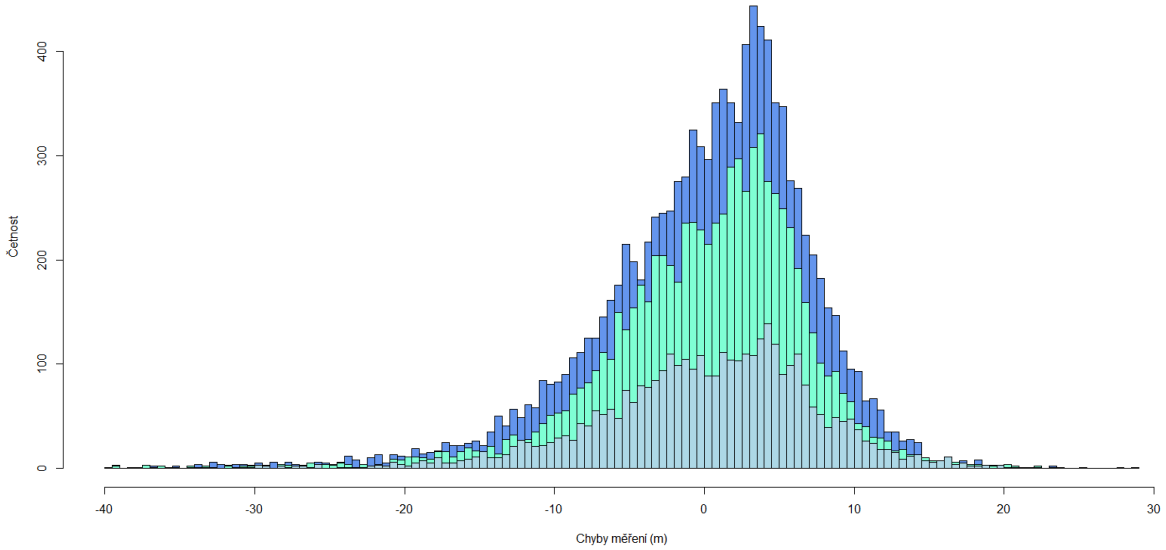
V Krušných horách zbylo po odebrání 70 % trénovacích dat 7 821 záznamů. Na tomto území z nich vyšla ME -0,02 metru a RMSE 7,25 metru. Vyjádřeny

procentuálně vyšly tyto metriky -0,08 % a 37,81 %. Mimo rozsah osy y leží 23 záporných odlehlých hodnot. Minimum a maximum těchto dat bylo v bodech -59,56 a 28,87 (maximum zde je zároveň nejvyšší hodnotou celého testovacího datasetu).

Z oblasti Krkonoš zbylo v testovací části 11 011 záznamů. Z nich vypočítaná ME byla -0,08 metru a RMSE 7,58 metru. Z chyb predikcí na tomto území bylo pro potřeby zobrazení histogramu odebráno 22 odlehlých hodnot, které byly i zde záporné. Minimum tohoto souboru chyb se nachází v bodě -77,30 a maximum 23,15. Procentuálně vyjádřená RMSE vyšla na tomto území 41,86 % a procentuální ME -0,45 % (viz tabulka 3).

Metriky chyb v zájmových územích				
	RMSE	%RMSE	ME	%ME
Šumava	8,34	43,03	-0,10	-0,50
Krušné hory	7,26	37,81	-0,02	-0,08
Krkonoše	7,58	41,86	-0,08	-0,45

Tabulka 3: Výsledné chyby v jednotlivých zájmových územích.



Obrázek 16: Histogram chyb predikcí v jednotlivých zájmových územích. Šedá barva reprezentuje chyby z území Šumavy, zelená z Krušných hor a tmavě modrá chyby predikcí vůči měřením v Krkonoších.

5. Diskuze

Na všech vybraných územích platí, že nejčetnější jsou kladné chyby nedaleko nuly, avšak celková průměrná chyba vyšla vždy záporná. Přestože se dle výsledných ME zdá, že model má tendence podhodnocovat, výsledný poměr kladných a záporných hodnot může být velmi vyrovnaný, nebo mohou dokonce převažovat chyby kladné. Celkový negativní trend bude nejspíše dán extrémními zápornými hodnotami (outliery). Dle grafu na obrázku 13 se na základě většiny bodů nad identity line spíše zdá, že model nadhodnocuje. Záporná průměrná chyba tedy nemusí vypovídat o celkové tendenci, ale tím, že je součástí výpočtu této metriky průměr, může být výsledek zavádějící, to je všeobecně známá vlastnost této metriky. Odlehlé hodnoty, přestože jich není příliš, mají velmi nízké hodnoty a „překlápí" chybu na zápornou stranu. Při odstranění hodnot menších -40 (z území Šumavy 19, z Krušných hor 23 a z Krkonoš 22) vyšla průměrná chyba již vždy kladná a vyšší než při jejím výpočtu ze všech hodnot (přibližně 0,1 metru). Je tedy otázkou, jestli vznik těchto velkých chyb spojoval jeden z faktorů prostředí, nebo vznikly nezávisle na sobě. Několik níže uvedených studií upozornilo na to, že jejich modely měly tendence nadhodnocovat nižší vegetaci a podhodnocovat vegetaci vyšší (Wang et al., 2021; Xi et al., 2022).

Vzhledem k tomu, že v této práci šlo pouze o modelování hodnot, je dosažená hodnota RMSE 7,6 m v rámci očekávatelných mezí, neboť i satelitní LiDARová měření v horském prostředí běžně dosahují RMSE okolo 4 metrů (Neuenschwander et Magruder, 2019). Avšak takto vysoká zde dosažená hodnota je překvapivá při porovnání s výsledky jiných prací – například se studií Wang et al. (2021), kde autoři dosáhli RMSE přibližně 2 metry, a se studiemi Li et al. (2020) a Lang et al. (2022), kde autorům vyšla RMSE pro různé předpovědní modely necelé 3 metry. Je otázkou, čím to může být zapříčiněno, protože v ostatních zmíněných pracích byly proměnné i prostředí mnohdy velmi podobné jako zde. Nejvyšší RMSE mezi predikovanými a testovacími daty vyšla v práci, kterou vydali Potapov et al. (2021) – 6,6 metru. Jako LiDAR zde byla také využita měření GEDI, avšak jako vysvětlující proměnné byla brána data Landsat. Nicméně u zmíněné práce lze nižší přesnost argumentovat tím, že byl vytvářen „globální model". Pracovalo se na území téměř celého světa, takže se i operovalo s nesrovnatelně větším množstvím dat a více typy lesních porostů. Potapov et al. (2021) zároveň pracovali s daty Landsat, která v jiné práci vyšla při porovnání s daty z platformy Sentinel pro modelování výšky vegetace jako méně přesná (Li et

al., 2020). Obecně lze očekávat, že v rámci menšího území by mělo být možné model lépe připravit pro predikce a chyby by měly být menší než při jeho použití ve větším měřítku. Z výsledků prezentované práce a jejich porovnání s obdobnými výstupy se však zdá, že tento předpoklad všeobecně neplatí a je předmětem k diskusi, zda to může souviset s vybranými různorodými pohořími nebo jiným faktorem; vybraná zájmová území zde použitá se vyznačují rozličnou a mnohdy velmi vysokou svažitostí, o které je známo, že přesnost predikcí znatelně komplikuje (Potapov et al., 2021). Výsledky této práce tedy korespondují s výsledky Potapov et al. (2021) i co se tendencí modelu k podhodnocování týče, ale i tak průměrná chyba dosažená modelem vyšla oproti jiným pracím znatelně nižší (Xi et al., 2022; Lang et al., 2022).

Obdobně jako tato práce byl zaměřený výzkum, který prováděli Li et al. (2020) na hornatém území v Číně. Testovali dva předpovědní modely pro výšku vegetace postavené na algoritmech random forest a deep learning, vysvětlovanou proměnnou byla LiDARová data ICESat-2 a jako vysvětlující proměnné byla použita multispektrální data družic Landsat-8 a Sentinel. Pearsonův korelační koeficient pro modely s použitím dat Sentinel ve studii vyšel 0,78 (deep learningový model) a 0,68 (algoritmu random forest) – obě hodnoty velmi blízké hodnotě korelace dosažené v této práci (0,7). Predikce obou modelů studie měly RMSE mezi 2,5 a 3 metry, oproti tomu hodnoty této metriky v předkládané práci vyšly téměř třikrát vyšší. Velikost zájmových území této práce i studie Li et al. (2020) byla srovnatelná a vertikální členitost území této práce pouze nepatrně vyšší než ve studii jejich. Tamější vegetace byla na rozdíl od vysokohorských smrkových lesů v České republice, tvořena především opadavými druhy, a i to například mohlo ovlivnit výsledky měření. V modelech zmiňované studie se jako slabší vysvětlující proměnné ukázaly ty z družice Landsat-8, nicméně se autorům jako funkční osvědčila možnost kombinace vstupních vysvětlujících proměnných z více zdrojů (doplnění dat Landsat-8 o koeficienty a Red edge pásma z družice Sentinel-2). Je otázkou, do jaké míry by spojení dat umožnilo přesnější predikce i při přípravě jiných modelů (například s jinými algoritmy a proměnnými), což by jistě stálo za další pozornost. Studie Li et al. (2020) navíc také validovala měření ICESat-2 pomocí leteckého laserového skenování, autoři uvádí, že největší korelace mezi těmito dvěma datovými sadami byla v měřítku 250 metrů, což je hrubší měřítko, než bylo použito v předkládané práci. Jak porovnání přesnosti predikcí mezi více měřítky, tak i zmiňované srovnání predikcí s daty pořízenými leteckým laserovým skenováním by bylo podnětným pokračováním

předkládané práce. Sami autoři uvádí, že data ICESat-2 nejsou časově ani prostorově rovnoměrně distribuovaná, a je tak vhodné je pro vyvozování širších vegetačních souvislostí využívat opatrně. Zároveň stojí za zmínku, že LiDAR GEDI použitý v této práci provádí mnohem četnější měření než ICESat-2, což na výsledky bude také mít podstatný vliv.

Jako rozlišení pro tvorbu modelu, jehož výstupy nejlépe odpovídají měření leteckého laserového skenování bylo i ve studii Xi et al. (2022) vyhodnoceno 250 metrů a ne vyšší. Studie tedy potvrzuje význam tohoto parametru a vhodnost zanesení srovnání přesnosti měření GEDI a leteckého laserového skenování do pracovního postupu. Takto by bylo navíc zjištěno, jak moc se predikce liší od „reality" a jestli by reálně model mohl být vůbec použitelný, protože je třeba počítat i s určitou chybou měření samotného satelitního LiDARu. Bez toho by hrozilo vytvoření modelu, který ale reálně nebude výhodné k jakýmkoli predikcím používat, jelikož by mohl extrapolovat hodnoty, které by samy o sobě nebyly dostatečně přesná. Xi et al. (2022) ve své studii v rámci buňky LiDARová měření také zprůměrovali a tyto hodnoty extrapolovali. To je stejný postup jako v této práci, nicméně jimi využitých měření ICESat-2 je obecně méně a nabízí se otázka, zda to pro tuto práci bylo ideální řešení, protože z měření GEDI do jedné 100 metrové buňky spadala například i 4 měření. Velmi zajímavé bylo ve zmiňované práci i použití informací o druhu lesa (data z čínského národního geografického monitoringu přírodních podmínek – National Geographical Condition Monitoring), autoři tak mohli porovnat „funkčnost" modelu mezi jednotlivými druhy vegetace (jehličnatá, opadavá, smíšená), přičemž tato možnost by jistě přinesla nové závěry i pro předkládanou práci. Z obou modelů studie vyšly predikce nejlepší pro jehličnaté lesy (ale z těchto pocházelo nejméně dat), korelační koeficient vyšel v zásadě stejný jako v této práci (0,72 – model random forest a 0,7 – model gradient boosting decision tree) a RMSE poloviční (3,15 m a 3,28 m). Oba modely pro smíšený les měly téměř totožné výsledky – korelační koeficient s měřenými daty 0,6 a RMSE 3,3 metrů, takže i tyto hodnoty řádově odpovídaly výsledkům této práce. Vzhledem k tomu, že jejich model pro všechny lesy bez rozdělení typů vegetace měl nejméně přesné výsledky, lze soudit, že vliv druhu porostu je značný a je vhodným námětem k dalšímu zkoumání této problematiky.

Co se nejlepších vysvětlujících proměnných týká, ve studii Xi et al. (2022) byly jako nejefektivnější vyhodnoceny vegetační indexy, po nich topografie (sklon svahů,

nadmořská výška a orientace ke světovým stranám) a na posledním místě radarová měření, studie nezkoumala vliv samostatných pásem. V jehličnatých lesích nejvíce variability vysvětlil index NDVI, po něm výše zmíněné topografické informace a dále radar s dvojitou vertikální polarizací, radar s vertikální a horizontální polarizací a index EVI. Protože NDVI se počítá z červeného pásma a NIR, měly by obsažené informace rámcově odpovídat pásmu Red edge, které vysvětlilo nejvíce variability ve výsledném modelu této práce. Až na topografii odpovídají vysvětlující proměnné zvolené ve studii Xi et al. (2022) těm v této práci. Ve smíšeném lese už zvolené prediktory v těchto pracích tolik nekorespondují.

Lang et al. (2022) jako jedni z prvních autorů udělali experiment s novými daty GEDI (použili data z prvních čtyř měsíců mise) a nacvičili na nich neuronové sítě k předpovědi výšky vegetace na lokalitách po celé Zemi. Data rozdělili na testovací a trénovací poměrem 1:9 (nápadně vyšší poměr než v ostatních pracích, který nejspíše umožnilo velké množství dat) a následně vypočítávali metriky chyb jako v této práci – RMSE (3,6 m), ME (-0,3 m) a procentuální MAE (26,5 %). Jednak jsou tyto hodnoty velmi obdobné výsledkům jiných předpovědních modelů zde zmiňovaných a jednak mimo RMSE v zásadě odpovídají i výsledkům této práce. Využití neuronových sítí redukuje možnost ručního nastavování parametrů modelu a zanesení chyb lidským faktorem. Je zajímavé, že i tak model vykazuje obdobnou chybovost jako jiné ručně připravované modely.

Wang et al. (2021) ve své práci porovnávali dva modelovací algoritmy (spatially-weighted geographical random forest a traditional random forest). Ve studii pracovali s buňkami o velikosti 30 metrů, extrapolovali zjednodušená (v rámci buňky zprůměrovaná) měření z leteckého laserového skenování a jako vysvětlující proměnné použili data Landsat-8 (zejména vegetační indexy jako např. MSR, MSAVI a DVI; a jednotlivá pásma – SWIR, NIR a zelené). Pro všechny modely v článku vyšla RMSE necelé 2,2 metry, a i zde se potvrdila tendence modelů podhodnocovat vysokou vegetaci (vyšší než 95% kvantil) a nadhodnocovat vegetaci nízkou (nižší než medián souboru). Vzhledem k tomu, že použití 98% kvantilu (jak tomu bylo v této práci) nebylo v ostatních pracích běžné a nejčastěji se bral právě kvantil 95 %, je možné, že i to může být důvod podhodnocování modelu v této práci. Hodnota RMSE studie od Wang et al. (2021) je opět nižší než v této práci, i když by vzhledem k velikosti buněk a použití bodového mračna z leteckého laserového skenování (přestože zprůměrovaného) nejspíše mohl být ještě nižší.

Je neočekávané, že v této studii nevyšel index NDVI jako podstatná vysvětlující proměnná, protože to tak ve studiích obdobného rázu bývá (Li et al., 2020; Xi et al., 2022). Jak je výše zmíněno, obsahuje podobnou informaci jako pásmo Red edge, takže pro účely modelu má nejspíše i obdobnou funkci. Využití radarových dat a indexu EVI (přestože ten vysvětlil minimum variability) a jejich význam pro predikce koresponduje se studií Xi et al. (2022). Souhrn zvolených prediktorů je ve srovnání s jinými studiemi nečekaný kvůli absenci více vegetačních indexů a samostatných pásem (NIR, zelené atd.), ale zároveň s nimi není přímo kontrastní a 51 % vysvětlené variability lze vzhledem k relativně malé modelované ploše vnímat jako obstojný výsledek (Popescu et al., 2004; Liu et al., 2019).

Reakce zde prezentovaného modelu na hodnoty Red edge jsou také lehce překvapivé. Obecně by se nejspíše dalo očekávat, že čím více (vyšší) vegetace, tím více chlorofylu a tím i vyšší odrazivost tohoto pásma, nicméně graf částečné závislosti na obrázku 8 ukazuje, že model zjevně vycházel z jiných předpokladů. Protože se vlnové délky Red edge pohybují mezi červeným a blízkým infračerveným pásmem, mohlo by možným vysvětlením nečekaného průběhu křivky být, že v datech Sentinel byl jako Red edge zachytáván v zájmových územích spíše projev viditelného spektra (červeného pásma). To by vysvětlovalo nízké hodnoty odrazivosti, které jsou pro vegetaci a červené pásmo přibližně 10 % (Huete, 2004). Také je možné, že podle modelu stromy nebyly zdravé – měly málo chlorofylu, nicméně i tak by křivka nejspíše vypadala jinak (vyšší odrazivost). Rovněž by se dalo předpokládat, že v průběhu křivky část dat chybí, a tedy křivka není úplná – má neobvyklý tvar pro normální rozdělení dat. To by mohlo být zapříčiněno například rozdělením dat na trénovací a testovací. Nejspíše zde však jde o kombinaci množství různých vlivů, které nejsou snadno shrnutelné. Pásmo Red edge jako vysvětlující proměnná ve výsledném modelu vysvětlilo nejvíce variability (87,74 %), nicméně s přihlédnutím k informacím z grafu částečné závislosti a argumentům výše se může jeho přesnost zdát méně diskutabilní. Avšak celková vysvětlená variabilita a hodnoty chyb RMSE a ME se nijak zásadně nevymykají hodnotám očekávatelným od tohoto typu výzkumu. Pro další práce by mohlo být výhodnější místo pásma Red edge kvůli interpretovatelnosti použít některý z jeho indexů, například Normalized Difference Red Edge Index.

S křivkou částečné závislosti výšky vegetace na odrazivosti pásma Red edge poněkud kontrastuje průběh křivky grafu závislosti výšky vegetace na hodnotách EVI indexu. Z druhé zmíněné lze usuzovat, že podle modelu se vitalita stromů zvyšuje s

jejich výškou, čemuž ale neodpovídají hodnoty odrazivosti Red edge, jak je popsáno výše. Mimo hypotézy o chybě v predikci pomocí Red edge výše si tuto nejasnost nejspíše lze vyložit tak, že jsou zavádějící hodnoty indexu EVI a vysoká vegetace na lokalitách obecně spíše strádala, každopádně je možné soudit, že tento nesoulad reálně reflektuje jistou chybu v předpokladech modelu.

Fakt, že jednotlivé hodnoty RMSE a ME jsou si v zájmových územích řádově velmi podobné, naznačuje, že buď je ve zmíněných horách obdobné prostředí (a vegetace), nebo že je model vůči těmto nuancím robustní a v předpovědích konzistentní a predikce nebyly tolik ovlivněny například různou vegetací (ať už druh či věk), svažitostí, orientací a vývojem na území. Nejvyšší RMSE vyšla na Šumavě, kde jednak bylo v testovacím datasetu nejméně měření, takže se zde dá předpokládat vysvětlující souvislost a jednak se tam v současné době nejspíše nedá očekávat většina zdravých porostů (Hubený, 2015). Nejnižší chyby vyšly v Krušných horách, to může být dáno například odlišnou orientací tohoto pohoří oproti zbylým dvěma, jistě by to nemělo být zapříčiněno například malým množstvím testovacích dat.

Je otázkou, nakolik výsledky mohlo ovlivnit zprůměrování hodnot měření GEDI v rámci jedné buňky. Tím byl celý dataset vlastně umenšen, protože se mnoho řádků po zprůměrování ve výsledku nepoužilo kvůli duplikaci. Na druhou stranu se v jiných studiích větší buňky ukázaly naopak predikovatelnější, což poukazuje a to, že by toto zprůměrování nemělo mít negativní vliv. Významný vliv také nejspíše měla volba vysvětlované proměnné modelu (relativní naměřená výška), v této práci byl brán 98% kvantil výšky vegetace, ale není to jediná všeobecně využívaná hodnota a většina prací naopak pracuje s nižším (relativně častý je 95%) (Li et al., 2020; Xi et al., 2022). Zvolená predikovaná proměnná výrazně ovlivňuje výslednou schopnost modelu vysvětlit co nejvíce variability (ta se zvyšuje s vyšším kvantilem a model může mít větší tendence se predikcemi vymykat). Další významné proměnné, které zde nebyly do modelu zaneseny a v navazující práci by jistě bylo dobré je zařadit jsou například vertikální členitost, topografie a typ vegetačního pokryvu území. Zároveň by bylo nepochybně přínosné porovnat efektivitu více různých algoritmů, protože se jejich chování velmi liší a každý má vlastní specifika, jimiž může být výsledek ovlivněn nemalou měrou. V neposlední řadě je třeba zopakovat, že by bylo velmi vhodné a přínosné provést srovnání predikcí modelu s výškami vegetace naměřenými leteckým laserovým skenováním. Protože takto zatím není zcela jasné, zda by model mohl být využitelný i v realitě a jestli není trénován na nedostatečných datech. Pro

doplnění lze uvést, že ve studiích, které realizovali Quiros et al. (2021) a Liu et al. (2021) byla při měření výšky vegetace v různých prostředích (nížiny až hory) LiDARem GEDI dosažena RMSE 3,56 metrů, tedy hodnota víceméně odpovídající chybě modelů několika výše zmíněných prací.

6. Závěr

Tato diplomová práce se zabývala tvorbou modelu k extrapolaci výšky vegetace naměřené LiDARem GEDI a následným vyhodnocením přesnosti jeho predikcí na třech zájmových územích v horách na pomezí České republiky. Pearsonův korelační koeficient pro vztah predikovaných a LiDARem naměřených výšek vegetace vyšel 0,7 a tento výsledek koresponduje s jinými porovnávanými pracemi. Hodnoty zvolených chybových metrik RMSE a ME vyšly od výsledků jiných studií překvapivě odlišně. Zde dosažená ME -0,06 m naznačuje celkový trend podhodnocování modelu, přičemž modely jiných prací obecně spíše nadhodnocovaly, naopak výsledná hodnota RMSE (7,6 m), je oproti jiným pracím přibližně jednou tak vysoká a je sporné, čím přesně je to zapříčiněno. Může to být dáno mnoha faktory, například nezahrnutím typu vegetace či topografie do modelu.

Zde vytvořený model by bylo třeba rozšířit o vysvětlující proměnné a lépe prozkoumat jeho předpovědi v kontextu dalších proměnných prostředí, pak by bylo jasnější, jaká je jeho schopnost předpovídat a bylo by spíše možné zhodnotit jeho reálnou využitelnost. V navazující studii by bylo přínosné lépe nastavit rozdělení dat na trénovací a testovací část – aby z každé oblasti byla data v obou datasetech rozdělená odpovídajícím poměrem, následně by bylo možné lépe spekulovat o tom, na čem je chybovost modelu závislá a jestli i zde mohou mít vliv proměnné prostředí (druh vegetace, svažitost a jiné). S těmito výsledky a bez další validace je sporné, jestli nepřesné výsledky zapříčiňuje nedostatek dat, nejednoznačné a nepřesně měření či například pro model špatně předvídatelné prostředí (v každém ze zvolených zájmových území jsou podmínky do určité míry různé).

Modelování v oboru dálkového průzkumu Země je stále relativně nový přístup, který však vykazuje slibné výsledky. Obzvláště když k laserovému „skenování“ dochází stále jen z letadel a žádná tak efektivní technologie není připravena na větší měřítko – pro snímání ze satelitů. Extrapolace měření běžně získávaných tímto způsobem se tedy ukazuje jako intuitivní postup k získávání více dat za minimální investice navíc. Jak ukázaly výsledky této práce, prostoru pro pokrok a vylepšení modelování i získávání dat je mnoho. LiDARová data získávaná ve velkém množství jsou zásadním vstupem pro natrénování takovýchto predikčních modelů a s tím, jak je jich získáváno více a více ve větší kvalitě (přesnosti měření) se dá očekávat prudký rozvoj modelování v dálkovém průzkumu Země.

7. Reference

AOPK, 2021. Připravovaná velkoplošná chráněná území. In: *AOPK* [online]. [cit. 2021-12-28]. Dostupné z: https://www.ochranaprirody.cz/uzemni-ochrana/uzemi-pripravovana-k-vyhlaseni/pripravovana-velkoplosna-chranena-uzemi/.

BATTEN, G. D., 1998. Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Australian Journal of Experimental Agriculture* [online]. **38**(7) [cit. 2021-12-22]. ISSN 0816-1089. Dostupné z: doi:10.1071/EA97146.

BCCVL, 2021. *Boosted Regression Tree* [online]. [cit. 2022-03-20]. Dostupné z: https://support.bccvl.org.au/support/solutions/articles/6000083202-boosted-regression-tree?fbclid=IwAR3XoCawVOEVfceQd7haEjlDK3QsElg47ULh0Fioj3W35Wy6hvex7zKzGXs.

BROWNLEE, Jason, 2019. What is Deep Learning?. In: *Machine Learning Mastery: Making Developers Awesome at Machine Learning* [online]. [cit. 2022-01-08]. Dostupné z: https://machinelearningmastery.com/what-is-deep-learning/.

CLARK, John, 2018. Pearson Correlation Coefficient. In: *Magoosh Statistics Blog* [online]. [cit. 2022-03-10].

COLE, Steve, Karen NORTHON, ed., 2014. NASA Selects Instruments to Track Climate Impact on Vegetation. In: *NASA* [online]. Washington [cit. 2021-11-29]. Dostupné z: https://www.nasa.gov/press/2014/july/nasa-selects-instruments-to-track-climate-impact-on-vegetation.

COLIN, Brigitte, Samuel CLIFFORD, Paul WU, Samuel RATHMANNER et Kerrie MENGERSEN, 2017. Using Boosted Regression Trees and Remotely Sensed Data to Drive Decision-Making. *Open Journal of Statistics* [online]. **07**(05), 859-875 [cit. 2022-03-20]. ISSN 2161-718X. Dostupné z: doi:10.4236/ojs.2017.75061.

COPERNICUS, 2021. O programu Copernicus. In: *Copernicus: Europe's eyes on Earth* [online]. [cit. 2021-12-28]. Dostupné z: https://www.copernicus.eu/cs/o-programu-copernicus.

DAVID, Petr et Vladimír SOUKUP, 2020. *Krušné hory známé i neznámé*. Vydání první. Praha: Euromedia Group. Universum (Euromedia Group). ISBN 978-80-242-6686-2.

DORON, Nakar, 2019. *Sentinel-2: Multispectral Instrument (MSI) design and system performance* [online]. [cit. 2021-12-28]. Dostupné z: doi:10.13140/RG.2.2.27242.67529.

ESA, 2021. Sentinel-1 Product Modes. In: *Sentinel Online* [online]. [cit. 2022-03-28]. Dostupné z: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes.

ESRI, 2020a. Train Random Trees Classifier (Spatial Analyst). In: *Esri: ArcGIS Pro* [online]. [cit. 2022-03-19]. Dostupné z: https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-analyst/train-random-trees-classifier.htm.

ESRI, 2020b. Indices gallery. In: *Esri: ArcGIS Pro* [online]. [cit. 2022-01-08]. Dostupné z: https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm.

FRIEDMAN, Jerome H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* [online]. **29**(5), 1189-1232 [cit. 2022-03-19]. Dostupné z: https://www.jstor.org/stable/2699986.

GARBADE, Michael, 2018. Understanding K-means Clustering in Machine Learning. In: *Towards data science* [online]. [cit. 2022-01-08]. Dostupné z: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1.

GUPTA, Shweta, 2021. RMSE: What does it mean?. In: *Great Learning* [online]. [cit. 2022-03-13]. Dostupné z: https://medium.com/@mygreatlearning/rmse-what-does-it-mean-2d446c0b1d0e.

HEIDENREICH, Hunter, 2018. Machine Learning for the Average Person: What are the types of machine learning?. In: *Hunter Heidenreich* [online]. [cit. 2021-12-28]. Dostupné z: http://hunterheidenreich.com/blog/breaking_down_ml_for_the_average_person/.

HERNDON, Kelsey, Franz MEYER, Africa FLORES, Emil CHERRINGTON et Leah KUCERA, 2020. What is Synthetic Aperture Radar?. In: *Earth Data: Open acces for open science* [online]. [cit. 2021-12-28]. Dostupné z: https://earthdata.nasa.gov/learn/backgrounders/what-is-sar.

Historie krušnohorských lesů, 2019. In: *Krušné hory* [online]. [cit. 2021-12-28]. Dostupné z: http://www.krusnohorsky.cz/2019/07/24/historie-krusnohorskych-lesu/.

HOLČÍK, Jiří et Martin KOMENDA, ed., 2015. *Matematická biologie: e-learningová učebnice* [online]. 1. vydání. Brno: Masarykova univerzita [cit. 2022-03-13]. ISBN 978-80-210-8095-9. Dostupné z: https://portal.matematickabiologie.cz/.

HUBENÝ, Pavel, 2015. Po kůrovci... (20 let poté). *Ochrana Přírody* [online]. **2015**(2), 19-21 [cit. 2022-03-28]. Dostupné z: https://www.casopis.ochranaprirody.cz/pece-o-prirodu-a-krajinu/po-kurovci-20-let-pote/.

HUBENÝ, Pavel et Pavla ČÍŽKOVÁ, 2016. *Šumavské lesy pod lupou: co jste nevěděli o šumavských lesích: v podrobném monitoringu za léta 2009-2014*. [Vimperk]: Správa Národního parku Šumava. ISBN 978-80-87257-31-9.

HUETE, A.R., 2004. REMOTE SENSING FOR ENVIRONMENTAL MONITORING. *Environmental Monitoring and Characterization* [online]. Elsevier, s. 183-206 [cit. 2022-03-21]. ISBN 9780120644773. Dostupné z: doi:10.1016/B978-012064477-3/50013-8.

HUETE, A, K DIDAN, T MIURA, E.P RODRIGUEZ, X GAO et L.G FERREIRA, 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* [online]. **83**(1-2), 195-213 [cit. 2022-01-08]. ISSN 00344257. Dostupné z: doi:10.1016/S0034-4257(02)00096-2.

CHEN, L., 2010. A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data. *International Journal of Remote Sensing* [online]. **24**(11), 2265-2275 [cit. 2021-12-21]. ISSN 0143-1161. Dostupné z: doi:10.1080/01431160210154966.

CHEN, Li, Keh-Chia YEH, Hsiao-Ping WEI et Gin-Rong LIU, 2011. An improved genetic programming to SSM/I estimation typhoon precipitation over ocean.

Hydrological Processes [online]. **25**(16), 2573-2583 [cit. 2021-12-21]. ISSN 08856087. Dostupné z: doi:10.1002/hyp.8132.

IMAM, Ekwal, 2019. *Remote Sensing and GIS: Modelling in Remote Sensing and GIS* [online]. University Grand Commission (UGC) [cit. 2022-03-17]. Dostupné z: https://www.researchgate.net/publication/330383498_Remote_Sensing_and_GIS_Modelling_in_Remote_Sensing_and_GIS.

KAJZAROVÁ, Eva, 2012. *Lesy Krkonošského národního parku a péče o ně: Rosteme kolem vás* [online]. Správa Krkonošského národního parku [cit. 2021-12-29]. ISBN 978-80-86418-90-2. Dostupné z: https://www.krnap.cz/data/Files/downloads/krnap-pece-o-lesy-web_144558678328.238.pdf.

KRAETZIG, Nikita Marwaha, 2020. *What can NDVI tell us about plants?* [online]. [cit. 2022-03-28]. Dostupné z: https://up42.com/blog/tech/5-things-to-know-about-ndvi.

KRNAP, 2021a. Natura 2000. In: *Krkonošský národní park* [online]. [cit. 2021-12-29]. Dostupné z: https://www.krnap.cz/natura-2000/.

KRNAP, 2021b. Maloplošná zvláště chráněná území. In: *Krkonošský národní park* [online]. [cit. 2021-12-29]. Dostupné z: https://www.krnap.cz/maloplosna-zvlaste-chranena-uzemi/.

Krušnohorské lesy, 2018. In: *Krušné hory* [online]. [cit. 2021-12-28]. Dostupné z: https://www.krusnohorsky.cz/2018/02/03/krusne-hory/.

KUMAR, Akash, 2020. What is Supervised Learning and its different types?. In: *Edureka!* [online]. [cit. 2022-01-08]. Dostupné z: https://www.edureka.co/blog/supervised-learning/.

L3HARRIS, 2020. Vegetation Indices. In: *L3Harris: Geospatial* [online]. L3Harris [cit. 2021-12-28]. Dostupné z: https://www.l3harrisgeospatial.com/docs/VegetationIndices.html.

LANG, Nico, Nikolai KALISCHEK, John ARMSTON, Konrad SCHINDLER, Ralph DUBAYAH et Jan WEGNER, 2022. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sensing of Environment* [online]. **268** [cit. 2022-03-26]. ISSN 00344257. Dostupné z: doi:10.1016/j.rse.2021.112760.

LARY, David, Amir ALAVI, Amir GANDOMI et Annette WALKER, 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* [online]. **7**(1), 3-10 [cit. 2021-12-21]. ISSN 16749871. Dostupné z: doi:10.1016/j.gsf.2015.07.003.

LIANG, Shunlin et Jindi WANG, ed., 2020. *Advanced Remote Sensing: Terrestrial Information Extraction and Applications* [online]. 2nd Edition. Elsevier [cit. 2022-01-08]. ISBN 9780128158265. Dostupné z: doi:10.1016/C2017-0-03489-4.

LIU, Aobo, Xiao CHENG et Zhuoqi CHEN, 2021. Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment* [online]. **264** [cit. 2022-03-26]. ISSN 00344257. Dostupné z: doi:10.1016/j.rse.2021.112571.

LIU, Yanan, Weishu GONG, Yanqiu XING, Xiangyun HU et Jianya GONG, 2019. Estimation of the forest stand mean height and aboveground biomass in Northeast

China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* [online]. **151**, 277-289 [cit. 2022-03-26]. ISSN 09242716. Dostupné z: doi:10.1016/j.isprsjprs.2019.03.016.

LI, Wang, Zheng NIU, Rong SHANG, Yuchu QIN, Li WANG et Hanyue CHEN, 2020. High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation* [online]. **92** [cit. 2022-02-24]. ISSN 03032434. Dostupné z: doi:10.1016/j.jag.2020.102163.

MAKKEASORN, Ammarin, Ni-Bin CHANG, Mark BEAMAN, Chris WYATT et Charles SLATER, 2006. Soil moisture estimation in a semiarid watershed using RADARSAT-1 satellite imagery and genetic programming. *Water Resources Research* [online]. **42**(9) [cit. 2021-12-21]. ISSN 00431397. Dostupné z: doi:10.1029/2005WR004033.

NASA, 2021a. Mission overview. In: *GEDI: Ecosystem Lidar* [online]. [cit. 2021-12-28]. Dostupné z: https://gedi.umd.edu/mission/mission-overview.

NASA, 2021b. Instrument Overview. In: *GEDI: Ecosystem Lidar* [online]. [cit. 2021-12-28]. Dostupné z: https://gedi.umd.edu/instrument/instrument-overview/.

NASA, 2021c. *Specifications*. In: *GEDI: Ecosystem Lidar* [online]. [cit. 2021-12-28]. Dostupné z: https://gedi.umd.edu/instrument/specifications/.

NASA, 2021d. Data products. In: *GEDI: Ecosystem Lidar* [online]. [cit. 2021-12-28]. Dostupné z: https://gedi.umd.edu/data/products/.

NASA, 2021e. GEDI's ground sampling pattern. In: *GEDI: Ecosystem Lidar* [online]. [cit. 2022-03-28]. Dostupné z: https://gedi.umd.edu/instrument/specifications/.

NEUENSCHWANDER, Amy et Lori MAGRUDER, 2019. Canopy and Terrain Height Retrievals with ICESat-2: A First Look. *Remote Sensing* [online]. **11**(14) [cit. 2022-02-20]. ISSN 2072-4292. Dostupné z: doi:10.3390/rs11141721.

NIELSEN, Michael, 2019. *Neural Networks and Deep Learning* [online]. [cit. 2022-01-08]. Dostupné z: http://neuralnetworksanddeeplearning.com/index.html.

PALATINE, Ill, 2017. *MIDOPT* [online]. [cit. 2022-03-28]. Dostupné z: https://midopt.com/new-triple-bandpass-filters-for-ndvi-and-endvi/.

POPESCU, Sorin, Randolph WYNNE et John SCRIVANI, 2004. Fusion of Small-Footprint Lidar and Multispectral Data to Estimate Plot- Level Volume and Biomass in Deciduous and Pine Forests in Virginia, USA. *Forest Science* [online]. **50**(4), 551–565 [cit. 2022-03-26]. Dostupné z: doi:https://doi.org/10.1093/forestscience/50.4.551.

QUIROS, Elia, Maria-Eugenia POLO et Laura FRAGOSO-CAMPON, 2021. GEDI Elevation Accuracy Assessment: A Case Study of Southwest Spain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* [online]. **14**, 5285-5299 [cit. 2022-03-26]. ISSN 1939-1404. Dostupné z: doi:10.1109/JSTARS.2021.3080711.

RIEDL, Marcel, Luděk ŠIŠÁK, Josef KAHUDA, Tomáš HOFMEISTER, Radomír ULRICH, Milena BĚLSKÁ, Jiří BÍLÝ, Petr DUŠEK, Petr DVOŘÁK, 2021. *ZPRÁVA O STAVU LESA A LESNÍHO HOSPODÁŘSTVÍ ČESKÉ REPUBLIKY V ROCE 2020* [online]. Praha: Ministerstvo zemědělství [cit. 2022-01-08]. ISBN 978-

80-7434-625-5. Dostupné z: https://eagri.cz/public/web/file/688968/Zprava_o_stavu_lesa_2020_web.pdf.

Sentinel Online [online], 2021. ESA [cit. 2021-12-28]. Dostupné z: https://sentinels.copernicus.eu/.

STEPHANIE, I., 2016. Boosted Regression Trees for Giant Gourami Distribution. In: *Oregon State University* [online]. [cit. 2022-03-20]. Dostupné z: https://blogs.oregonstate.edu/geo599spatialstatistics/2016/06/05/boosted-regression-trees-giant-gourami-distribution/.

STERNER, Robert et James ELSER, 2003. *Ecological Stoichiometry* [online]. Princeton University Press [cit. 2022-03-14]. ISBN 9781400885695. Dostupné z: doi:10.1515/9781400885695.

ŠTURSA, Jan, 2011. *Krkonoše/Karkonosze: Přeshraniční biosférická rezervace UNESCO* [online]. Správa Krkonošského národního parku [cit. 2021-12-28]. Dostupné z: https://www.krnap.cz/data/File/letaky_brozury/biorezervace_cz_03.pdf.

WANG, Hui, Travis SEABORN, Zhe WANG, Christopher CAUDILL et Timothy LINK, 2021. Modeling tree canopy height using machine learning over mixed vegetation landscapes. *International Journal of Applied Earth Observation and Geoinformation* [online]. **101** [cit. 2022-03-26]. ISSN 03032434. Dostupné z: doi:10.1016/j.jag.2021.102353.

XI, Zhilong, Huadong XU, Yanqiu XING, Weishu GONG, Guizhen CHEN et Shuhang YANG, 2022. Forest Canopy Height Mapping by Synergizing ICESat-2, Sentinel-1, Sentinel-2 and Topographic Information Based on Machine Learning Methods. *Remote Sensing* [online]. **14**(2) [cit. 2022-03-26]. ISSN 2072-4292. Dostupné z: doi:10.3390/rs14020364.

XUE, Jinru et Baofeng SU, 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors* [online]. **2017**, 1-17 [cit. 2021-12-22]. ISSN 1687-725X. Dostupné z: doi:10.1155/2017/1353691.