

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ
KATEDRA PROSTOROVÝCH VĚD

**Prostorová interpolace výšky vegetace z dat
GEDI a ICESat-2**
DIPLOMOVÁ PRÁCE

Vedoucí práce: Ing. Vojtěch Barták, Ph.D.
Odborný konzultant: doc. Ing. Vítězslav Moudrý, Ph.D.
Diplomant: Bc. Ondřej Coufal

2024

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Ondřej Coufal

Prostorové vědy v životním prostředí

Název práce

Prostorová interpolace výšky vegetace z dat GEDI a ICESat-2

Název anglicky

Spatial interpolation of vegetation height from the GEDI and ICESat-2 data

Cíle práce

Provést prostorovou interpolaci výšky vegetace z dat satelitní laserové altimetrie misí GEDI a ICESat-2.

Metodika

Existující přístupy k modelování výšky vegetace z dat satelitní laserové altimetrie v místech chybějících měření doplňují hodnoty na základě modelovaného vztahu mezi výškou vegetace a dalších, externích proměnných (např. optická či radarová data z misí Sentinel). V této práci bude na pilotní lokalitě vyzkoušen přístup využívající prostorovou interpolaci, založenou na prostorové autokorelaci výšky vegetace. Interpolační metodou bude kriging. Práce jednak posoudí míru autokorelace přítomné v datech, dále provede interpolaci a výsledky porovná jednak z výškou vegetace z letského laserového skenování, jednak s existujícími produkty vegetačně výškových map vytvořených z dat GEDI a ICESat-2.

Doporučený rozsah práce

40 – 60 stran

Klíčová slova

Spatial interpolation, Kriging, Satellite LiDAR, Vegetation height, Canopy model

Doporučené zdroje informací

- Lang N., Schindler K., Wegner JD. 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sensing of Environment* 233. <https://doi.org/10.1016/j.rse.2019.111347>
- Lefsky MA., 2009. A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*. <https://doi.org/10.1029/2010GL043622>
- Potapov P., et.al. 2020 Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment* 253. <https://doi.org/10.1016/j.rse.2020.112165>
- Simard M., 2011. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research*. <https://doi.org/10.1029/2011JG001708>

Předběžný termín obhajoby

2023/24 LS – FŽP

Vedoucí práce

Ing. Vojtěch Barták, Ph.D.

Garantující pracoviště

Katedra prostorových věd

Elektronicky schváleno dne 18. 3. 2024

prof. Ing. Petra Šimová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 19. 3. 2024

prof. RNDr. Michael Komárek, Ph.D.

Děkan

V Praze dne 27. 03. 2024

Čestné prohlášení

Prohlašuji, že jsem diplomovou práci na téma: „Prostorová interpolace výšky vegetace z dat GEDI a ICESat-2“ vypracoval samostatně a citoval jsem všechny informační zdroje, které jsem v práci použil a které jsem rovněž uvedl na konci práce v seznamu použitých informačních zdrojů.

Jsem si vědom, 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 vědom, 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 dne
(podpis autora práce)

Poděkování

Rád bych poděkoval svému vedoucímu práce Ing. Vojtěchu Bartákovi, Ph.D., a mému odbornému konzultantovi doc. Ing. Vítězslavu Moudrému, Ph.D., za přátelský a profesionální přístup a trpělivé vedení této diplomové práce. Děkuji také své rodině a přítelkyni za neustálou podporu a motivaci.

Abstrakt

Tato diplomová práce se zabývá metodami vytváření modelů výšky vegetace (Canopy Height Model – CHM) z dat pořízených v rámci misí satelitní laserové altimetrie GEDI a ICESat-2. Cílem je porovnat různé geostatistické metody prostorové interpolace, jako jsou ordinary kriging a universal kriging, s ohledem na jejich efektivitu při modelování výšky vegetace. Práce také klade důraz na analýzu prostorové autokorelace a její přítomnost v měřených bodech výšky vegetace z obou misí satelitních LiDARů. Pro výsledné modely CHM bylo zvoleno prostorového rozlišení 10 metrů pro zajištění porovnatelnosti s existujícími globálními CHM, vytvořenými z dat GEDI pomocí strojového učení. Použité metody filtrace mohou spolu s výsledky analýzy prostorové autokorelace, modely výšky vegetace a jejich porovnání s globálními CHM, představovat základ pro další výzkum v oblasti využití krigingu, jako alternativní metody prostorové interpolace pro tvorbu CHM, které dávají porozumění dynamice vegetačního pokryvu a jeho vlivu na globální klimatické procesy. Zavedeny byly pokročilé metody filtrace kvalitních měření výšky vegetace, které vykazují menší redukci dat GEDI a ICESat-2 než tradiční používané filtrace dat pro tyto satelitní mise. Dále byla v datech obou misí ověřena vysoká míra autokorelace, která byla popisována pomocí empirických variogramů. Na tomto základě byly vytvořeny čtyři varianty CHM, které kombinovaly přístupy Ordinary Krigingu (OK) a Universal Krigingu (UK). Pro UK modely vyšly v křížové validaci, a ve validaci s modelem výšky vegetace z ALS pro rok 2012, nejmenší chyby ME, MAE a RMSE. Kvalita všech CHM, prezentovaných v práci, byla také větší než u existujících globálních modelů výšky vegetace z dat GEDI a ICESat-2.

Klíčová slova: prostorová interpolace, kriging, satelitní LiDAR, výška vegetace, model výšky vegetace, GEDI, ICESat-2, CHM

Abstract

This thesis deals with the methods of creating Canopy Height Models (CHM) from data acquired by the GEDI and ICESat-2 satellite laser altimetry missions. The aim is to compare different geostatistical spatial interpolation methods, such as ordinary kriging and universal kriging, with respect to their effectiveness in vegetation height modelling. The paper also emphasizes on the analysis of spatial autocorrelation and its presence in the measured vegetation height points from both satellite LiDAR missions. A spatial resolution of 10 meters was chosen for the resulting CHM models to ensure comparability with existing global CHMs created from GEDI data using machine learning. The filtering methods used, together with the results of the spatial autocorrelation analysis, vegetation height models and their comparison with global CHMs, can form the basis for further research into the use of kriging as an alternative spatial interpolation method for the production of CHMs that give an understanding of vegetation cover dynamics and its influence on global climate processes. Advanced methods for filtering high-quality vegetation height measurements have been introduced that show less reduction of GEDI and ICESat-2 data than the traditional data filtering used for these satellite missions. In addition, a high degree of autocorrelation was verified in the data from both missions and described using empirical variograms. On this basis, four CHM variants were developed combining the Ordinary Kriging (OK) and Universal Kriging (UK) approaches. For the UK models, the ME, MAE and RMSE errors came out the smallest in the cross-validation, and in the validation with the 2012 ALS vegetation height model. The quality of all CHMs presented in the paper was also greater than that of existing global vegetation height models from GEDI and ICESat-2 data.

Key words: spatial interpolation, kriging, satellite LiDAR, vegetation height, canopy height model, GEDI, ICESat-2, CHM

Obsah

1. Úvod.....	12
2. Literární řešerše	14
2.1 Modely výšky vegetace.....	14
2.2 Satelitní lidary a jejich využití	18
2.2.1 GEDI	18
2.2.2 ICESat-2.....	24
2.3 Modely výšky vegetace z dat satelitní laserové altimetrie.....	29
2.4 Metody prostorové interpolace	33
2.4.1 Prostorová interpolace.....	33
2.4.2 Geostatistika a koncept variogramu	36
2.4.3 Základy krigingu	41
2.4.4 Typy krigingu.....	43
3. Metodika	46
3.1 Stanovení kritérií a metod tvorby modelu.....	46
3.2 Získávání dat	49
3.2.1 Stahování a výběr dat	49
3.2.2 Porovnání zdrojů dat GEDI.....	51
3.3 Předzpracování dat	54
3.3.1 Výběr vzorku.....	54
3.3.2 Metoda filtrace a korekce dat	57
3.3.3 Předzpracování prediktorů	59
3.4 Geostatistické modelování	61
3.4.1 Analýza prostorové struktury a korelace.....	61
3.4.2 Ordinary kriging.....	63
3.4.3 Universal kriging.....	65

3.5	Postup validace modelu.....	67
4.	Výsledky	70
4.1	Výsledky filtrace	70
4.2	Prostorová struktura a korelace.....	74
4.3	Modely výšky vegetace.....	75
4.4	Porovnání modelů s globálními CHM	79
5.	Diskuse	80
6.	Závěr.....	84
7.	Seznam použité literatury.....	87
	Seznam obrázků	98
	Seznam tabulek	101
	Přílohy	102

Seznam zkratek

CHM – Canopy Height Model (model výšky vegetace)

LiDAR – Light Detection And Ranging

GEDI – Global Ecosystem Dynamics Investigation (mise satelitního LiDARu)

ICESat-2 – The Ice, Cloud, and land Elevation Satellite-2 (mise satelitního LiDARu)

ALS – Airborne Laser Scanning (letecké laserové skenování)

DN – Digital Number

ABA – Area-Based Approach

AGB – Above Ground Biomass (nadzemní biomasa)

nDSM – Normalised Digital Surface Model (normalizovaný digitální model povrchu)

DSM – Digital Surface Model (digitální model povrchu)

DEM – Digital Elevation Model (digitální výškový model)

DTM – Digital Terrain Model (digitální model terénu)

SAR – Synthetic Aperture Radar (radar se syntetickou aperturou)

DAP – Digital Airborne Photogrammetry (digitální letecká fotogrammetrie)

ISS – International Space Station (Mezinárodní vesmírná stanice)

NIR – Near Infrared (blízké infračervené spektrum)

RH – Relative Height (metrika relativní výšky vegetace)

CDF – Cumulative Distribution Function (kumulativní distribuční funkce)

EOSDIS – NASA's Earth Observing System Data and Information System

DAAC – Data Archive and Analysis Center

LP DAAC – Land Process DAAC

ORNL DAAC – Oak Ridge National Laboratory DAAC

ATBD – Algorithm Theoretical Basis Documents (dokumentace algoritmů)

ATLAS – Advanced Topographic Laser Altimeter System

PRF – Pulse Repetition Frequency (pulzní frekvence u LiDARu)

PMT – Photomultiplier Tube (fotonásobič)

NISAR – NASA-ISRO Synthetic Aperture Radar (radarová mise)

DOE – Diffractive Optical Elements (difrakční optický element)

ISF – Instrument Support Facility (oddělení NASA)

NSIDC – National Snow and Ice Data Center

UMD GLAD – The University of Maryland, Global Forest Canopy Height

ARD – Landsat Analysis Ready Data (GLAD ARD)

ME – Mean Error (střední chyba)

RMSE – Root Mean Square Error (střední kvadratická chyba)

MAE – Mean Absolute Error (střední absolutní chyba)

CNN – Convolutional Neural Network (konvoluční neuronové sítě)

UMD – The University of Maryland

NN – Nearest Neighbor (nejbližší soused)

TIN – Triangulated Irregular Network (triangulovaná nepravidelná síť)

IDW – Inverse Distance Weighting (metoda vážené inverzní vzdálenosti)

GRF – Gaussian Random Field (Gaussovo náhodné pole)

SK – Simple Kriging (jednoduchý kriging)

OK – Ordinary Kriging (obyčejný kriging)

UK – Universal Kriging (univerzální kriging)

RK – Regression Kriging (regresní kriging)

KRNAP – Krkonošský národní park

EDSC – Earthdata Search Client

GEE – Google Earth Engine

SF – Simple Feature (standard pro formát prostorových objektů)

IQR – Interquartile range (mezikvartilové rozpětí)

NDVI – Normalized Difference Vegetation Index (norm. rozdílový vegetační index)

NDMI – Normalized Difference Moisture Index (norm. rozdílový index vlhkosti)

EVI – Enhanced Vegetation Index (vegetační index EVI)

1. Úvod

Modely výšky vegetace (CHM) se staly nepostradatelným nástrojem pro sledování dynamiky vegetace. Ve většině případů se jedná o data pořízená leteckým laserovým skenováním (ALS), která poskytují přesný a detailní popis struktury vegetace. Omezením těchto dat jsou však vysoké náklady na pořízení a limitace v pravidelné aktualizaci. V tomto kontextu data ze satelitních misí GEDI a ICESat-2 nabízí cennou alternativu. Jejich klíčové výhody jsou globální pokrytí, neustálá aktualizace a snadná dostupnost, čímž otevírají nové možnosti tvorby CHM.

Tato diplomová práce se zaměřuje na zavedení a ověření nové metodiky využívající kriging pro tvorbu CHM z dat těchto satelitních misí. Výhodou krigingu je jeho přímá kvantifikace nejistot spojených s odhady. Cílem je tak nejen prozkoumat potenciál krigingu, jako alternativní metody prostorové interpolace, ale také položit základy pro budoucí výzkum v této oblasti. Přestože existují některé studie zaměřené na tvorbu digitálních výškových modelů (DEM) pomocí krigingu (Narin, Gullu, 2023), aplikace této metody pro tvorbu modelů výšky vegetace z dat GEDI a ICESat-2 představuje nový směr využití geostatistických metod. Zásadním krokem pro ověření této metodiky bylo vybrání území Krkonošského národního parku (KRNAP), které díky své členitosti a specifickým podmínkám poskytuje ideální prostředí pro testování zvolených metod. Vyšší hustota dat GEDI v severních Čechách, způsobená specifickými orbitami ISS (s přístrojem GEDI) měla poskytnout důležitý předpoklad pro kriging. Původní hypotéza předpokládala, že by se díky použití adekvátního postupu filtrace v měřeních výšky vegetace z GEDI a ICESat-2 mohla vyskytovat prostorová autokorelace. Tato práce si tak zároveň klade za cíl prozkoumat její přítomnost a vhodně ji modelovat tak, aby kriging mohl poskytnout přesné a spolehlivé odhady na neznámých místech. Výsledkem by byly modely výšky vegetace (CHM), které jsou důležitým nástrojem pro sledování vegetace a dynamiky lesů. Důležitým aspektem práce bylo také porovnání vytvořených modelů CHM s existujícími globálními modely od autorů Lang et al. (2022a) a Potapov et al. (2021), což by umožnilo nejen ověření této nově navržené metodiky, ale i její přímé srovnání s již etablovanými přístupy.

Celkově by tato práce měla představovat pokus o zavedení krigingu jako nového nástroje pro tvorbu CHM z volně dostupných dat satelitních LiDARů. Měla

by také být důležitým krokem vpřed v porozumění možnostem a limitacím této techniky. Cílem je nejen ověřit schopnost krigingu vytvářet přesné modely výšky vegetace, ale také poskytnout základ pro další výzkum v oblasti využití geostatistiky pro tvorbu CHM z veřejně dostupných dat misí satelitní laserové altimetrie.

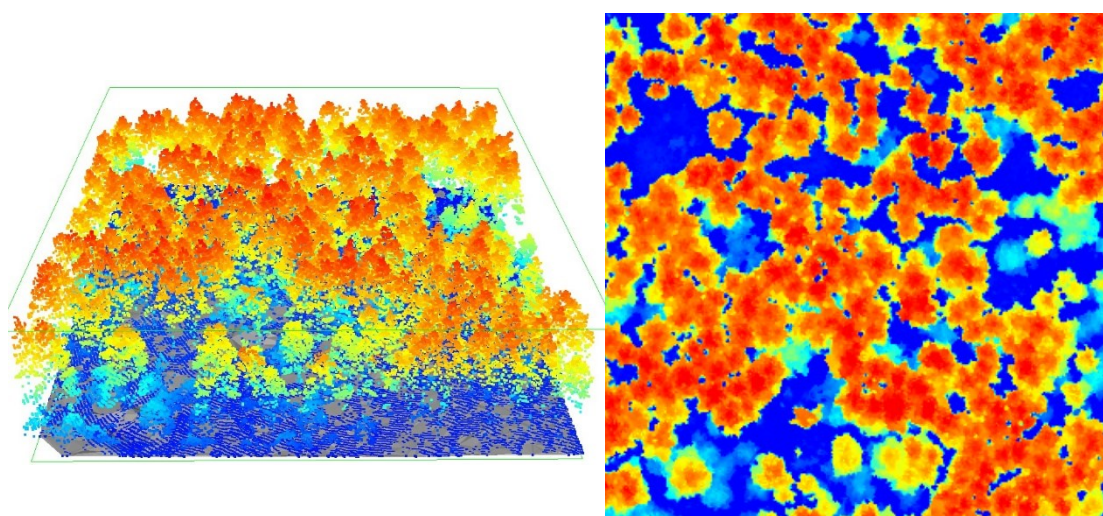
Cíle práce

Hlavním cílem práce tak bylo provést prostorovou interpolaci výšky vegetace z dat satelitní laserové altimetrie misí GEDI a ICESat-2 s využitím krigingu a zavést tak metodiku využívající geostatistické přístupy spolu s průzkumem jejich limitů. Současně je cílem práce také prozkoumat přítomnost prostorové autokorelace ve výšce vegetace z dat GEDI a ICESat-2 a posoudit kvalitu modelu porovnáním s referenčními daty z leteckého laserového skenování, včetně porovnání s existujícími globálními CHM využívající podobná data pro modely strojového učení.

2. Literární rešerše

2.1 Modely výšky vegetace

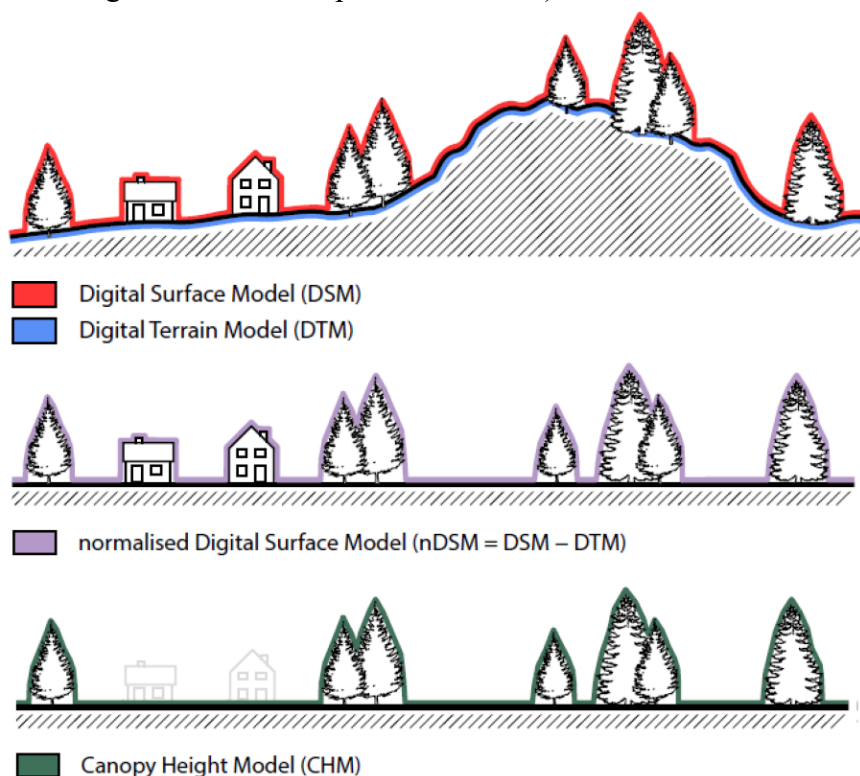
Model výšky vegetace, zkráceně CHM (z anglického Canopy Height Model), je produkt, nejčastěji odvozený z dat pořízených metodou LiDAR (Light Detection And Ranging), který je digitální reprezentací výšky nebo struktury vegetačního pokryvu. Ve formě rastru poskytuje pro každý pixel informace o absolutní výšce vegetace v hodnotách digital number (DN) (Wang et al., 2021). Tímto CHM nabízí detailní pohled na strukturu vegetace, jejíž studium je klíčové pro nejrůznější aplikace v lesnictví, ekologii a environmentálních vědách. Bakx et al. (2019) zdůrazňují důležitost modelů výšky vegetace pro efektivní odvození různých vegetačních parametrů nebo dendrometrických veličin, které jsou klíčové pro tvorbu nebo analýzu lesní inventarizace, modelování struktury a monitorování těžby či obnovy lesa, nebo při posuzování vhodnosti stanovišť pro lesní živočichy v kontextu modelování druhové distribuce (Bakx et al., 2019). Pokud se tyto kvantitativní plošné vegetační parametry počítají na větších plochách, často se tato metoda v lesnictví či ekologii označuje jako Area-Based Approach (ABA). Z hlediska ABA přináší práce s CHM velké výhody v rychlosti zpracování, protože jsou rastrové analýzy výpočetně jednodušší než analýzy bodového mračka. Modely výšky vegetace se také dají použít k výpočtu přesného odhadu nadzemní biomasy (AGB), což je například zásadní pro pochopení dynamiky uhlíku v lesních ekosystémech (Loh et al., 2020).



Obrázek 1: Ukázka normalizovaného bodového mračka obsahujícího strukturu vegetace v absolutních výškách nad terénem (vlevo) a odvozeného rastru CHM (vpravo). Zdroj: Isenburg, 2014

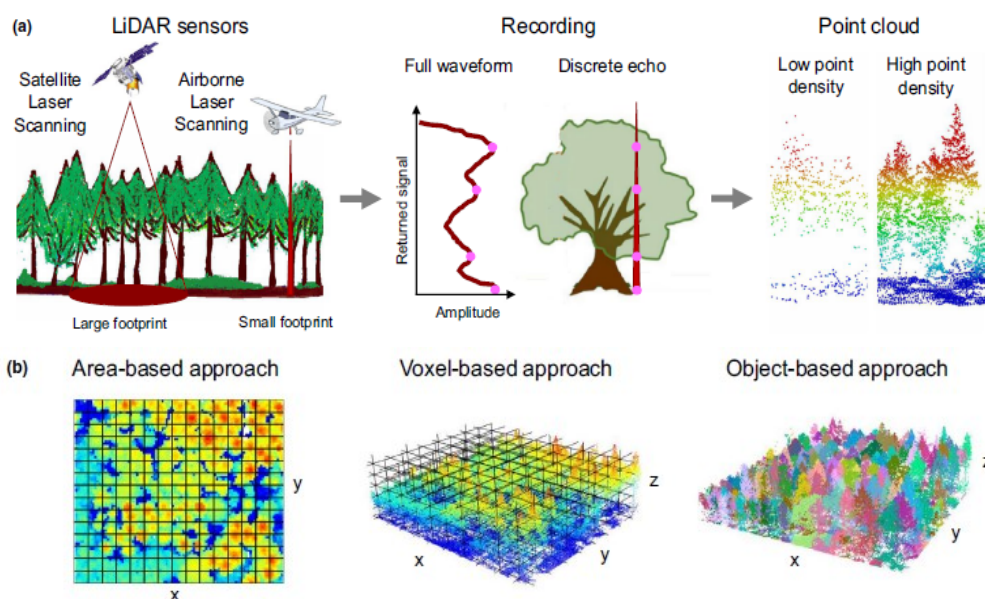
Ze studie Vandendaele et al. (2021) vyplívá, že má využití CHM v lesní inventarizaci také velký potenciál díky možnosti efektivní detekce, měření či segmentaci jednotlivých stromů. Pro identifikaci jednotlivých stromů se často využívá detekce špiček korun, například pomocí filtru lokálního maxima.

CHM je speciální formou normalizovaného digitálního modelu povrchu (nDSM), ve které je však obsažena pouze výška vegetace. Při tvorbě nDSM se odečítá digitální model terénu (DTM) od digitálního modelu povrchu (DSM), tedy $nDSM = DSM - DTM$. Tímto krokem se sice získá absolutní výška vegetace, ale kromě ní se v nDSM promítají i výšky jiných objektů nad terénem, například zástavby. Pro vytvoření plnohodnotného CHM, který zahrnuje pouze vegetaci, je tedy nutné tyto nežádoucí objekty z nDSM odstranit (viz. Obrázek 2). Existuje mnoho metod, jak vygenerovat CHM, ale většina z nich pracuje s rozdílem mezi DSM a DTM. Pokud máme například bodové mračno z leteckého laserového skenování (ALS), nebo jiných lidarových metod zachytávajících diskrétní návraty, často se například uplatňuje princip prostorové interpolace prvních návratů a určování jejich absolutní výšky nad DTM (Liu, Dong, 2014; Khosravipour et al., 2014)



Obrázek 2: Schematické znázornění rastrů, které zahrnují nadmořskou výšku terénu (DSM, DTM) a absolutní výšku nad terénem (nDSM, CHM), odvozených z lidarových bodových mračen. Zdroj: Moudrý, 2021

Tvorbě CHM předchází bodové mračno, které je reprezentováno nepravidelným rozložením návratů (tj. datových bodů) v trojrozměrném prostoru. Kromě toho, že se nejčastěji jedná o výstup z leteckého laserového skenování (ALS), mohou být data pořízena také pomocí radaru se syntetickou aperturou (SAR) nebo digitální letecké fotogrammetrie (DAP) a využít se přitom dá i různých nosičů senzorů (bezpilotních letadel, leteckých nebo kosmických) (Valbuena et al., 2020; Bergen et al., 2009). Z možných kombinací senzorů a platform má však použití letadla jako nosiče senzorů LiDAR pro většinu aplikací velkou výhodu, protože poskytuje souvislé a husté pokrytí poměrně velkých studovaných ploch. To je v kontrastu se satelitními laserovými altimetry, jako je GEDI nebo ICESat-2, které poskytují větší pokrytí, ale obsahují jen řídká diskrétní měření (Marselis et al., 2022; Moudrý et al., 2022). GEDI je navíc omezen na oblast mezi 51,6° s. š. a 51,6° j. š. (Dubayah et al., 2020). Drony na druhé straně poskytují daleko větší podrobnosti a hustotu bodového mračna než ALS, dokážou ale pokrýt jen malé oblasti (Calders et al., 2020). Ze zmíněných metod pořízení dat (LiDAR, DAP a SAR) má však jednoznačnou výhodu LiDAR díky schopnosti zachytit terén pod vegetačním krytem (Stereńczak et al., 2016).



Obrázek 3: Sběr a zpracování dat LiDAR: a) Srovnání satelitního a leteckého LiDARu, zahrnující rozdíly v návratových signálech a hustotě bodů. b) Metody zpracování dat: plošné jako ABA (např. CHM), voxelové a objektové segmentace. Zdroj: Bakx et al., 2019

Globální data výšky vegetace (CHM) jsou od roku 2020 celosvětově dostupná v rámci databáze Global Canopy Forest Height 2019 v rozlišení 30 m (Potapov et al., 2021). Lang et al. (2022a) dále vytvořili z dat GEDI pro rok 2020 první globální model výšky vegetace s rozlišením 10 m. Toto globální CHM v roce 2022 zároveň zpřístupnili veřejnosti. Význam těchto globálních modelů může vyplívat z toho, že

lesy, které pokrývají více než 4,1 miliardy hektarů zemského povrchu, hrají klíčovou roli v globálním koloběhu uhlíku (Pan et al., 2011). V kontextu Pařížské dohody o změně klimatu a Agendy pro udržitelný rozvoj 2030 se otázka zásoby uhlíku v lesích a její dynamiky stává klíčovou pro mnoho států a data z CHM pro ně mohou sloužit při tvorbě akčních plánů v oblasti klimatu (Hein et al., 2018). Lesy zkrátka fungují jako mohutný rezervoár uhlíku a ovlivňují tak klima naší planety (Pan et al., 2011). Ze studie Goetz, Dubayah (2011) vyplývá, že jsou údaje o zásobách uhlíku a jejich dynamice v lesích (například odlesňování, přírůst atp.) omezené zvláště v rozvojových zemích a pomohla by tak daleko větší integrace satelitních lidarů pro odhad nadzemní biomasy (AGB). Další studie zmiňují, že ačkoliv je metoda leteckého skenování (ALS) přesná a data jsou některými vyspělými zeměmi poskytována zdarma, data nejsou příliš často aktualizována, což je pro monitorování přírůstu a úbytku lesa nezbytné. Data z ALS mají navíc poměrně velké pořizovací náklady a terénní lesní inventarizace je zase náročná na pracovní sílu, proto má využití dat ze satelitních lidarů velký potenciál, zvláště při mapování lesů na velkých plochách nebo v globálním měřítku (Mulverhill et al., 2022; Herold et al., 2019).

Moudrý et al. (2023) v článku diskutují obavu, že snadná dostupnost globálních CHM v rastrovém formátu může znamenat riziko, že se tato data začnou využívat neuváženě a začnou nahrazovat využití přesnějších dat ALS. Jako příklad autoři uvádějí studii Lewis et al. (2022), ve které se počítaly průměry a rozptyly výšky stromů z databáze Global Canopy Forest Height 2019 (Potapov et al., 2021), ačkoliv byla pro dané území za stejné období veřejně dostupná i data z ALS. Použití globálních CHM přitom při modelování distribuce druhů nebo studiu biodiverzity na lokálním měřítku může vést k chybným výsledkům (Meyer, Pebesma, 2022). Bodová mračna z ALS, narozdíl od CHM, zvláště v globálním měřítku, navíc poskytují mnohem bohatší informace o struktuře vegetace, které jsou cenné pro analýzy využití stanovišť a rozšíření druhů, ale zpracování těchto dat je obtížné pro většinu ekologů (Moudrý et al., 2023). V článku Moudrý et al. (2023) se proto dále navrhuje konkrétní soubor proměnných popisujících strukturu vegetace (vegetačních metrik), který by byl snadno dostupný výzkumníkům a zainteresovaným stranám prostřednictvím datových portálů (jedinými dostupnými odvozenými produkty z dat ALS jsou zatím jen DTM a DSM), protože se autoři domnívají, že je tento přístup důležitý pro pokrok v ekologickém výzkumu a výzkumu biodiverzity.

2.2 Satelitní lidary a jejich využití

2.2.1 GEDI

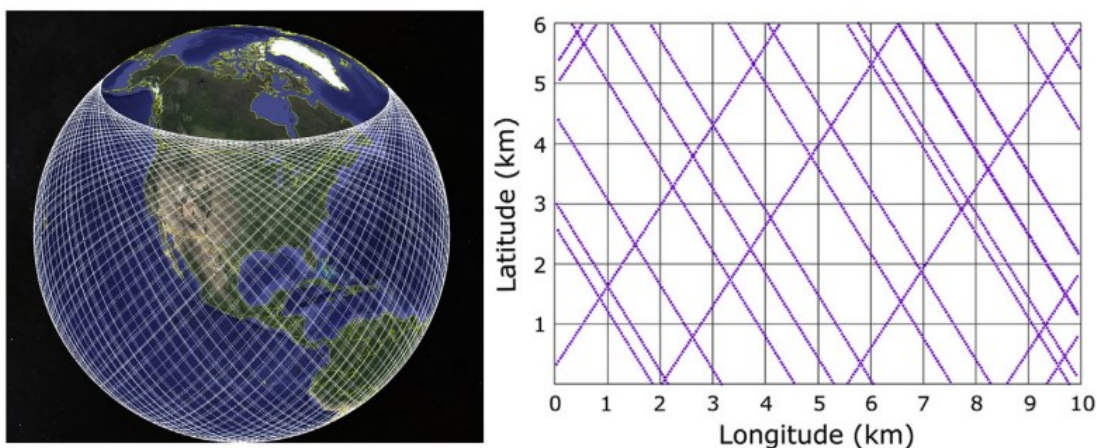
Global Ecosystem Dynamics Investigation (GEDI) je významná satelitní lidarová mise, při které NASA v roce 2018 vyslala stejnojmenný přístroj na Mezinárodní vesmírnou stanici (ISS) (Dubayah et al., 2022). Primárním cílem GEDI je zlepšit naše chápání role lesních ekosystémů v globálním uhlíkovém cyklu (Lang et al., 2022b). GEDI dosahuje těchto cílů sběrem dat o výšce a vertikální struktuře vegetace společně s výškou terénu (Dubayah et al., 2020). Kromě toho pomáhá globálně monitorovat a zaznamenávat množství nadzemní biomasy (AGB) v různých klimatických oblastech a vegetačních typech, což ukazuje svou univerzálnost v environmentálním výzkumu (Morin et al., 2022).

Jedná se tedy o tzv. full waveform LiDAR (FW), který se formou ukládání návratu odlišuje od často využívané technologie s diskrétními návraty (např. v ALS) a poskytuje tím unikátní možnost pozorovat globální strukturu vegetace (Fayad et al., 2021a). Satelitní mise pokrývá geografický rozsah od 51,6° severní šířky do 51,6° jižní šířky (orbitální hranice ISS) a efektivně pracovat může pouze v oblastech s relativně malou oblačností. Hustá oblačnost blokuje průchod laserových paprsků a znemožňuje tím přesná měření (Dubayah et al., 2020; Hancock et al., 2019). Dubayah et al. (2020) ve svém článku přímo uvádějí podrobné specifikace GEDI:

- 3 lasery (*Nd:YAG*)
- Vlnová délka: 1064 nm (~NIR)
- Rychlost pulzu (PRF): 242 Hz
- Velikost stopy (footprint): ~25 m (tvar kruhu o průměru ~25 m)
- Vzdálenost mezi středy stop (shots): 60 m
- 4 paprsky (2x coverage beams, 2x power beams)
- 8 pozemních drah (ground tracks)
- Vzdálenost mezi dráhami: 600 m

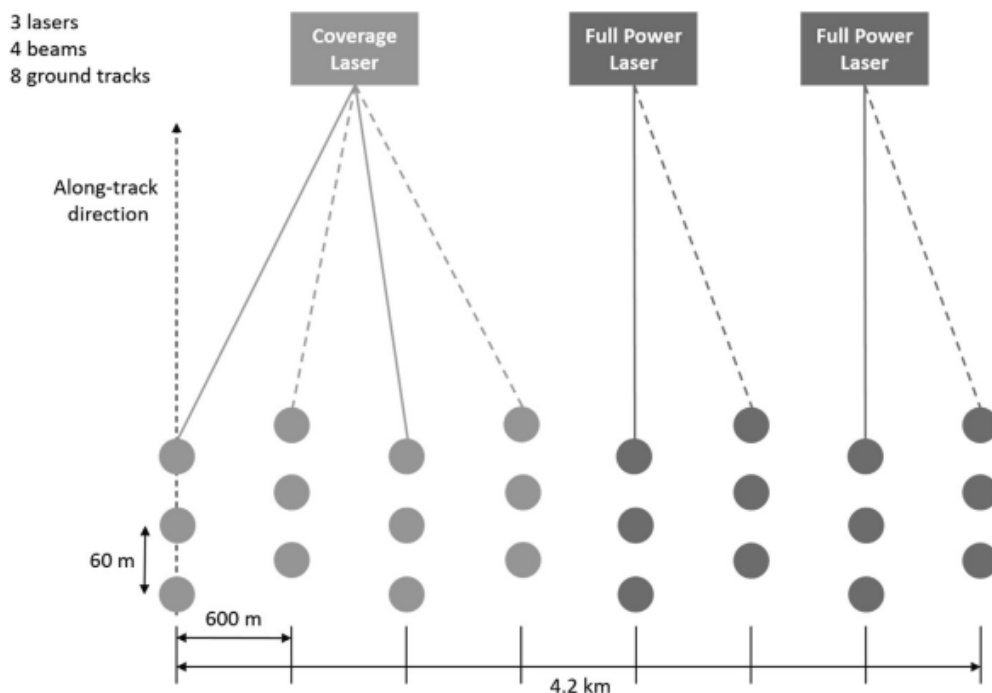
System se tedy skládá ze 3 laserů *Nd:YAG* z nichž jeden je rozdělen na dva paprsky (*coverage beams*), zatímco ostatní dva zůstávají v plném výkonu (*power beams*). Každý laser tak vystřelí 242krát za sekundu 4 paprsky o vlnové délce 1064 nm. Posunutím každého druhého výstřelu vznikne dohromady 8 pozemních drah ve vzdálenosti 600 m od sebe (šířka pokrytí v jednom orbitu ~4,2 km). Jeden laserový pulz po dopadu na zem (shot) má tvar kruhu o průměru ~25 m a středy těchto jednotlivých stop (shots) jsou od sebe vzdáleny 60 m (Dubayah et al., 2020).

Na následujícím Obrázek 4 je patrné, že pokrytí dat GEDI je nejhorší na rovníku, a naopak u horní ($51,6^\circ$ s. š.) nebo spodní ($51,6^\circ$ j. š.) hranice orbitu ISS jsou dráhy letu zhuštěny. Z toho vyplývá i hypotéza této práce, že na území ČR (sever Čech je blízko vrchní hranice orbitu) by GEDI mělo mít hodně měření, protože se tam ISS otáčí. Datové body (shoty) z GEDI mají však stále poměrně malou hustotu, a proto je například při tvorbě CHM nutno použít metody prostorové interpolace.



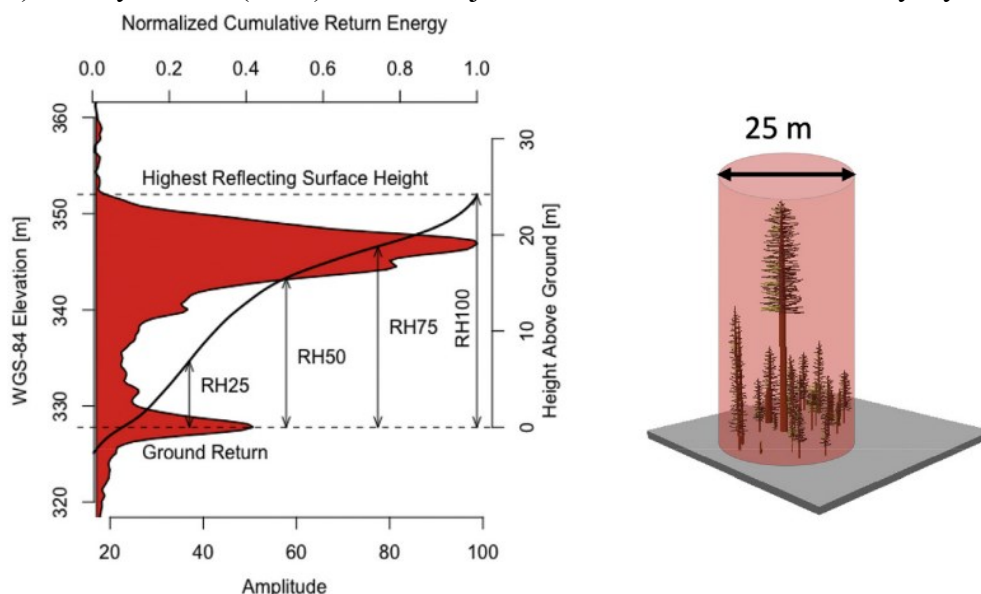
Obrázek 4: Vlevo – GEDI orbity dohromady za dva týdny, každý obsahuje 8 paprsků. Vpravo – simulované rozložení pozemních drah GEDI po dvou letech (na rovníku, kde bylo 50 % drah odstraněno jako simulace vlivu mraků). Nepravidelnost pozemních drah souvisí s nepravidelností dráhy letu ISS. Zdroj: Dubayah et al., 2020

Každý orbit má již popisovaný vzor 8 pozemních drah s tím, že jednotlivé dráhy mají mezi sebou 10x větší rozestup, než má dopad jednotlivých shotů.



Obrázek 5: Vzor pozemních drah GEDI s parametry rozmístění. Zdroj: Dubayah et al., 2020

Navracený pulz je zaznamenán jako kontinuální návratová vlna, ze které lze odvodit celou řadu vegetačních a vlnových metrik, jako je výška vegetace, vertikální profil vegetace a metriky relativní výšky (RH metriky) (Tang et al., 2012; Drake et al., 2002). Dubayah et al. (2020) dále zmiňuje možnost extrakce nadmořské výšky terénu.



Obrázek 6: Návratová vlna GEDI v NIR spektru, se odrazí od vegetace v oblasti kruhu o průměru ~25 m (vpravo) a následně je z křivky odvozena výška terénu, vegetace (korun stromů) a metriky relativní výšky (RH metriky). Zdroj: Dubayah et al., 2020

Z metrik získaných z návratové vlny lze pak odvodit mnoho dalších produktů, včetně hustoty vegetace, pokrytí korun (canopy cover) a odhadu nadzemní biomasy (Dubayah et al., 2020). Zásadními metrikami nesoucími informaci o struktuře vegetace jsou RH metriky (relative height metriky). Návratová vlna obsahuje záznam odraženého množství energie z vyslaného laserového pulzu v jednotlivých nadmořských výškách vzhledem k elipsoidu WGS84. Množství této energie se dá také vyjádřit kumulativní distribuční funkcí (CDF). Pomocí této funkce, podle podílu navracené energie, pak RH metriky označují výšku v jednotlivých percentilech RH0-RH100 od nejnižšího určeného modu (špičky) v návratovém signálu, který obvykle označuje terén (viz Obrázek 6). Hodnoty RH tedy popisují, v jaké výšce nad terénem se nachází jednotlivá procenta podílu celkové biomasy v dané stopě (footprintu) měření. RH75 pak například znamená výšku vegetace (v m), ve které se nachází 75 % její dřevní hmoty. Celkovou výšku vegetace (vrcholek korun) by měla teoreticky obsahovat hodnota RH100, ale první část návratové vlny bývá často zatížena značným šumem např. vlivem atmosféry. Řada studií tak považuje RH100 za nestabilní metriku, jelikož výšku vegetace v mnoha případech příliš nadhodnocuje a navrhuje místo ní používat hodnoty nižších percentilů RH98 nebo RH95 (Duncanson et al., 2022; Lang

et al., 2022a; Potapov et al., 2021). Pokud se naopak příliš velká část energie odrazila od terénu, může dojít k značnému podhodnocení spodních RH metrik (RH0-RH10), které mohou obsahovat dokonce i záporné hodnoty (Beck et al., 2021; Spracklen, Spracklen, 2021).

GEDI produkty

Data z mise GEDI jsou dostupná od dubna 2019, kdy provedena poslední zkouška na orbitu. Vědecký tým GEDI je zavázán dodávat data do datových center Land Process (LP) a Oak Ridge National Laboratory (ORNL) Data Archive and Analysis Centers (DAACs), které fungují v rámci systému EOSDIS (NASA Earth Observing System Data and Information System) (Dubayah et al., 2020). Datové soubory z GEDI jsou pak zpracovány a distribuovány v několika úrovních (Levels), od surových geolokovaných dat (Level 1), po data spolu s metrikami získanými podrobnou analýzou a zpracováním návratových vln. Pro každou stopu (datový bod) GEDI je tak dostupné velké množství atributů (Fayad et al., 2021b).

Tabulka 1: Přehled datových produktů mise GEDI. Zdroj: Hofton et al., 2019; Products Overview - GEDI, 2019

Úroveň (Level)	Popis	Rozlišení	Dostupnost
L1A	RAW vlnové průběhy	průměr stopy 25 m	nedostupné
L1B	geolokované vlnové průběhy	průměr stopy 25 m	LP DAAC
L2A	výška terénu, výška vegetace, RH metriky	průměr stopy 25 m	LP DAAC
L2B	Canopy Cover Fraction (CCF), Leaf Area Index (LAI), CCF a LAI profily	průměr stopy 25 m	LP DAAC
L3	rasterizované metriky z úrovně Level 2	1 km grid (rastr)	ORNL DAAC
L4A	úroveň nadzemní biomasy	průměr stopy 25 m	ORNL DAAC
L4B	rastr hustoty nadzemní biomasy (AGBD)	1 km grid (rastr)	ORNL DAAC

Všechny produkty dostupné v jednom z datových center DAAC jsou zároveň veřejně dostupné ke stažení (po přihlášení) na stahovacím portále NASA Earth Data Search (<https://search.earthdata.nasa.gov/search>). Data jsou uložena v hierarchickém formátu HDF5 (Hierarchical Data Format version 5) s příponou .h5 a distribuována po jednotlivých orbitech v souřadnicovém systému WGS84 (EPSG:4326). Soubor obsahuje metadata o daném orbitu včetně dat z jednotlivých osmi paprsků ve formě vrstev (Beck et al., 2021).

Účelem produktu GEDI Level 2A (GEDI02_A) je poskytnout interpretaci průběhu návratové vlny poskytnutím z produktu GEDI Level 1B (GEDI01_B) odvozených metrik, které zahrnují nadmořskou výšku terénu, výšku vegetace, RH metriky a další vrstvy, kterých je dohromady 156 pro každý z osmi paprsků. Pro pochopení, jakou informaci nesou jednotlivé atributy (vrstvy) v produktu je k dispozici seznam atributů „GEDI L2A Product Data Dictionary“ v GEDI dokumentaci na stránkách produktu GEDI Level 2A (GEDI02_A) v002 (https://lpdaac.usgs.gov/products/gedi02_av002/). Metodika generování produktu je převzata z algoritmu Land, Vegetation, and Ice Sensor (LVIS) (LP DAAC, 2021).

V produktech L1 a L2 jsou dále používány různé algoritmy zpracování návratové vlny a geolokace stopy. Na obě problematiky existuje podrobná dokumentace, která zahrnuje fyzikální teorie, matematické postupy a modelové předpoklady, které se používají při vytváření těchto produktů (Hofton et al., 2019; Luthcke et al., 2019; Products Overview - GEDI, 2019). Tyto dokumenty se nazývají Algorithm Theoretical Basis Documents (ATBD). V jednom z nich Hofton et al. (2019) například uvádí podrobný popis šesti různých algoritmů pro odvození RH metrik a jejich parametry. RH metriky v produktu L2A jsou pak k dispozici ve všech šesti algoritmech zpracování s jedním označeným jako preferovaný pro danou lokaci. Z hlediska správné geolokace středu stop, první verze produktů GEDI Version 1 (V001) nesplnila své původní očekávání. Průměrná chyba horizontální geolokace byla 20,9 metru s 95% týdnů mise, kdy byla chyba geolokace 1-sigma menší než 25,3 m. Původním očekávaná průměrná chyba přitom byla pouze ~10 m. V druhé verzi produktů GEDI Version 2 (V002) jsou však geolokace stop značně zpřesněny a jejich průměrná horizontální chyba geolokace 1-sigma je pouze 10,3 metrů, přičemž 95 % týdnů mise má chybu geolokace 1-sigma menší než 11,9 metru (Beck et al., 2021).

Filtrace kvalitních měření

Pokud je oblast stopy (kam paprsek dopadne) pokryta hustou vegetací, která korunami stromů pokrývá 95 % plochy a více, nemá šanci žádný ze čtyř slabších paprsků (coverage beams) proniknout skrz až na terén. V tomto případě je měření chybné, protože bez kvalitního určení nadmořské výšky terénu není možné určit ani RH metriky. Silnější paprsky (power beams) mohou proniknout vegetací s maximálně

98% pokrytím prostoru korunami stromů. Určitou možností filtrace by tak bylo v místech s hustější vegetací vybrat pouze silnější paprsky (Beck et al., 2021).

Beck et al. (2021) v uživatelském manuálu GEDI (GEDI Level 2 User Guide) dále popisuje jednotlivé atributy označující kvalitu každého měření podle kvality návratového signálu, geolokace nebo zaměření terénu za různých podmínek při dopadu paprsku. Základním atributem uvnitř produktů L2, komplexně posuzujícím kvalitu měření, je `quality_flag`, který v sobě nese binární hodnotu. Pokud je `quality_flag=1`, měření je tak označeno za kvalitní, po uvážení energie návratu, sensitivity při průchodu vegetací, odchylky měřené výšky terénu vůči digitálnímu výškovému modelu (DEM) a dalších kritérií. Jedná se o nejjednodušší způsob filtrace, který bývá v mnoha studiích proveden na začátku práce s daty GEDI (Huettermann et al., 2022; Milenković et al., 2022).

Atribut sensitivity, který obsahuje hodnoty od 0 do 1, je také součástí produktů L2. Označuje schopnost signálu vyslaného pulzu proniknout korunami stromů a dostat se tak až na terén. Aby byl `quality_flag=1` musí být sensitivity rovna 0,90 a vyšší (Beck et al., 2021). Studie Huettermann et al. (2022) například zjišťuje, že se koeficient determinace R^2 s referenčními daty z ALS zvyšuje s vyšší sensitivity. Data s větší sensitivity průniku paprsku skrze vegetaci tak podle studie vedou k přesnějšímu odhadu výšky vegetace, jelikož mají silnější lineární vztah s daty ALS. Adam et al. (2020) však předtím ve své studii žádnou přímou souvislost mezi mírou sensitivity a zkreslením výšky vegetace nenalézají, což naznačuje, že i když existují určité trendy růstu přesnosti učení výšky vegetace s vyšší sensitivity signálu, nejsou tyto vztahy přímočaré a univerzálně platné pro všechny typy vegetace.

Při ověřování kvality dat, například ověřováním vztahu s referenčním měřením výšky terénu či vegetace z ALS, je také nutné brát v úvahu chyby geolokace dat GEDI. Faktory zhoršení kvality geolokace, jako například vychýlená trajektorie letu, jsou uloženy v atributu `degrade_flag` pod různými kódy. Nenulová hodnota `degrade_flag` tak značí vždy ovlivněnou nebo chybnou geolokaci. Lang et al. (2022b) ve své studii zmiňuje, že při nenulovém `degrade_flag` může být u GEDI V001 poloha dopadu paprsku posunuta až o 60 m mimo skutečné místo měření. Beck et al. (2021) dále například zmiňují negativní vliv slunečního šumu a radí používat jen data, která byla pořízena v noci (podle atributu `solar_elevation < 0`). Další studie však prokázaly, že je tento vliv na přesnost měření zanedbatelný (Liu et al., 2021; Potapov et al., 2021).

2.2.2 ICESat-2

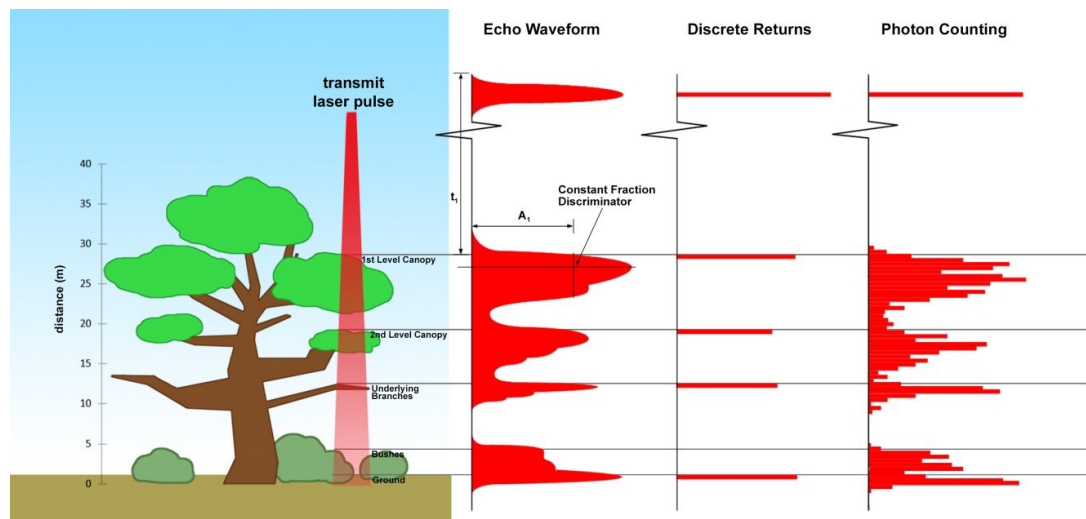
ICESat-2 (Ice, Cloud, and land Elevation Satellite 2) je satelitní mise, při které NASA 15. září 2018 vypustila na oběžnou dráhu nový satelitní LiDAR, jako nástupce původního satelitu ICESat (Markus et al., 2017). Mise ICESat byla přitom první misí využívající satelitní laserovou altimetrii pro účely environmentálních věd (Schutz et al., 2005). Podle průvodních studií projektu (Neumann et al., 2019; Markus et al., 2017) má satelit ICESat-2 stanoveny následující čtyři vědecké cíle:

- Kvantifikovat podíl polárního ledového příkrovu na současných a nedávných změnách mořské hladiny a jeho souvislosti s klimatickými podmínkami
- Kvantifikovat regionální příznaky změn ledového příkrovu s cílem posoudit mechanismy, které tyto změny způsobují, a zlepšit jejich prediktivní modely
- Odhadnout tloušťku mořského ledu a prozkoumat výměnu energie, hmoty a vlhkosti mezi ledem, oceánem a atmosférou
- Měřit výšku vegetace jako podklad pro odhad velkoplošné biomasy a její změny

Díky svému orbitu dokáže ICESat-2 pokrýt geografické území od 88° s. š. po 88° j. š. a pohybuje se zhruba ve výšce 500 km. Jednou za 91 dní provede měření na identickém dráze, což je dobré pro pozorování změn v ledových příkrovech nebo v množství biomasy. Nad oceány a polárními regiony je navíc frekvence opakovaného měření ještě vyšší a nová měření se provádí jednou za měsíc. Životnost satelitu ICESat-2 byla navržena na 3 roky, proto některé z požadavků a cílů mise musely být omezeny ve prospěch delší životnosti (Markus et al., 2017).

ICESat-2 dokáže měřit tloušťku mořského ledu, výšku oceánů, včetně výšky vnitrozemských vodních ploch, globální výšku vegetace, výšku různých typů oblaků, vrstev či útvarů v atmosféře a další. Ke všem těmto měřením satelit využívá svého jediného přístroje Advanced Topographic Laser Altimeter System (ATLAS), jehož způsob měření se liší od klasického analogového laserového altimetru, který byl například na předchozím satelitu ICESat nebo GEDI. ATLAS je tzv. mikropulzní laser s vysokou pulzní frekvencí (PRF) 10 kHz, která sice zapříčiní, že bude energie každého pulzu malá, ale zároveň se díky tomu vytvoří velmi hustá síť měření s malým odstupem ~0,7 m. Minimální mezery mezi měřeními jsou vhodné pro věrnou reprezentaci topografie u heterogenních nebo drsných povrchů, jako jsou ledovce nebo mořské hladiny. Kvůli velmi malé energii každého pulzu je potřeba, aby byla detekce zpětného odrazu vysoce citlivá. ICESat-2 zpětně odražené pulzy detekuje za pomoci

fotonásobičů (PMT) na úrovni jednotlivých fotonů, které jsou následně geolokovány. Tento způsob se také výrazně liší od zaznamenávání celého průběhu návratové vlny, jako odraženého signálu, který se používá v satelitních misích GEDI nebo ICESat (Markus et al., 2017). Odražené fotony přijaté přístrojem ATLAS však ve svém vertikálním rozložení, pokud jsou nakumulované ve velkém množství (stovky a více), také dokáží reprezentovat návratovou vlnu (Neuenschwander et al., 2022).



Obrázek 7: Přehled způsobů LiDAR detekce. Zdroj: Neuenschwander et al., 2022

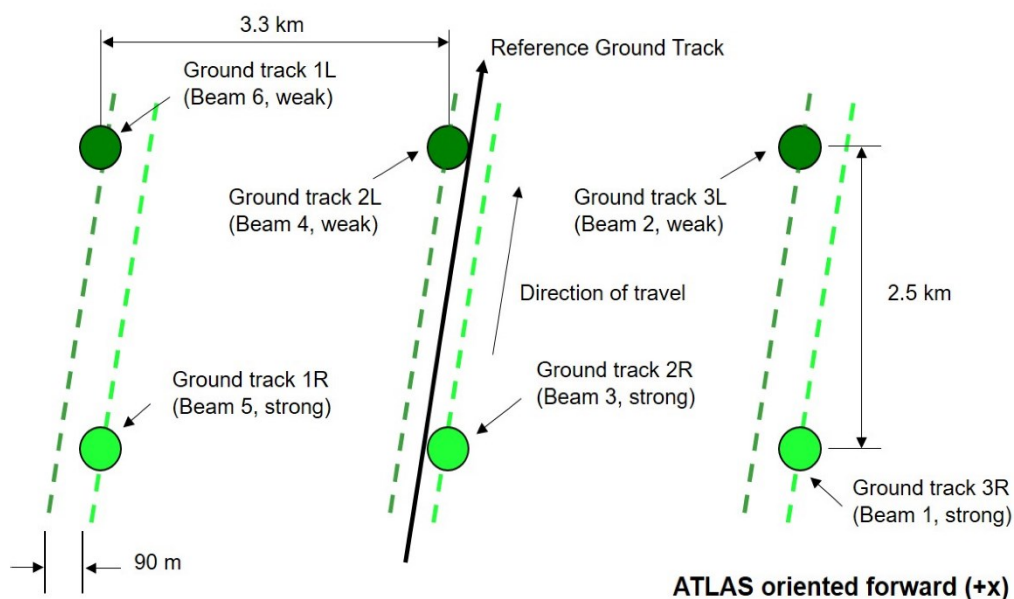
ICESat-2 může vysokou hustotou měření v dráze, po odrazu pulzů s nízkou energií, uchovávat podobný záznam vertikální struktury jako GEDI, který ji získá po odrazu pulzu s vyšší energií v rámci jedné stopy. Zatímco ICESat-2 má tedy v jednotlivých dráhách přesnější horizontální rozlišení, protože je více zaměřen na pozorování kryosférických procesů, GEDI záznamem kontinuálního průběhu návratové vlny dosahuje většího vertikálního rozlišení, které umožňuje podrobnější studium struktury vegetace. Monitorování uhlíkových zásob v lesích je však stále v rámci mise ICESat-2 jedním z požadavků, jehož součástí je také možnost využití měření ICESat-2, jako doplněk k datům z jiných probíhajících misí mapování biomasy a vegetace. Markus et al. (2017) přímo zmiňují synergii použití dat ICESat-2 s daty GEDI nebo NASA-ISRO SAR Mission (NISAR).

Podle článků (Neumann et al., 2019; Markus et al., 2017) lze dále klíčové specifikace ICESat-2/ATLAS shrnout v jednotlivých bodech:

- Vlnová délka: 532 nm (~zelená)
- Rychlost pulzu (PRF): 10 kHz
- Velikost stopy (footprint): <17,5 m (~13 m)
- Vzdálenost mezi stopami (shots): 0,7 m
- 3 páry paprsků (3x strong beam, 3x weak beam)
- 6 pozemních drah (ground tracks)
- Poměr energie mezi silnými a slabými paprsky: 4:1
- Vzdálenost mezi páry paprsků: 3,3 km
- Vzdálenost drah uvnitř párů: 90 m

Na jeden pulz o vlnové délce 532 nm připadne zhruba 20 trilionů fotonů (Ranndal et al., 2021). Každý z pulzů se po dopadu na zem odrazí od plochy o průměru menším než 17,5 m (někdy uváděná velikost stopy cca 13 m). Při již zmiňované pulzní frekvenci 10 kHz se mezi stopami vytvoří malé mezery 0,7 m, při kterých se však stopy budou překrývat. Z jednoho paprsku se vytvoří šest paprsků, pomocí difrakčního optického elementu (DOE). Každý z paprsků může mít díky tomu specifické parametry. Z hlediska celkové dostupné energie bylo možné vytvořit jen 3 silné paprsky, které doplňují 3 slabé paprsky při poměru energie 4:1. Tyto paprsky na zemi tvoří 6 pozemních drah, kdy je uvnitř každého páru drah (slabého/silného paprsku) mezera 90 m a vzdálenost mezi dvěma páry je 3,3 km (Neumann et al., 2019). Podle Markus et al. (2017) byla konfigurace silných/slabých paprsků ATLAS navržena tak, aby se zvýšil radiometrický dynamický rozsah k dosažení primárních vědeckých cílů. Neumann et al. (2019) doplňuje, že se od silných i slabých paprsků očekává, že budou poskytovat dostatečný poměr signálu k šumu pro výšková měření nad povrchy s vyšší odrazivostí, jako je mořský led nebo ledové příkrovy, zatímco silné paprsky budou primárním prostředkem pro zaměřování objektů s nízkou odrazivostí, jako jsou oceány, nebo ve většině případů i vegetace.

Přibližně dvakrát ročně dojde ke změně orientace dráhy letu ICESat-2. Pokud je ATLAS orientován dopředu, slabé paprsky jsou v páru paprsků na levé straně a případnou jim pozemní dráhy 1L, 2L a 3L. Zároveň jsou k sobě slabé a silné paprsky skloněny tak, že pokud je ATLAS orientován dopředu, slabé paprsky jsou vysílány o zhruba 2,5 km před silnými. Pokud je orientace přístroje ATLAS změněna, na levé straně páru budou naopak silné paprsky, které budou zároveň o ~2,5 km napřed před slabými paprsky (Neumann et al., 2019).



Obrázek 8: Schéma rozložení paprsků ICESat-2 při dopředné orientaci přístroje ATLAS. Zdroj: Zhang et al., 2021

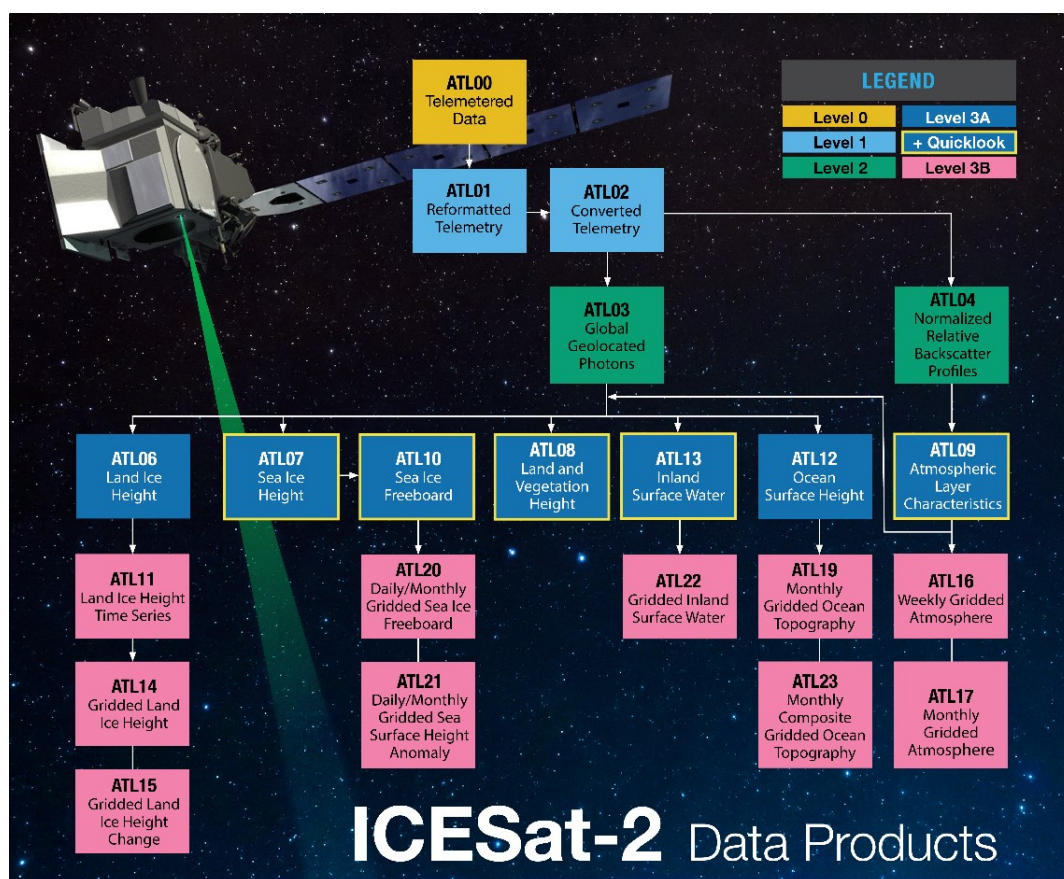
Fotony samy o sobě neposkytují přímou informaci o výšce objektu, od kterého jsou odraženy, a do teleskopu se mohou odrážet i jiné fotony (např. ze Slunce). Nutné je proto přijaté fotony akumulovat a neustále statisticky analyzovat. Od země odražené fotony mají mnohem větší hustotu a specifické statistické parametry, kterými je možné je odlišit od šumu. Skutečná přesnost měřené výšky pomocí ATLAS závisí na poměru signálu k šumu (SNR), na délce nebo vzdálenosti, na kterou jsou laserové pulzy zaznamenány, a na přesnosti, s jakou lze každý návrat načasovat (Markus et al., 2017).

ICESat-2 produkty

V Goddardově kosmickém středisku NASA je oddělení Instrument Support Facility (ISF), které se stará o monitorování, řízení a kontrolu stavu přístroje ATLAS. Pro tvorbu a distribuci datových produktů mise ICESat-2 je využíván systém SIPS (Science Investigator-led Processing system). Ze systému SIPS jsou datové produkty zasílány do datového centra NASA Distributed Active Archive Center (DAAC) umístěného v National Snow and Ice Data Center (NSIDC) ve městě Boulder, Colorado. Odtud jsou data distribuována veřejnosti v různých úrovních zpracování (Markus et al., 2017). V hierarchickém datovém formátu HDF5 nebo CSV jsou dostupná buď přímo z NSIDC (<https://nsidc.org/data/icesat-2/>) nebo na portálu NASA Earth Data Search (<https://search.earthdata.nasa.gov/search>) či NASA OpenAltimetry (Khalsa et al., 2020).

Datové produkty z mise ICESat-2 sahají od surových telemetrických dat z přístroje ATLAS ve formě paketů (ATL00) po například produkt ATL03, který

obsahuje jednotlivé geolokované fotony v souřadnicovém systému WGS84 s výškou vztaženou k IRTF2014. ATL03 má fotony organizovány do ~20 m dlouhých segmentů podél pozemních drah a je zásadním produktem pro odvození, již tematicky specifických, produktů úrovně Level 3A (Markus et al., 2017). Jedním z odvozených produktů je například produkt s označením ATL08, který obsahuje výšku terénu a vegetace v 100 m dlouhých segmentech pro každý paprsek podél pozemních drah, které jsou složeny z pěti segmentů ATL03. Datové body v ATL08 představují centroidy segmentů, tedy obdélníků o rozměrech ~13x100 m orientovaných ve směru příslušné dráhy. Výška vegetace je pro každý z těchto segmentů v ATL08 uložena např. v atributu `h_canopy`, kde označuje 98. percentil relativní výšky vegetace nad terénem v metrech. Pro získání těchto atributů z fotonů se využívá v produktu ATL08 sofistikovaných postupů filtrace šumu a algoritmů popsanych v příslušné dokumentu ATBD pro ATL08 (Neuenschwander et al., 2022).



Obrázek 9: Přehled produktů mise ICESat-2 v jednotlivých úrovních zpracování. Zdroj: NASA, NSIDC DAAC

Filtrace kvalitních měření

Předpokladem pro použití dat ATL03 ke zpracování a tvorbě produktu ATL08 jsou vhodné podmínky na orbitu, kvalita fotonů a geolokace. Požita jsou například

data, která mají atribut podmínek při sběru dat $POD_PPD = 0$ a atribut kvality fotonů $quality_ph = 0$. Hodnota 0 v obou případech označuje normální podmínky a nepřítomnost chyb. Základní filtrace v produktu ATL08 pak může být založená čistě na odstranění dat příliš vysoko nad terénem nebo pokud bylo pro výpočet dat o výšce v daném segmentu použito příliš málo fotonů. Atribut $canopy_rh_conf$ například indikuje spolehlivost výškových údajů o vegetaci na základě procenta pozemních a vegetačních fotonů v segmentu. Hodnota $canopy_rh_conf$ je nastavena na 0, pokud je počet fotonů vegetace menší než 5% celkového počtu klasifikovaných fotonů v segmentu. Dále například atribut $h_canopy_uncertainty$ vyjadřuje nejistotu výšky vegetace, která zahrnuje všechny systematické nejistoty (např. časování, podmínky na orbitech, geolokace atd.) a nejistotu identifikace fotonů (Neuenschwander et al., 2022).

2.3 Modely výšky vegetace z dat satelitní laserové altimetrie

V této kapitole se budeme zabývat existujícími modely výšky vegetace (CHM), vytvořenými především z dat GEDI. V některých případech byla data kombinována s produkty mise ICESat-2. Modely, nejčastěji vycházející ze strojového učení, využívaly také velmi často data z radarových či optických snímků. Od roku 2018, kdy byla mise GEDI zahájena, vzniklo z jejích produktů několik modelů výšky vegetace. Nejznámějšími z nich jsou globální CHM zmiňované v kapitole 2.1.

Potapov et. al. (2021)

Jako první byl publikován článek (Potapov et al., 2021), který pojednává o tvorbě globálního CHM z dat GEDI v kombinaci se snímky Landsat s výsledným rozlišením 30 m pro rok 2019. Rozlišení celého modelu přitom vychází z 30 m rozlišení snímků Landsat. Pokrytí dat z mise GEDI umožnilo natrénovat model a provést prostorovou interpolaci jen mezi 52° s. š. a 52° j. š, přičemž v boreálních oblastech (taiga) bylo využito extrapolace. Potapov et al. (2021) dále o výběru dat uvádí, že tým Global Land Analysis and Discover na University of Maryland (UMD GLAD) integroval v té době dostupná data GEDI v001, od dubna do října 2019, s časovou řadou optických snímků Landsat pro celý rok 2019. Cílem bylo vytvořit co nejuniverzálnější model strojového učení tak, aby ho bylo možné aplikovat i na Landsat data z jiných let. To pak by umožnilo sledování globálního vývoje struktury vegetace, což by bylo klíčovým přínosem pro ochranu životního prostředí a klimatický výzkum. Studie přímo zdůrazňuje důležitost důsledného a rozsáhlého monitorování

výšky lesů pro odhad emisí uhlíku souvisejících s lesy, analýzu odlesňování a úbytku lesů včetně kvantifikace účinnosti iniciativ na obnovu lesů (Potapov et al., 2021).

V globálním CHM (Potapov et al., 2021) byla z dat GEDI v001 využita konkrétní metrika RH95, jelikož vykazovala největší korelaci s validačními daty z leteckého laserového skenování (ALS), konkrétně s 90 % percentilem z ALS CHM pro rozlišení 30 m. Uvedeno je, že má GEDI tendenci výšku vegetace nadhodnocovat zejména na svazích v horských travnatých oblastech mírného a subtropického klimatického pásu. Za příklad je uvedeno velké nadhodnocení na Novém Zélandu či v Lesothu. Při filtraci byla zachována pouze data GEDI, která pocházela ze silného paprsku (power beam), nasnímaná byla během noci (solar_elevation < 0), jejich sensitivita byla ≥ 0.9 a rozdíl mezi RH metrikami u 6 algoritmů zpracování nebyl větší než ≤ 2 m. Pro oblasti tajgy a mírného pásu byly také použity data pouze ve vegetačním období. Pro Landsat byly využity 16denní časové řady z GLAD Landsat Analysis Ready Data (ARD) (Potapov et al., 2020) pro rok 2019, ze kterých bylo odvozeno dohromady 546 různých metrik v $1^\circ \times 1^\circ$ dlaždicích. Mezi metrikami byly minimální, maximální a mediánové hodnoty odrazivosti pro spektrální kanály (B, G, R, NIR, SWIR1, SWIR2) včetně vegetačních metrik spočítaných pro start, vrchol a konec vegetační sezóny (především NDVI) a také bylo například využito výškových dat SRTM. Všechna data z byla využita bezmračná a pro některé byla využita lineární regrese pro vyplnění prázdných hodnot. Před modelováním byla data rozdělena do $1^\circ \times 1^\circ$ dlaždic, kde 90 % tvořili trénovací data a 10 % validační. Při modelování bylo využito lokálně kalibrovaného modelu strojového učení nazvaného „bagged regression tree ensemble“, což prakticky znamená soubor rozhodovacích stromů v pohyblivém okně. Porovnání modelu s validačními daty GEDI ukázalo chyby RMSE (Root Mean Square Error) = 6,6 m, MAE (Mean Absolute Error) = 4,45 m, a koeficient determinace $R^2 = 0,62$. Pomocí předchozí ALS CHM v 30 m rozlišení pak byla odhadnuta vyšší nejistota modelu RMSE = 9,07 m; MAE = 6,36 m, $R^2 = 0,61$. V urbánních a suburbánních oblastech může být zároveň výška vegetace zaměněna s výškou zástavby a velký vliv na nejistotě v modelu měla také velká chyba geolokace u dat GEDI verze 1 nebo kvalita učení nadmořské výšky terénu (Potapov et al., 2020)

Data výšky vegetace z tohoto modelu jsou nyní veřejně dostupná jako Global Forest Canopy Height, 2019 na adrese <https://glad.umd.edu/dataset/gedi>. Data lze stáhnout ve formátu GeoTiff v souřadnicovém systému WGS84.

Lang et. al. (2022)

Další dostupný globální CHM (Lang et al., 2022a) si klade za cíl monitorovat globální koloběh uhlíku, podobně jako model (Potapov et al., 2021). Výhoda tohoto modelu je, že má rozlišení 10 m, protože využívá dat vyššího rozlišení ze satelitu Sentinel-2. Původní studie z roku 2022 využívala data GEDI pro období dubna 2019 až srpna 2020 spolu se snímky Sentinel-2 L2A pro rok 2020. V roce 2023 se stejný tým zabýval aktualizací modelu a vydána byla rozšiřující studie (Lang et al., 2023). Provedlo se vylepšení techniky strojového učení s využitím novějších dat GEDI L2A (duben 2019 až březen 2023). Pro tvorbu globálního CHM byla použita metrika RH98, která nejlépe korelovala s výškami stromů z ALS a podle studie nebyla tolik zatížena geolokační chybou. Při filtraci byly zachovány pouze měření na nevegetačních plochách, tj. s 70% pravděpodobností výskytu vegetace. Podle studie může zástavba ve městech způsobit velké zvýšení RH98, což by model zatížilo chybou. Dále byl použit nejzákladnější filtr pro zachování pouze kvalitních měření (`quality_flag=1`) a odstraněny negativní a nulové RH98. Při modelování pomocí konvolučních neuronových sítí (CNN) bylo dále využito náhodné křížové validace, aby bylo odstraněno 30% měření s největší neurčitostí, čímž se vylepšila přesnost modelu. Data ze Sentinel-2 L2A byla využita bezmračná a konkrétně všechny spektrální pásma, která byla sjednocena na rozlišení 10 m kubickou interpolací. Veškerá data byla následně rasterizována, rozdělena na dlaždice 100 x 100 km. Z těchto dlaždic bylo 80% dat vyhrazeno na trénink modelu a zbylých 20% na validaci. (Lang et al., 2023; Lang et al., 2022a).

Výsledky validace dat GEDI z modelu v první studii (Lang et al., 2022a) vyšli s $RMSE = 6,0$ m a $ME = 1,3$ m. U výsledků se hodně řešil problém tzv. saturace, kdy model nedokáže správně odhadovat vyšší výšky stromů, protože je celosvětově 95 % procent stromů nižších než 30 m. Vyšší stromy jsou pak ve výsledku velmi podhodnoceny nebo nezachyceny. Dále se uvádí, že to byl jeden z dalších důvodů pro výběr metriky RH98, jelikož představuje vyšší výšky než např. RH95. Cílem modelu CNN bylo tedy přesněji odhadovat vysoké stromy, které mohou ukládat větší množství uhlíku. Podle studie tedy chyba $ME = 1,3$ m znamená nadhodnocení nižší vegetace, která je průměrem. Model byl zároveň hodnocen pro jednotlivé výškové intervaly zvlášť, aby bylo zjištěno, že pro vzácnější vysoké stromy bylo po optimalizacích modelu docíleno minimálního nadhodnocení $ME = -1,8$ m a $RMSE = 7,3$ m. Validace

byla dále provedena pomocí dat NASA Land, Vegetation, and Ice Sensor (LVIS). Jedná se o data leteckého LiDARu (ALS), která poskytují systematické pokrytí, zahrnující několik velkých regionů na světě. Při porovnání s těmito daty model CNN dosáhl větší chyby $RMSE = 7,8$ m a $ME = 0,6$ m. Dále se uvádí zvlášť chyby pro oblasti mimo dosah GEDI, $RMSE = 4,7$ m a $ME = 1,6$ m a pro oblasti v dosahu dat GEDI, $RMSE = 8,8$ m a $ME = 0,2$ m. (Lang et al., 2022a).

V aktualizované studii z roku 2023 porovnává Lang et al. (2023) výsledky validace CNN modelu pomocí dat LVIS s modelem UMD (Potapov et al., 2021). Globální CHM (UMD; Potapov et al., 2021) mělo chyby v porovnání s LVIS, $RMSE = 9,1$ m a $ME = 1,7$ m. CNN model (Lang et al., 2023) měl přitom pouze $RMSE = 7,9$ m a $ME = 1,7$ m. V této studii se tak mělo docílit výrazného zlepšení přesnosti. Původní i aktualizovaná globální CHM byla zároveň zpřístupněna veřejnosti jako ETH Global Canopy Height 2020 (<https://langnico.github.io/globalcanopyheight/>) pod licencí CC BY 4.0 v systému WGS84 (Lang et al., 2023; Lang et al., 2022a). K tomuto projektu stejný tým publikoval mnoho studií. Například již v roce 2019 se zabývali modelováním CHM na základě CNN pomocí dat z ALS v kombinaci s daty ze Sentinelu-2 ve Švýcarsku a v Gabonu. Pro Švýcarsko jim vyšla chyba $RMSE = 3,4$ m a $MAE = 1,7$ m a pro Gabon $RMSE = 5,6$ m a $MAE = 4,3$ m (Lang et al., 2019).

Z dat GEDI vzniklo také několik lokálních CHM, mezi které patří především CHM s vysokým rozlišením z roku 2022 (Schwartz et al., 2022) pro lesní oblasti Landes ve Francii. Tento model využívá dat RH95 z GEDI, radaru se syntetickou aperturou Sentinel-1 (S1) a multispektrálních snímků Sentinel-2 (S2) v kombinaci s hlubokým učením pro vytvoření CHM s rozlišením 10 m (Schwartz et al., 2022). Na druhé straně se například Fayad et al. (2021b) věnuje hodnocení dat RH100 z GEDI pro odhady výšek a objemu plantáží eukalyptu v Brazílii, kde byly pro tyto odhady využity různé lineární i nelineární regresní modely. Objemy byly zároveň odhadovány různými dalšími metrikami, které GEDI poskytuje, proto je tato studie hodnotná pro pochopení využití dat z GEDI a minimalizace chyb jeho měření (Fayad et al., 2021b). Další studie se na lokální úrovni například zabývají kombinací dat GEDI s daty ICESat-2 za použití prediktorů z misí PALSAR a Sentinel (Sothe et al., 2022) nebo tvorbou CHM z dat ICESat-2 v kombinaci s daty ze S1 a S2, včetně dat Landsat-8 (Li et al., 2020). Tvorbou globálních CHM z mise ICESat-1, snímáním návratové vlny podobné misi GEDI, se také věnují studie Lefsky (2010) a Simard et al. (2011).

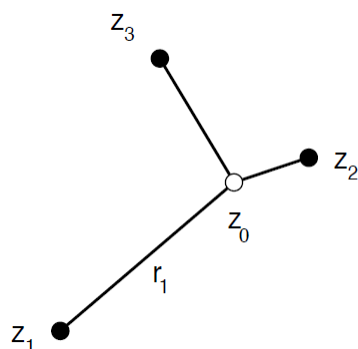
2.4 Metody prostorové interpolace

2.4.1 Prostorová interpolace

Pro tvorbu modelu výšky vegetace jako souvislého povrchu, který bylo nutné vytvořit z řídkých měření GEDI a ICESat-2 v nepravidelné síti bodů, bylo nezbytné využít metody prostorové interpolace. Prostorovou interpolaci lze vymezit jako skupinu metod sloužících k vytváření spojitých povrchů z bodových měření (Dobrovolný, 2005a). „Datová síť většinou není dostatečně hustá, dat je málo, a je nemožné zjistit průběh prostorové veličiny přesně... ..může být proto rozumné hledat nějakou reprezentativní plochu, jakýsi prostorový průměr závislý na souřadnicích“ (Ježek, 2015, s. 13). Prostorové interpolace mohou probíhat i ve více než 2 dimenzích.

Základem některých interpolačních metod jsou matematické postupy založené na: „...prokládání vhodných funkcí nebo na vážených průměrech“ (Ježek, 2015, s. 13). Takovéto metody Dobrovolný (2005a) nazývá deterministickými. Dále uvádí, že tyto metody umožňují data extrapolovat (interpolovat za hranice známých datových bodů), pokud známe fyzikální podstatu jejich jevu. Jin a Heap (2008) v knize uvádí tabulku, kde mezi deterministické metody prostorové interpolace řadí například metody nejbližších sousedů (NN), triangulaci (TIN) nebo metodu vážené inverzní vzdálenosti (anglicky Inverse Distance Weighting; IDW). Kromě deterministických metod existují ještě statistické metody, také známé jako stochastické metody. Jsou to přístupy k prostorové interpolaci, které se opírají o pravděpodobnostní a statistické úvahy a zahrnují koncept náhodnosti. Výsledkem tohoto typu interpolace jsou statistické odhady, které jsou náhodnými veličinami a lze je snadno statisticky popisovat. Určit lze rozdělení pravděpodobnosti, či jiné statistické parametry nebo intervaly spolehlivosti. Díky tomuto snadnějšímu odhadu nejistot je v tomto přístupu možné lépe se zaměřit na minimalizaci chyb (Ježek, 2015; Dobrovolný, 2005a). Jin a Heap (2008) mezi stochastické metody interpolace řadí například regresní modely, techniku rozhodovacích stromů nebo kriging. Řadit se sem mohou i metody strojového učení.

Metoda inverzních vzdáleností (IDW) je podle Ježka (2015) jednou z nejjednodušších a zároveň často používaných deterministických metod prostorové interpolace. Konkrétně o metodě uvádí: „Jedná se o metodu, která je založena na váženém průměru, jehož váhy jsou nepřímo úměrné nějaké mocnině vzdálenosti mezi interpolovaným bodem a datovými body.“ (Ježek, 2015, s. 17). Na IDW lze zároveň snadno vysvětlit několik důležitých konceptů a problémů, které se týkají i jiných metod prostorové interpolace, využívajících vzdálenosti mezi body.

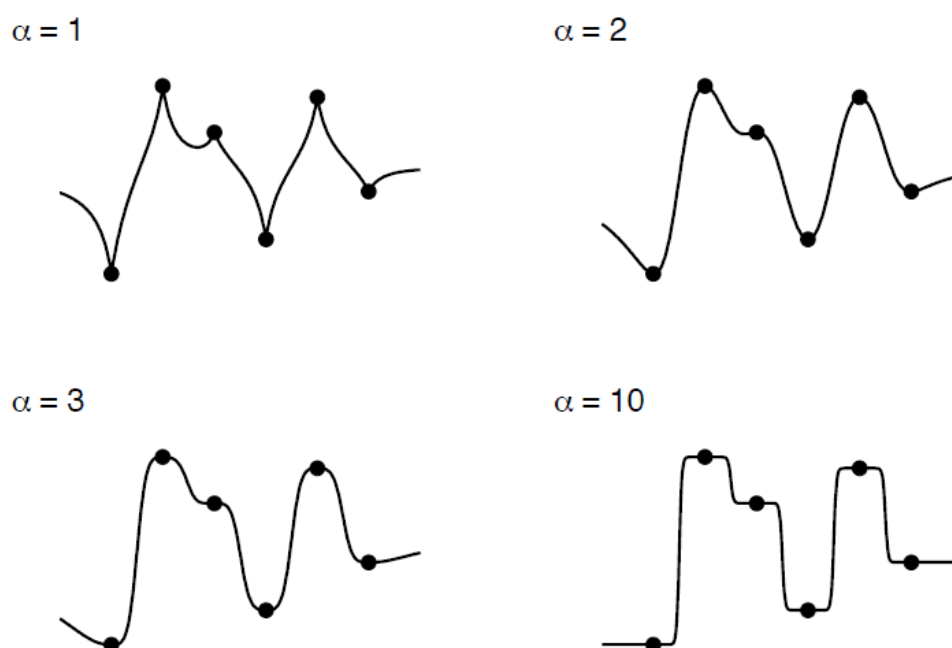


Obrázek 10: Odhad hodnoty bodu z_0 jako váženého průměru z hodnot v jeho okolí z_1, z_2, z_3 . Zdroj: (Ježek, 2015)

Vzorec pro odhad hodnoty z_0 na Obrázek 10 z váženého průměru je tedy:

$z_0^* = \sum_{i=1}^n \lambda_i z_i$. Body měření jsou přitom navzájem různě vzdáleny, což IDW promítá do vah λ_i tím, že jsou voleny nepřímo úměrné této vzdálenosti r_i . Poměr vah v součtu je převráceným poměrem vzdáleností okolních bodů. Součet vah je zároveň roven 1, čímž je zajištěno, že pokud by měly body v okolí naměřené stejné hodnoty, výsledek interpolace bych se od těchto hodnot nelišil. Ze vzorce pro výpočet vah v IDW vyplývá, že pokud je bod, ve kterém provádíme interpolaci, totožný s datovým bodem známé hodnoty, váha daného bodu bude rovna 1 a ostatní váhy v okolí budou rovny 0. Výsledná interpolovaná plocha tak bude vždy přesně procházet známými body, z čehož z definice vyplývá, že metoda IDW je tzv. přesný interpolátor (Ježek, 2015).

U vah metody IDW se zároveň určuje koeficient mocniny α , který udává, jak velký vliv mají měřené body na okolí neznámých (interpolovaných) bodů. Větší mocninou je obvykle zvýšen vliv nejbližších měřených bodů a interpolace může vést až k strmějším přechodům v polovině vzdálenosti mezi body, kdy se pro interpolovaný bod vybere hodnota nejbližšího měření. IDW tak získává spíše charakter metody nejbližšího souseda (NN). Shlazení je možné docílit modifikací vah, přičemž interpolovaná plocha nebo křivka v tomto případě nemusí nutně procházet přesně všemi datovými body (Ježek, 2015).



Obrázek 11: Ukázka vlivu mocniny α na podobě interpolované křivky pomocí IDW (ukázka v 1D).
Zdroj: Ježek, 2015

Metoda IDW je považována za lokální interpolátor, protože využívá k výpočtu neznámé hodnoty pouze měření z předem definovaného okolí interpolovaného bodu. Důležitý vliv na výsledek interpolace má také vliv konfigurace měřených bodů v okolí interpolovaného bodu. IDW je citlivé na shluky nebo odlehle hodnoty těchto bodů, ze kterých navíc zároveň vždy vychází minimální a maximální hodnota pro interpolovanou veličinu, což může být jistou nevýhodou (Dobrovolný, 2005a). Omezením okolního rozsahu zahrnutí bodů do výpočtu vah, se například může docílit odfiltrování odlehlejších hodnot, jelikož mohou být vzdálenější hodnoty zároveň odlišnějšími a nemusí mít na interpolovanou hodnotu takový vliv. Tím, že se omezí výpočet vah jen na několik nejbližších sousedů, se také zmenší výpočetní náročnost algoritmu. Pokud se však okolí omezí příliš a bodů bude málo, mohou v interpolaci naopak vznikat artefakty nebo bude výsledek znehodnocen v případě přítomnosti

velmi odlišné hodnoty v měřených bodech. Dalším důležitým pojmem je anizotropie, která určuje jistou závislost prostorové veličiny nebo výskytu měřených bodů na směru. Pokud hodnoty v měřených bodech nebo jejich poloha vykazují jistou míru anizotropie, může to mít negativní vliv na kvalitu výsledků interpolace. U IDW je výchozím předpokladem naopak izotropie veličiny, a proto má ve 2D prohledávané okolí tvar kružnice (Ježek, 2015). Anizotropie se však dá zohlednit změnou tohoto tvaru okolí tak, že se do výpočtu váh více zahrnou body, které leží ve směru, kde jsou si více podobné. Pokud je například prostorová veličina ovlivněna směrem větru, může se okolí interpolovaného bodu upravit na elipsu, jejíž hlavní osa bude rovnoběžná s tímto směrem. Vychází se z předpokladu, že si budou interpolované veličiny v tomto směru na větší vzdálenosti více podobné než ve směru kolmém (Dobrovolný, 2005a).

Deterministické metody, jako IDW, si tedy zakládají na jednoduchosti parametrizace interpolační funkce, snadném výpočtu a spojitosti výsledného povrchu, což lze využít i v případě, kdy nemáme příliš měření. Pokud však průběh interpolovaného prostorového jevu ovlivňují námi neznámé procesy, může být výhodnější zvolit statistický přístup, který v mnoha případech poskytuje objektivnější odhad a zároveň kvantifikuje s ním spojenou nejistotu. Oproti tomu by například u IDW vznikla otázka, jak zvolit správnou hodnotu koeficientu mocniny α , který zásadně ovlivňuje výsledek interpolace. Ačkoliv by šlo pro zvolené α určit chyby pomocí křížové validace, odhad optimální hodnoty tohoto koeficientu by byl zbytečně náročným procesem. V mnoha případech je proto volba mocniny subjektivní a metoda IDW, na rozdíl od několika statistických metod, nebere v potaz ani vzájemnou konfiguraci měřených bodů nebo chování prostorové veličiny (Ježek, 2015).

2.4.2 Geostatistika a koncept variogramu

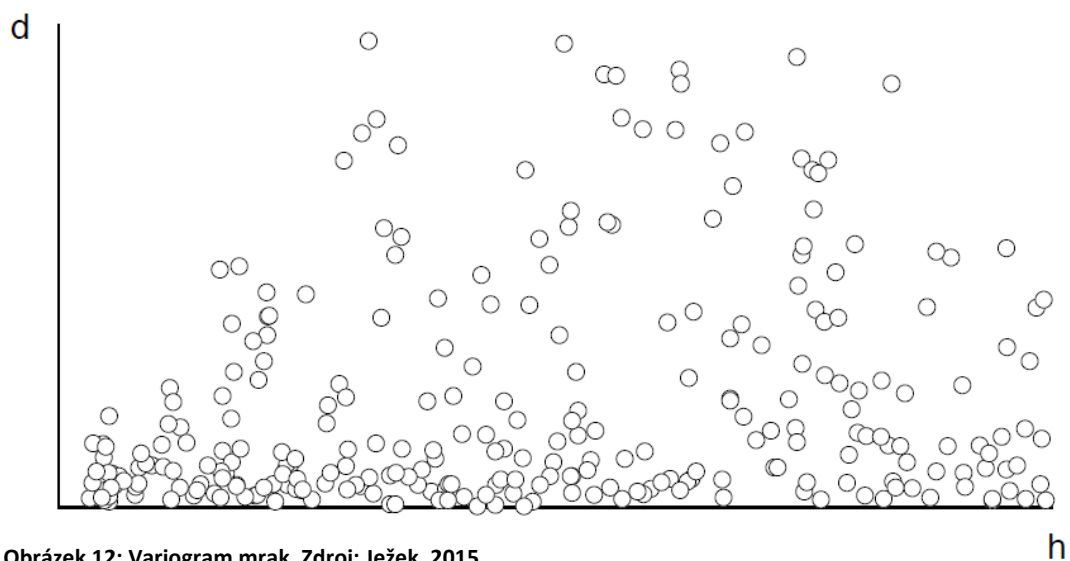
Jelikož je interpolačních metod velké množství, bylo nutné provést volbu vhodné metody, která by nejlépe odpovídala vlastnostem interpolované prostorové veličiny. Na základě předpokladu, že je na území severních Čech hustota měření výšky vegetace z dat GEDI a ICESat-2 vyšší, mohla být vybrána statistická metoda prostorové interpolace, zohledňující prostorovou strukturu měřených bodů.

Většinu takovýchto nástrojů zastřešuje pojem geostatistika. Tento pojem znamená podle Ježka (2015, s. 9) „označení specifické metody zpracování prostorových dat“. Geostatistika je dnes jedna ze základních metod statistiky

prostorových jevů, jelikož se postupem času začala používat ve většině oborech pracujících s prostorovými daty. Její typické aplikace se skládají ze dvou kroků. Prvním krokem je strukturní analýza, která popisuje vlastnosti veličiny pomocí tzv. variogramu. Druhým krokem je pak využití znalostí prostorové struktury z variogramu pro odhad neznámých veličin pomocí geostatistické prostorové interpolace zvané kriging. Variogram a kriging jsou tak oba základními nástroji geostatistiky (Ježek, 2015). Zároveň je v geostatistice klíčovým konceptem tzv. autokorelace, která předpokládá, že hodnoty měřené veličiny jsou si na kratších vzdálenostech více podobné než hodnoty pozorované na delších vzdálenostech (Dobrovolný, 2005a). Prostorový jev je tedy v geostatistice považován spíše za soubor korelovaných náhodných veličin, na které se aplikuje teorie náhodných funkcí (Ježek, 2015). Hlubšímu pochopení těchto statistických konceptů a pravděpodobnostnímu popisu prostorové veličiny, ze kterého vychází další geostatistické nástroje, se věnují kapitoly 6 a 7 v knize Geostatistika a prostorová interpolace od prof. Ježka (2015).

Stručně řečeno se koncept náhodné funkce používá v geostatistice pro popis prostorově kontinuálních jevů, protože se předpokládá, že jejich měření na konkrétních místech mohou být považována za realizaci náhodného procesu. Pro popis této měřené náhodné veličiny se pak nejčastěji využívají pravděpodobnostní rozdělení. Při opakovaném měření vycházejícím z náhodného procesu se dá předpokládat normální rozdělení a vztahy mezi hodnotami bodů jsou potom charakterizovány pomocí prostorové kovariance, která popisuje míru vzájemné závislosti mezi hodnotami měřenými na různých místech v prostoru. Tento geostatistický přístup tak vede k tomu, že má celkový proces mnohonásobné normální rozdělení, což je mimo jiné typickým znakem Gaussových procesů. Pro soubor náhodných měření, které jsou pozorovány v kontinuálním prostorovém jevu, se z tohoto hlediska také někdy uvádí pojem Gaussovo náhodné pole (GRF). Základním parametrem pro simulaci nebo modelování náhodného procesu pomocí Gaussových procesů je kovarianční funkce, která je ekvivalentem již zmiňované prostorové kovariance. Kovariančních funkcí je mnoho typů a používají se v závislosti na charakteru náhodných měření a prostorové struktuře. Jedním z klíčových nástrojů geostatistiky pro analýzu této prostorové struktury je variogram (Moraga, 2023; Ježek, 2015).

Variogram nebo kovariance se obvykle obecně používají pro popis vlastností náhodných měření, jako je jejich prostorová korelace nebo chování v různých směrech (izotropie/anizotropie). Hodnoty bližších bodů měření mají většinou mezi sebou daleko menší rozdíly než hodnoty vzdálené, což je předpoklad autokorelace. Abychom ji mohli ve variogramu zobrazit, musí být na ose X vzdálenost mezi jednotlivými páry bodů a na ose Y rozdíl hodnot těchto bodů. Pro každý pár měřených bodů v prostoru pak vznikne jeden bod, který bude obsahovat informaci o vzájemném rozdílu a vzdálenosti. Takovému grafu se říká variogram mrak, přičemž je rozdílnost mezi hodnotami měřených bodů z vyjádřena mírou rozdílnosti (d), někdy nazývanou také míra nepodobnosti. Nepodobnost (d) je polovinou kvadrátu rozdílu hodnot (z) mezi libovolnými dvěma body (x): $d = \frac{1}{2}[z(x_i) - z(x_j)]^2$. Euklidovská vzdálenost mezi těmito body je ve variogram mraku vyjádřena jako (h) na ose X (Ježek, 2015). Již z následujícího grafu je patrné, že nepodobnost se vzdáleností mezi body vzrůstá.



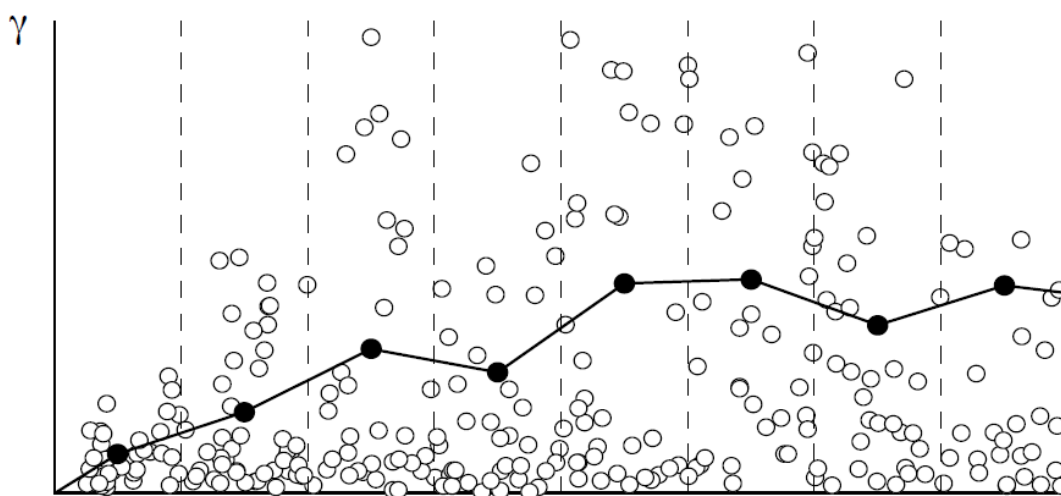
Obrázek 12: Variogram mrak. Zdroj: Ježek, 2015

Pro větší přehlednost vztahu nepodobnosti a vzdálenosti je z variogram mraku odvozen tzv. empirický variogram, který se mnohdy označuje pouze jako variogram. Někdy se uvádí také pojem semivariogram, který vyjadřuje, že se jedná o polovinu kvadrátů rozdílů hodnot, jedná se ale o totéž. Empirický variogram se konstruuje podobně jako je tomu například u histogramu. Data z variogram mraku se rozdělí podle vzdálenosti (h) na distanční třídy (anglicky lags), jejímž středům se vždy přiřadí vypočtená průměrná nepodobnost, která se pro přehlednost ještě mnohdy spojuje lomenou čarou. Někdy se k bodům průměrné nepodobnosti přidává ještě informace o počtu párů, nacházejících se uvnitř intervalu, ze kterých byl průměr vypočítán. Vzorec

pro empirický variogram, kde h je vzdálenost odpovídající středům distančních tříd a $N(h)$ je počet dvojic měřených bodů spadajících do dané třídy, lze zapsat jako:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{|x_i - x_j| \approx h} [z(x_i) - z(x_j)]^2$$

Na následujícím schématu konstrukce empirického variogramu je jasně vidět tendence růstu nepodobnosti se vzdáleností. Další zajímavostí je, že průběžná hodnota nepodobnosti první třídy není nulová. Tato počáteční nepodobnost na velmi krátkých vzdálenostech je nazývána mikrovariabilita a může být buď přirozená (volbou měřítka) nebo způsobená chybou měření (Ježek, 2015).

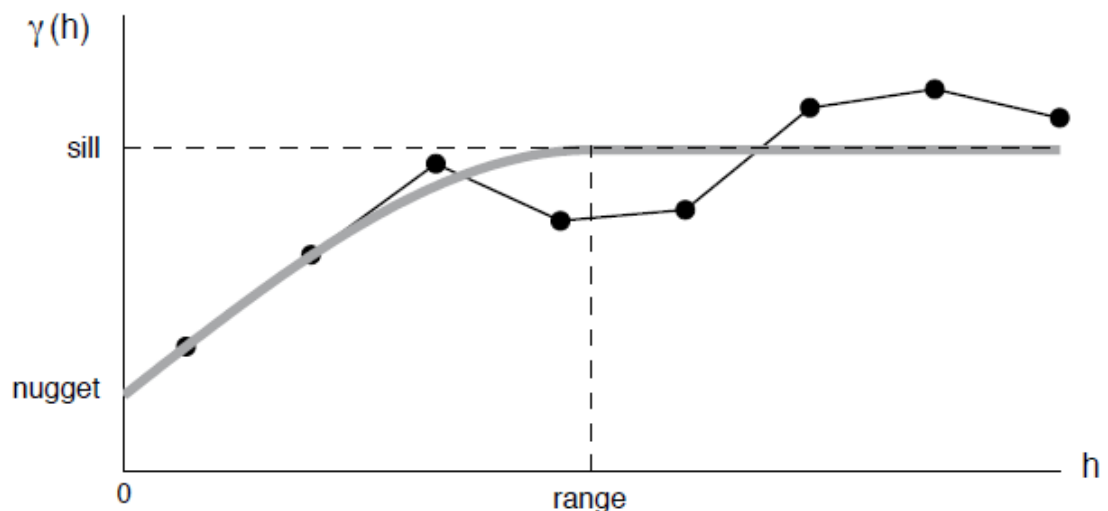


Obrázek 13: Schématické vyjádření výpočtu empirického variogramu. Zdroj: Ježek, 2015

h

Empirický variogram tedy odráží prostorovou autokorelaci v datech tím, že zobrazují nepodobnost jako funkci vzdálenosti. Prostorová struktura lze také popsat například korelogramem, který vyjadřuje, jak se se vzdáleností mění kovariance nebo korelace mezi dvojicemi měření. V obou případech se jedná o míry podobnosti, korelace přitom nezachovává původní rozsah dat. Korelogram má tedy, na rozdíl od empirického variogramu, opačný průběh, kdy podobnost dat se vzdáleností klesá. Jak korelogram, tak variogram, se dají modelovat pomocí kovariančních funkcí, jejichž použití je zásadní pro geostatistické analýzy a predikce. Jedná se například o parametrické modely, jako je Maternova funkce, jejíž parametry se z empirického variogramu odhadují nejčastěji pomocí metody nejmenších čtverců. Modelováním přitom vznikne tzv. teoretický variogram (Moraga, 2023; Ježek, 2015).

Ačkoliv je modelů variogramu mnoho, výsledné teoretické variogramy mají jen několik parametrů. Nepodobnost mezi páry bodů, s jejich vzdáleností od sebe, ve variogramu narůstá. V určitý moment se však růst nepodobnosti zastaví nebo výrazně zpomalí a vzdálenost bodů od sebe přitom přestane mít vliv. Prvním parametrem je tedy vzdálenost, na které se ztrácí závislost nepodobnosti na vzdálenosti, která se nazývá dosah variogramu (anglicky range). Hodnotu variogramu (nepodobnosti) v tomto bodě označuje tzv. práh variogramu (anglicky sill). Znamená to tedy práh míry nepodobnosti (sill), s dosahem vzdálenosti (range), do které lze danou veličinu ještě brát za prostorově korelovanou. Třetím ze základních parametrů je tzv. nugget efekt, který označuje mikrovariabilitu (Ježek, 2015). Mikrovariabilita se podle Ježka (2015, s. 64) vyskytuje na vzdálenostech menších, „nežli je nejmenší vzdálenost datových bodů, takže zde nemáme informaci o tom, jak by mohl variogram vypadat“.



Obrázek 14: Teoretický variogram s označením základních parametrů. Zdroj: Ježek, 2015

Na základě znalosti prostorového jevu se můžeme rozhodnout, zda chceme nugget efekt zahrnout do autokorelace, pokud je mikrovariabilita přirozenou součástí prostorového jevu, nebo ho modelovat čistě zvlášť, pokud se spíše jedná např. o chyby měření. Tento druhý případ je vidět na Obrázek 12, kdy je teoretický variogram součtem modelu samotného nugget efektu a modelu s nulovým nugget efektem. Již zmiňovaný Matérnův model má, na rozdíl od jiných modelů variogramů (např. sférický model), výhodu ve své flexibilitě. Obsahuje navíc parametr kappa, který určuje hladkost průběhu funkce. Vyšší kappa znamená vyšší flexibilitu (Ježek, 2015).

2.4.3 Základy krigingu

Kriging je statistická metoda prostorové interpolace spadající mezi základní nástroje geostatistiky. Poskytuje odhady neznámých hodnot a díky tomu, že vychází z teorie Gaussových procesů, může určit také s těmito odhady spojené nejistoty. Princip krigingu je založen na váženém průměru, kde jsou krigingové váhy odvozeny z odhadované prostorové struktury, popisované teoretickým variogramem (Moraga, 2023; Ježek, 2015).

Podle studijních materiálů (Dobrovolný, 2005b) se dá teoreticky říci, že kriging na interpolovanou prostorovou veličinu nahlíží jako na náhodný prostorový jev, který je tvořen ze tří složek. První složka, označující se jako obecný trend (anglicky drift), je komponentou popisující strukturu konstantním průměrem nebo trendem. Někdy se označuje také za deterministickou složku. Druhá složka vyjadřuje naopak náhodnou (stochastickou) část, která se dá popsat pomocí prostorové korelace vyjádřené variogramem. Třetí složka je označována jako náhodný šum, protože její náhodné hodnoty neprojevují prostorovou autokorelaci. V libovolném bodě x_i lze tak hodnotu interpolované proměnné $Z(x_i)$ popsat jako: $Z(x_i) = \mu(x_i) + \varepsilon'(x_i) + \varepsilon''$. Hodnota Z je náhodnou funkcí, nikoliv měřenou hodnotou (Dobrovolný, 2005b).

V krigingu může složka průměru $\mu(x)$ vykazovat nějaký trend (drift), který se dá modelovat zvláště deterministicky nebo se dá v jednodušším případě předpokládat, že je průměr v konstantní (Dobrovolný, 2005b). Předpokladem pro odhad je, že jsou hodnoty náhodné funkce (Z) v prostoru tzv. vnitřně stacionární. Vnitřní stacionarita je rozšířením konceptu slabé stacionarity neboli stacionarity druhého řádu. Slabá stacionarita předpokládá, že je střední hodnota a kovariance mezi dvěma body nezávislá na posunutí, při zachování jejich vzájemné konfigurace. Změny těchto parametrů mohou být závislé tedy jen na směru nebo vzdálenosti obou bodů od sebe. „Stacionarita kovariance implikuje též stacionaritu rozptylu, neboť rozptyl je dán hodnotou kovariance v počátku, $\text{var } Z = C(0)$ “ (Ježek, 2015, s. 75). Jelikož nemají některé děje v geostatistice konečný rozptyl, slabá stacionarita je formulována poněkud odlišně, jako stacionarita přírůstků prostorové proměnné. Jedná se o vnitřní stacionaritu, kde se střední hodnota označuje nejčastěji jako E namísto μ (Ježek, 2015).

Očekávaný rozdíl mezi náhodnými hodnotami $Z(x)$ a $Z(x+h)$, které jsou od sebe vzdálené o vektor vzdálenosti h , by tedy měl být nula: $E[z(x) - z(x + h)] = 0$.

Druhým předpokladem je, že by měl rozptyl rozdílů záviset pouze na vzdálenostech mezi místy (x) a $(x+h)$: $var[z(x) - z(x+h)] = 2\gamma(h)$. Funkce $\gamma(h)$ je rovna polovině variance a označuje se jako semivariogram. Pokud je tedy odhadnut příspěvek obecného trendu $\mu(x)$ a zbývající složky mají konstantní rozptyl, rozdíly mezi hodnotami náhodné funkce Z , jsou jen funkcí jejich vzájemné vzdálenosti. Předchozí vzorec pro hodnotu v náhodném bodě x_i lze tedy přepsat jako: $Z(x_i) = \mu(x_i) + \gamma(h) + \varepsilon''$. Hodnoty $\varepsilon'(x)$ a $\gamma(h)$ jsou v tomto případě ekvivalentní (Ježek, 2015; Dobrovolný, 2005b).

V kapitole 2.4.2 věnované geostatistice a variogramu bylo již naznačeno, že krigingu předchází analýza prostorové struktury pomocí variogramu. Z dříve uvedených vzorců pro vnitřní stacionaritu tedy vyplývá, že lze předpoklad stacionarity pro kriging uvažovat, jako konstantní střední hodnotu a neměnnou podobu variogramu v celém prostoru. Variogram je tak závislý pouze na vzdálenosti párů od sebe a může být i na směru, ačkoliv jsou data uvažována jako izotropní. Konstrukce empirického variogramu lze omezit pro různé směry. Porovnáním variogramů pro všechny směry lze ověřit, zda jsou data izotropní, případně lze všechny variogramy modelovat jedním modelem a modelovat tak případnou anizotropii. Modelováním anizotropie bude výsledný odhad přesnější, stejně jako kvalitou nafitování modelu variogramu pomocí metody nejmenších čtverců. Průběh teoretického variogramu v podstatě popisuje průběh interpolované plochy, proto například při volbě parametru hladkosti κ u Matérnovy kovarianční funkce, volíme i hladkost výsledného povrchu. Při zahrnutí mikrovariability do autokovariance, například pomocí exponenciálního modelu variogramu, vytvoříme zase nevyhlazenou interpolovanou plochu plnou detailů, které mohou být jindy považovány za chybu a odstraněny modelem nugget efektu. Přístupů pro popis prostorové struktury pro kriging je tedy mnoho (Moraga, 2023).

V rámci procesu krigingu se popis vzájemných vztahů hodnot, pomocí teoretického variogramu, přepočítá na krigingové váhy, kterými poskytuje nejlepší lineární nevychýlený odhad. Proto se kriging anglicky často označuje také jako Best Linear Unbiased Prediction (BLUP), což označuje základní podmínky krigování. Odhad hodnoty krigingu $Z^*(x_0)$ je, na základě váženého průměru, v bodě x_0 realizován pomocí lineární kombinace:

$$Z^*(x_0) = \sum \lambda_i \times Z(x_i)$$

Pro váhy zároveň musí platit, že je jejich součet roven jedné, $\sum \lambda_i = 1$. Pro nevychýlený odhad se předpokládá, že je střední hodnota chyby odhadu nulová, $E[Z(x_0) - Z^*(x_0)] = 0$. Za nejlepší odhad se zase považuje ten, který dává nejmenší rozptyl kolem správné hodnoty, $\sigma_k^2 = \text{var}[Z(x_0) - Z^*(x_0)] = \min$. Všechna tato kritéria nakonec určují výslednou podobu soustavy lineárních rovnic krigingu (Ježek, 2015; Dobrovolný, 2005c).

Ježek (2015, s. 109) upřesňuje, že je rozptyl krigingu σ_k^2 (anglicky kriging variance), který se snažíme minimalizovat, zároveň vyjádřitelný pomocí kovariance nebo variogramu. Tato hodnota vyjadřuje teoretickou míru nejistoty odhadů a může stejně jako interpolované odhady tvořit souvislý povrch. Ačkoliv se tento způsob vyjádření nejistoty objevuje nejčastěji, někde se používá odmocnina z tohoto rozptylu, tedy směrodatná chyba (odchylka) krigingu (anglicky standard error). Ta má výhodu, že je ve stejných jednotkách jako interpolovaná hodnota (Dobrovolný, 2005c).

2.4.4 Typy krigingu

Za základní typ krigingu se považuje tzv. ordinary kriging (OK), ze kterého lze odvodit i ostatní typy krigingu. Nejčastěji se provádí jako tzv. bodový kriging (point kriging), což vyjadřuje, že se interpoluje mezi body. Bodový odhad lze určitým způsobem vztáhnout i na plochu nebo objem v prostoru, čemuž se říká blokový kriging (block kriging). Nejčastěji se využívá, pokud chceme pomocí krigingu interpolovat strukturu pravidelných buněk (grid). V krigingu lze také, stejně jako u metody IDW, omezit okolí, ve kterém se budou hledat nejbližší sousedé pro výpočet krigingových vah. Okolo neznámého bodu se vytvoří region hledání (anglicky search neighborhood), ve kterém se kriging chová jako lokální interpolátor a počítá pouze s kovarianční strukturou okolních bodů v regionu. Při lokálním omezení to také znamená, že data mohou být z hlediska střední hodnoty stacionární jen v rámci daného regionu a není potřeba, aby byly stacionární globálně. Omezením výpočtu vah na vybraný region se zároveň odstraní vliv vzdálených bodů a výrazně urychlí čas výpočtu krigingu, který je jinak výpočetně velmi náročný (Dobrovolný, 2005c).

Pro charakterizování jednotlivých typů krigingu lze uvažovat zjednodušený model: $Z(x_i) = \mu(x_i) + \varepsilon(x_i)$. Proměnná vycházející z náhodné funkce $Z(x_i)$, se v bodě x_i skládá z deterministické hodnoty trendu $\mu(x_i)$ a stochastické náhodné veličiny $\varepsilon(x_i)$, vykazující autokorelaci. Pro náhodnou složku přitom platí vnitřní

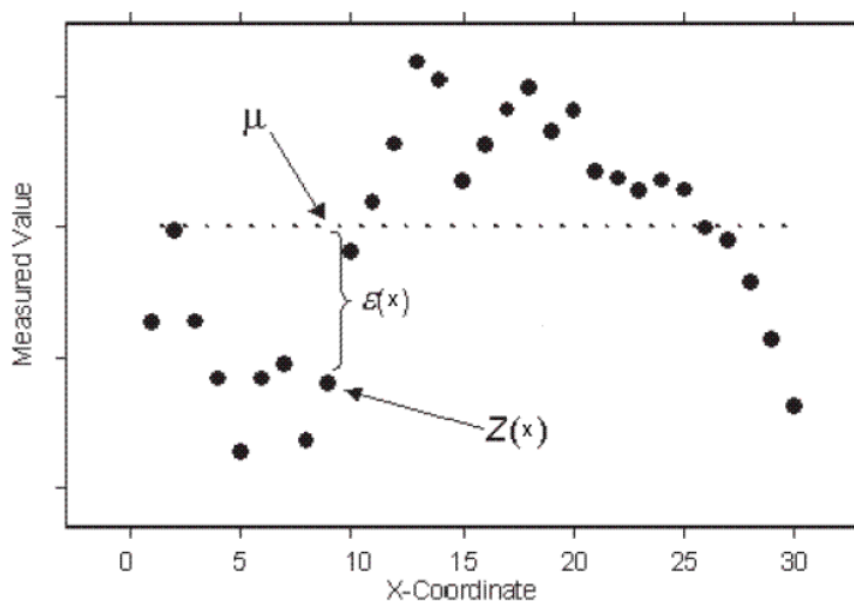
stacionarita. V závislosti na tom, jestli je hodnota $\mu(x_i)$ známá, či neznámá, a také podle toho, zda představuje konstantní hodnotu průměru $\mu(x_i) = \mu$ nebo trend, jsou stanoveny různé typy metod krigování (Dobrovolný, 2005c). Některé z typů krigingu byly uvedeny v následující tabulce, je jich však mnoho:

Tabulka 2: Základní typy krigingu. Upraveno podle: Ježek, 2015

Typ metody	Popis metody
Simple kriging (SK)	předpokládá známou střední hodnotu v celé oblasti
Ordinary kriging (OK)	předpokládá neznámou střední hodnotu, konstantní v celé oblasti
Universal kriging (UK)	uvažuje trend stejného typu v celé oblasti
Regression kriging (RK)	kombinace lineární regrese trendu s krigingem residuí
Indicator kriging (IK)	namísto veličiny se pracuje s indikátorovou funkcí
Lognormal kriging (LOK)	kriging lognormální veličiny (transformace veličiny)
Cokriging (CK)	využívá další autokorelovanou veličinu
External drift kriging (KED)	využívá korelaci se známou prostorovou veličinou

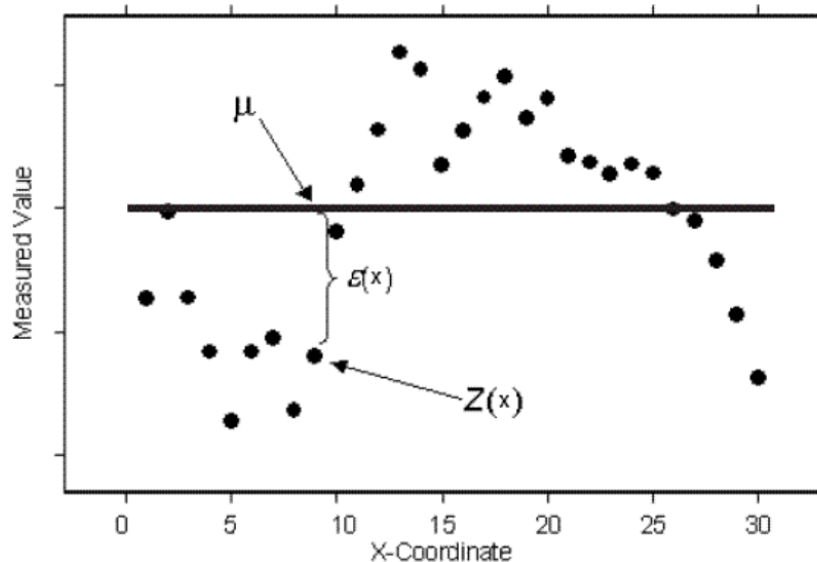
Za základní tři typy krigingu, lze považovat simple kriging (SK), ordinary kriging (OK) a universal kriging (UK). Všechny tyto typy vychází ze zmiňovaného obecného modelu: $Z(x_i) = \mu(x_i) + \varepsilon(x_i)$.

Základní krigování, obyčejný kriging, nebo ordinary kriging (OK), označuje koncept krigingu, kterému se věnuje předchozí kapitola 2.4.3. Jedná se o případ, kdy neznáme střední hodnotu μ a pokoušíme se ji odhadovat, jako konstantní v celé oblasti.



Obrázek 15: Princip ordinary krigingu (OK). Zdroj: Dobrovolný, 2005c

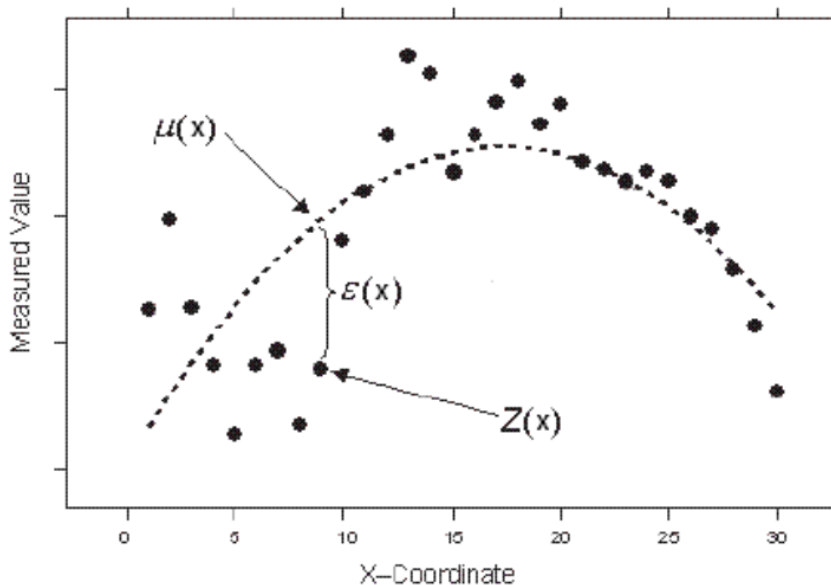
Jednoduché krigování, neboli simple kriging (SK), je nejjednodušší variantou krigování. Pro výpočet je nutné znát střední hodnotu veličiny μ v celé oblasti. Narozdíl od OK je tedy μ známou konstantou, kterou není nutné odhadovat, a proto SK nabízí přesnější odhad autokorelace, tedy náhodné složky $\varepsilon(x)$ (Dobrovolný, 2005c).



Obrázek 16: Princip simple krigingu (SK). Zdroj: Dobrovolný, 2005c

Předpoklad znalosti μ je však málokdy reálný. SK se často využívá jako součást regression krigingu (RK). Pomocí RK se nejprve modeluje deterministická část, například pomocí lineární regrese nebo jiných metod, a na rezidua těchto modelů se pak aplikuje simple kriging, který pro ně uvažuje nulovou střední hodnotu.

Universal kriging (UK) odhaduje $\mu(x)$ pomocí deterministické funkce (trend) a zároveň modeluje autokorelovanou náhodnou složku $\varepsilon(x)$ pomocí variogramu.



Obrázek 17: Princip universal krigingu (UK). Zdroj: Dobrovolný, 2005c

3. Metodika

3.1 Stanovení kritérií a metod tvorby modelu

Tato kapitola je věnovaná procesu stanovení kritérií pro model výšky vegetace (CHM), dále v práci označovaný jen jako model, vytvořený prostorovou interpolací z dat satelitní laserové altimetrie misí GEDI a ICESat-2, jehož tvorba byla hlavním cílem diplomové práce. Veškerá kritéria a předpoklady, rozebírané v této kapitole, pak poskytují důležitý rámec pro volbu metod popsanych v dalších kapitolách.

Práce se především věnuje využití dat výšky vegetace z mise GEDI, přičemž data z mise ICESat-2 sloužila spíše ke zhuštění pokrytí vybraného území. Měření provedená pomocí satelitního lidarů bývají typicky řídká a obecně spojena s vysokou mírou nejistoty (viz kapitola 2.2), proto bylo vhodné zvolit statistickou metodu prostorové interpolace, která by umožňovala kvantifikovat nejistoty při predikci hodnot na nezměřených místech. Z tohoto důvodu byla vybrána metoda známá jako kriging, která explicitně zahrnuje odhad nejistoty jako součást výstupu (viz kapitola 2.4.3 o základech krigingu).

Kriging vychází z předpokladu výskytu prostorové autokorelace, jejíž míru bylo v rámci této práce nezbytné zanalyzovat. Tento předpoklad zdůrazňuje potřebu zajistit co nejhustší pokrytí vybraného území dostupnými daty o výšce vegetace a zároveň maximalizovat kvalitu těchto dat. Výběr vhodné filtrační metody byl klíčový pro minimalizaci neautokorelované variability a zajištění maximální přesnosti výsledků interpolace. Tvorbou modelů výšky vegetace z dat z GEDI a ICESat-2 se věnuje již více studií, a to jak na globální (Lang et al., 2022a; Potapov et al., 2021), tak na lokální úrovni (Schwartz et al., 2022; Sothe et al., 2022; Li et al., 2020). Tyto studie preferují spíše metody strojového učení než geostatistické metody odhadu, jako je kriging. Společně se studii zabývajícími se tvorbou CHM z ICESat-1 (Simard et al., 2011; Lefsky, 2010), byly však podrobně prostudovány, protože pojednávají o kvalitách a problémech dat ze satelitního lidarů. Schwartz et al. (2022) se například věnuje posouzení kvality produktu GEDI L2A Version 2 a porovnává ho se starší verzí GEDI L2A Version 1 z hlediska geolokačních chyb. Ve studii potvrzuje téměř dvakrát přesnější zaměření u verze 2 s geolokační chybou (~11 m).

Výběr vhodných produktů (GEDI L2A v002) a (ICESat-2 L3A – ATL08, v006) a výběr vhodných zdrojů dat byl také podmíněn dostupností důležitých atributů

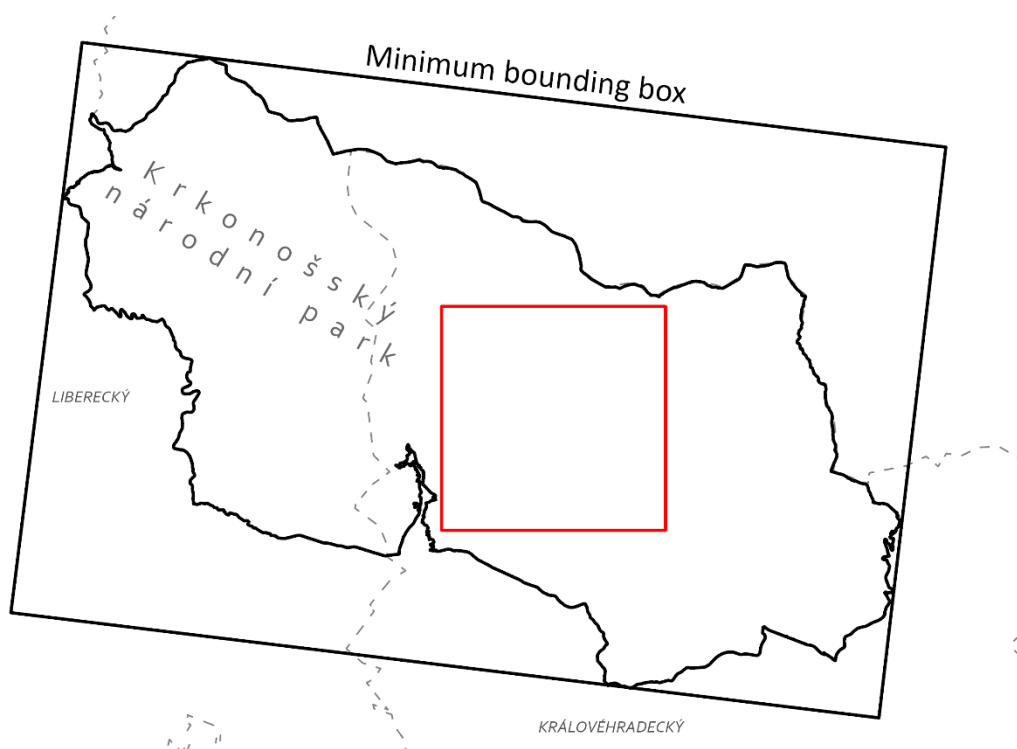
pro posouzení kvality. Z dokumentace GEDI (Beck et al., 2021) bylo posouzeno, které pak budou vybrány pro filtraci kvalitních měření ve fázi výběru dat (subsetting). Jako další krok k dosažení optimálních výsledků byla zvolena metoda univerzálního krigingu (UK). Tato technika umožňuje modelování deterministické složky neboli globálního trendu v datech, s využitím prediktorů získaných pomocí jiných metod dálkového průzkumu Země (optická či radarová data). Pro tento účel byla zvolena kombinace dat ze satelitu Sentinel-1, což je radar se syntetickou aperturou (SAR), a optických snímků ze Sentinel-2, z nichž byly odvozeny různé vegetační metriky.

Na základě očekávané vyšší hustoty dat z GEDI v severních Čechách, kvůli specifické orbitě ISS, která se otáčí na 51,6 s. š., bylo pro tvorbu modelu výšky vegetace zvoleno území Krkonošského národního parku (KRNAP). Výběr této oblasti byl motivován jedinečnými ekologickými a terénními charakteristikami, kde se vyskytuje velká proměnlivost ve výšce vegetace a terén je v mnoha místech poměrně náročný. Tato rozmanitost prostředí představovala jak výzvu, tak příležitost pro testování zvolených metod. Další výhodou je případná využitelnost výsledného modelu výšky vegetace pro potřeby správy KRNAP. Otáčení ISS na 51,6° s. š. způsobuje, že se orbity potkávají častěji, což vede k hustějšímu datovému pokrytí. Toto husté bylo dalším faktorem, který umožnil využití pokročilých metod prostorové interpolace, jako je například kriging.

V neposlední řadě byla snaha vytvořit vlastní CHM tak, aby bylo porovnatelné s globálními CHM (Lang et al., 2022a; Potapov et al., 2021). Model Lang et al. (2022a) disponoval prostorovým rozlišením 10 m a využíval data z GEDI metriky RH95 (95. percentil výšky vegetace), které integroval do modelu strojového učení. Autoři modelu volili 10 m rozlišení, protože jako vysvětlující proměnné pro RH95 používali optická data ze satelitu Sentinel-2. Naproti tomu model Potapov et al. (2021) kombinoval výškovou metriku RH98 s časovou řadou optických dat z misí Landsat, výsledkem čehož byl model s prostorovým rozlišením 30 m. Oba uvedené globální modely byly zpřístupněny veřejnosti, což umožnilo využít jejich data pro srovnání s modelem vytvořeným pomocí krigingu. Aby se minimalizovalo riziko nadhodnocení výšky vegetace a zároveň zajistila kompatibilita s globálními modely, byl pro určení výšky vegetace v této práci zvolen 95. percentil (atribut RH95 z GEDI). Pro data získaná z mise ICESat-2 byl specificky vybrán atribut h_canopy, který odráží 98. percentil, a byl uznán za spolehlivou metriku výšky vegetace. Aby se dosáhlo konzistence

rozlišení s daty získanými ze satelitů Sentinel-1 a Sentinel-2, která byla využita jako vysvětlující proměnné v procesu univerzálního krigingu, bylo rovněž v souladu se studií (Lang et al., 2022a) preferováno prostorové rozlišení CHM 10 metrů.

V průběhu této práce se ukázalo, že výběr kritérií a metodologie vyžadoval pružnou adaptaci, která byla možná díky pokračujícímu zkoumání dané problematiky. Například bylo zásadní optimalizovat výpočetní dobu pro kriging, který je metodou s vysokými nároky na výpočetní výkon. Kvůli minimalizaci náročnosti výpočtu bylo tedy přistoupeno k výběru menšího vzorku území 10x10 km v oblasti KRNAP.



Obrázek 18: Vybrané území pro stahování dat (minimum bounding box), území pro realizaci cílů práce (KRNAP) a vybrané území 10x10 km pro tvorbu modelu (červený čtverec)

Po extrakci dat z HDF5 souborů pocházejících z mise GEDI a eliminaci nekvalitních bodů měření s využitím atributu `quality_flag` bylo také zaznamenáno významné snížení objemu dat, přičemž se zachovalo pouze přibližně 30 % původního množství. V důsledku toho byly pro maximalizaci využitelných dat studovány různé alternativní zdroje a produkty dat GEDI, stejně jako byly testovány rozličné metody filtrace kvalitních měření. Každé měření z GEDI představovalo oblast s průměrem 25 m (tzv. footprint), zatímco data z mise ICESat-2 byla rozdělena do segmentů o rozměrech 13x100 m. Vzhledem k tomu, že všechna data získaná satelitním lidarem tvoří rozsáhlé agregace, je evidentní, že vytvořené CHM disponují relativně omezeným prostorovým rozlišením, které neumožňuje rozlišit jednotlivé stromy tak, jak je to možné u detailnějších CHM získaných pomocí ALS.

3.2 Získávání dat

3.2.1 Stahování a výběr dat

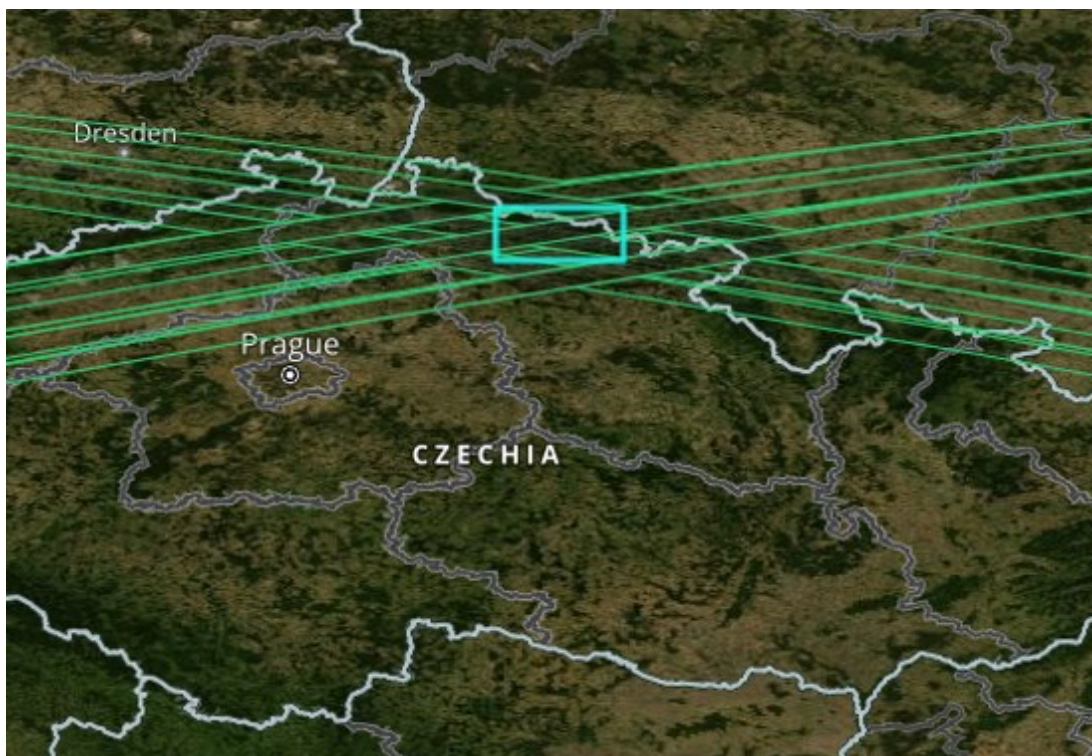
Před zahájením stahování dat bylo nutné provést důkladnou přípravu, vzhledem k širokému spektru dostupných služeb a nástrojů pro výběr a stažení dat. Bylo klíčové definovat specifické požadavky, jako jsou požadované atributy dat, způsob jejich zpracování a požadovaný formát. Z vybraného datového produktu pak bylo důležité určit, které atributy jsou klíčové a které lze vyřadit za účelem redukce velikosti souborů pro stažení – proces známý jako subsetting. Pro efektivní omezení geografického rozsahu dat byla vytvořena minimální ohraničující oblast (minimum bounding box), konkrétně obdélník zahrnující Krkonošský národní park (KRNAP). Navíc bylo zvaženo rozšíření této oblasti, aby se minimalizoval možný okrajový efekt, který může nastat při krigingu. Tento jev se projevuje zejména na okrajích studované oblasti, kde dochází k extrapolaci v důsledku nedostatečného pokrytí datovými body.

Pro získávání dat z misí GEDI a ICESat-2 lze využít online platformy, jako jsou NASA Earthdata Search (search.earthdata.nasa.gov), Google Earth Engine (earthengine.google.com), a specificky pro data z mise ICESat-2 je k dispozici NASA's OpenAltimetry (openaltimetry.earthdatacloud.nasa.gov). Stahovací portál NASA Earthdata Search zahrnuje klienta Earthdata Search Client (EDSC), který umožňuje efektivní vyhledávání a stahování požadovaných dat v rozsáhlých objemech. Uživatelé mohou vytvořit vlastní objednávku, jež nabízí možnost subsettingu a komprimuje stahovaná data do formátu ZIP. Kromě toho jsou k dispozici další nástroje pro dotazování a stahování dat z mise GEDI, včetně GEDI_finder, DAAC_downloader, APPEARS a DAAC2Disk. Například pomocí GEDI_Finder je možné stahovat data ve formátu CSV s využitím jazyků R nebo Python. Pro Python existuje také nástroj GEDI_Subsetter, který umožňuje vybírat relevantní vzorky z datové sady. Nicméně veškeré zpracování dat v rámci této práce bylo realizováno v programovacím jazyce R.

Stahování bylo nakonec realizováno prostřednictvím portálu NASA Earthdata Search z důvodu předností, které tento způsob nabízí, jako možnost stažení dat ve formátu HDF5 pro jednotlivé orbity nebo hromadně na základě specifického výběru, subsettingu umožňujícího přímý výběr atributů, a komprimace dat do formátu ZIP, což vedlo k jejich významnému zmenšení. Následné manuální rozbalování souborů a

konverze dat HDF5 do formátu CSV nebo formátu vhodného pro R však vyžadovalo značné časové úsilí.

Produkt GEDI L2A, tedy úroveň 2, obsahující zpracovaná data a metriky RH (percentily výšky vegetace), byl stahován vždy ve verzi 002, jež poskytovala výrazně přesnější geolokaci měření (geolokační chyba ~11 m). Každé bodové měření v těchto datech reprezentuje stopu s průměrem 25 metrů. Pro obdélníkové území Krkonošského národního parku byla data tohoto produktu stažena za všechna dostupná období s cílem maximalizovat hustotu pokrytí území. Dostupnost dat byla pro období od 14. dubna 2019 do 22. února 2023. V tomto časovém rozmezí bylo staženo celkem 245 orbitů (souborů), v nichž bylo zaznamenáno 870 005 měření, z nichž 869 772 bylo vyhodnoceno jako platných. Mezi neplatná měření patřila například ta, která neobsahovala informace o geolokaci (hodnoty `lat_lowestmode` a `lon_lowestmode`).



Obrázek 19: Vybrané území pro stahování dat s orbity GEDI. Zdroj: NASA, search.earthdata.nasa.gov

Stahování dat z mise ICESat-2 bylo provedeno metodou analogickou k té, jaká byla aplikována na data GEDI, za účelem jejich doplnění a zvýšení hustoty datové sady získané z mise GEDI. Pro stejné území byl zvolen produkt ICESat-2 (L3A – ATL08, v006), který představuje 3. úroveň zpracování, nabízející specializované produkty pro rozličné environmentální studie. Produkt ATL08 poskytuje údaje o výšce terénu a vegetace, přičemž každé měření je reprezentováno centroidem segmentu o

velikosti 100 x ~13 m, v jehož rámci dochází k agregaci údajů o výšce vegetace v důsledku analýzy fotonů. Data byla opět ve formátu HDF5 a byla stahována po jednotlivých orbitech. V období od 23. listopadu 2018 do 14. srpna 2023 bylo staženo celkem 64 orbitů (souborů), obsahujících dohromady 32 703 měření, z nichž 32 622 měření neslo informace o geolokaci (atributy longitude a latitude). Celkově byla stažena data z misí GEDI a ICESat-2 pro období od roku 2018 do roku 2023.

Jako alternativní zdroj dat z mise ICESat-2 mohl sloužit portál NASA's OpenAltimetry, který však umožňoval stahování pouze jednotlivých orbitů zvlášť nebo výběr specifických vzorků dat ve formátu CSV.

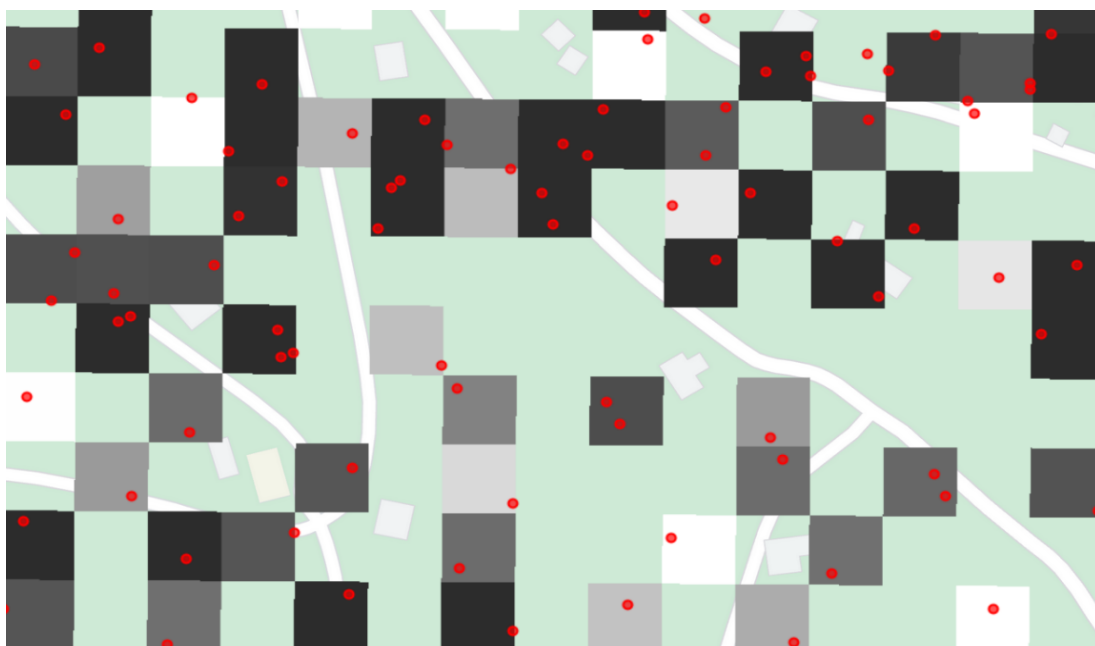
3.2.2 Porovnání zdrojů dat GEDI

V rámci snahy o maximalizaci hustoty dat na zvoleném území byly dále prozkoumány alternativní nástroje pro stahování a zpracování dat z mise GEDI. Datová sada GEDI představovala na území základní vrstvu měření, díky jejímu většímu pokrytí, které převyšovalo pokrytí dat z mise ICESat-2. Hlavním alternativním zdrojem těchto dat byl Google Earth Engine (GEE), což je platforma navržená pro analýzu a zpracování nejrůznějších dat dálkového průzkumu Země. Data GEDI L2A z Google Earth Engine, ač pochází ze stejného původního zdroje (NASA's Land Processes DAAC) jako data dostupná skrze NASA Earthdata, jsou poskytována v odlišně předzpracované formě, optimalizované pro rychlou vizualizaci (na rozdíl od původních HDF5 souborů). Další motivací pro průzkum alternativních zdrojů bylo přitom najít platformu, která umožňuje efektivnější stahování a zpracování objemných datových sad GEDI, aby byly rychle dostupné pro tvorbu CHM v prostředí RStudia.

Kvalita produktů GEDI získaných prostřednictvím Google Earth Engine byla posuzována srovnáním s nezpracovanými daty ve formátu HDF5, dostupných na NASA Earthdata. Aby bylo zajištěno relevantní porovnání, vybraná data z GEE byla omezena tak, aby odpovídala stejnému geografickému rozsahu a časovému rámci. Specificky pro období 2019–2023 a území KRNAPu byly zvoleny produkty GEDI L2A Vector Canopy Top Height (Verze 2) a GEDI L2A Raster Canopy Top Height (Verze 2), reprezentující vektorovou a rastrovou verzi produktu GEDI L2A.

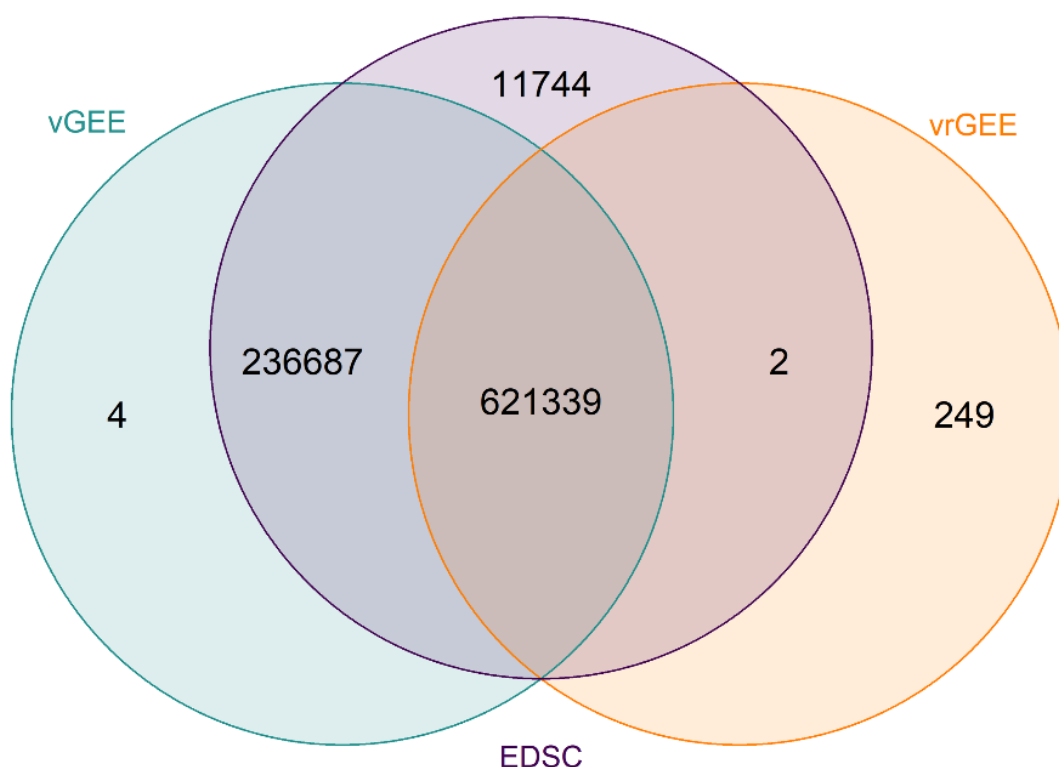
Pro načtení vektorového produktu bylo v JavaScriptu použito celkem 261 objektů `ee.FeatureCollection`, které reprezentovaly individuální orbity. Import dat byl možný pouze skrze explicitní specifikaci cesty k danému orbitu, přičemž některé

orbity chyběly nebo je nebylo možné naimportovat. Následně bylo provedeno sloučení a zpracování těchto dat do CSV, s přidáním sloupců X a Y pro geolokaci v souřadnicovém systému WGS84. Export bylo také možné provést ve formátu KML, avšak import rozsáhlého KML souboru do R by vyžadoval značné množství času. Rastrová verze produktu GEDI L2A byla distribuována jako měsíční průměry ze všech orbitů s prostorovým rozlišením 25 m, což by umožnilo rychlé stahování a snadnou manipulaci s daty při redukované velikosti souborů na disku. Avšak pro účely srovnání byla vyžadována vektorizace těchto dat. Dostupný polohový atribut u tohoto produktu v GEE byl jen `lat_highestreturn` (nikoliv `lat_lowestmode`), který označuje polohu nejvyššího odrazu, u GEDI často zatíženou velkou chybou. Tato skutečnost znemožňovala určení přesné polohy měření z pixelů. Nezbytné proto bylo provést vektorizaci těchto 25 m pixelů formou vytváření jejich centroidů, což zase vedlo k značné odchylce od původních měřicích pozic. Vektorový produkt oproti tomu nabízel souřadnice ve formátu WGS84, kde byla poloha definována s přesností na centimetry (nutné zohlednit geolokační chybu ~11 m). V rastrové verzi je však jakákoliv přesnost geolokace neznámá. Potvrzeno také bylo, že je rastrový produkt silně agregovaný, protože se několik měření z vektorového produktu nacházelo v oblasti jednoho pixelu. Vzhledem k tomu, že šlo o měsíční průměry ze všech orbitů, nebylo vhodné tyto data používat v oblastech kde se ISS otáčí, protože pokud by na plochu jednoho pixelu připadlo za měsíc více měření, data se zprůměrují.



Obrázek 20: Ukázka agregace rastrového produktu GEDI L2A v Google Earth Engine v porovnání s vektorovým

Jak už bylo naznačeno, pro účely analýzy a porovnání produktů GEDI získaných z Google Earth Engine (GEE) s daty získanými přes NASA Earthdata bylo nezbytné zajistit jejich vzájemnou konzistenci. V produktech z GEE bylo například zaznamenáno odlišné označení paprsků, kde byla náhrada názvu jednotlivých z osmi paprsků číslicemi: BEAM0000 = 0, BEAM0001 = 1, BEAM0010 = 2, BEAM0011 = 3, BEAM0101 = 5, BEAM0110 = 6, BEAM1000 = 8, BEAM1011 = 11. Rozdíly byly také pozorovány v počtu orbitů, měření a dostupných atributů mezi oběma zdroji. Atribut `shot_number`, který představuje unikátní identifikátor každého měření GEDI, byl využit pro vytvoření Vennova diagramu. Zajímavým zjištěním je, že produkt vektorizovaného rastru GEDI L2A z GEE obsahuje v porovnání s daty z NASA Earthdata 249 unikátních měření navíc a oproti vektorové verzi GEDI z GEE je rozdíl 250 unikátních objektů. Zkratkou EDSC jsou v grafu označována data L2A z Earth Data Search Client (.h5), vGEE jsou vektorová data L2A z Google Earth Engine a vrGEE jsou rastrová data L2A z GEE (vektorizovaná).



Obrázek 21: Vennův diagram zobrazující počty a shodu měření v každém z datasetů

Data z NASA Earthdata ve formátu HDF5 mají tedy více měření a kompletnější sadu atributů oproti omezenějším datům GEDI z GEE.

3.3 Předzpracování dat

3.3.1 Výběr vzorku

V rámci této diplomové práce byly veškeré operace s daty, jejich zpracování a tvorba CHM prováděny v prostředí RStudio s využitím programovacího jazyka R. Vzhledem k tomu, že zdrojem dat pro mise GEDI a ICESat-2 byla platforma NASA Earthdata, vyvstala potřeba vytvoření R skriptu pro efektivní práci s hierarchicky strukturovanými soubory HDF5. Tyto soubory, sloužící jako distribuční formát dat, bylo nutné zpracovat tak, aby bylo možné s nimi dále pracovat v R nebo je exportovat do CSV. Formát HDF5 je charakteristický svou strukturou, připomínající systém složek uchovávajících hodnoty jednotlivých atributů měření.

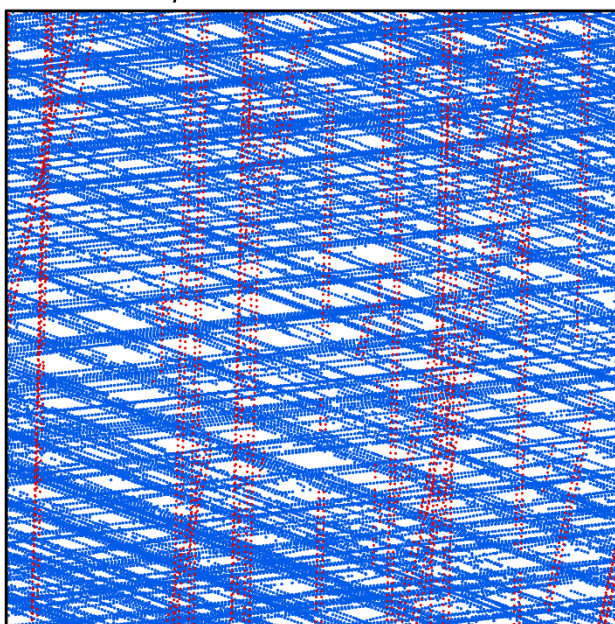
Konkrétně u dat z mise GEDI bylo nutné zpracovat 245 souborů HDF5, zastupujících jednotlivé orbity, přičemž každý z nich detailně popisoval atributy měření rozdělené mezi osm paprsků. K lepšímu pochopení obsažených atributů produktu GEDI L2A byly prozkoumány jejich seznamy poskytnuté distributorem (GEDI L2A Product Data Dictionary, *n.d.*). Tyto seznamy a příslušná dokumentace (Beck et al., 2021) umožnily identifikovat přímé cesty k atributům, které bylo možné v R načíst pomocí balíčku `rhdf5`. Pro každé měření bylo například přiřazeno unikátní ID, známé jako atribut `shot_number`.

Přístroj GEDI obsahuje osm paprsků – čtyři "silné" (power beams) a čtyři "slabé" (coverage beams), které byly v HDF5 souborech identifikovány specifickými názvy. Pro dosažení potřebné hustoty měření byly využity všechny paprsky. Dále byly vybrány atributy relativní výšky vegetace (RH metriky), zejména `RH95`, `RH98`, a `RH100`, umístěné ve složkách odpovídajících jednotlivým paprskům, například pro paprsek `BEAM0101` ve složce `BEAM0101/rh`. Souřadnice měření byly uchovány v attributech `lat_lowestmode` a `lon_lowestmode` v souřadnicovém systému WGS84. Podobným způsobem byly extrahovány další základní atributy, jako je `quality_flag`, `sensitivity`, `solar_elevation`, `degrade_flag`, `num_detectedmodes` a `elev_lowestmode`, označující nadmořskou výšku terénu. Většina z nich byla zpracována více algoritmy `a1-a6`, z nichž byly vybrány pouze první dva (viz vysvětlení atributů v kapitole 2.2.1). Výběr specifických parametrů návratové vlny, energie návratového signálu, nebo rozdílů od terénu, byl inspirován studií Kellner et al. (2023) s cílem umožnit pozdější filtrace kvalitních měření. Pomocí cyklu `FOR` v R bylo takto z jednotlivých 8 paprsků GEDI vybráno 54 různých atributů (sloupců) pro celkem 870 005 měření (řádků).

Při zpracování dat z mise ICESat-2 ATL08 byly atributy extrahovány z formátu HDF5 podobným způsobem. Data byla rozdělena do šesti paprsků, s rozlišením na levé (gt1l, gt2l, gt3l) a pravé (gt1r, gt2r, gt3r). Z těchto paprsků byly primárně extrahovány atributy *latitude* a *longitude* pro určení geografické polohy měření ve formátu WGS84, *h_te_best_fit*, který reprezentuje nadmořskou výšku terénu, *h_canopy* určující výšku vegetace, a *sc_orient* sloužící k identifikaci orientace trajektorie družice ICESat-2, což umožnilo následnou klasifikaci paprsků na slabé a silné podle studie Liu et al. (2022). Podobně jako u dat GEDI byl kladen důraz na výběr atributů relevantních pro následnou filtraci dat. V produktu ICESat-2 ATL08 se vedle výškových atributů nacházela také nejistota spojená s procesem odvozování těchto atributů z přijatých fotonů. Hodnoty výšky vegetace *h_canopy*, označující 98. percentil, měly například míru nejistoty uloženou v atributu *h_canopy_uncertainty*. Z jednotlivých šesti paprsků ICESat-2, uložených v 67 HDF5 souborech, bylo nakonec vybráno 12 atributů pro celkem 32 703 měření.

Po vytvoření tabulek z vybraných datových sad GEDI a ICESat-2, kde jednotlivá měření tvořila řádky a atributy byly reprezentovány jako proměnné ve sloupcích, bylo možné dále provádět explorativní analýzu. Pro práci s prostorovými objekty byl používán balíček SF (simple feature), pomocí kterého také probíhal převod všech souřadnic na systém S-JTSK (EPSG:5514). S ohledem na snížení výpočetní náročnosti prostorové interpolace, konkrétně metody krigingu, byl dále vybrán vzorek území 10x10 km v Krkonošském národním parku, který je vidět na Obrázek 22.

Vybrané území 10x10 km

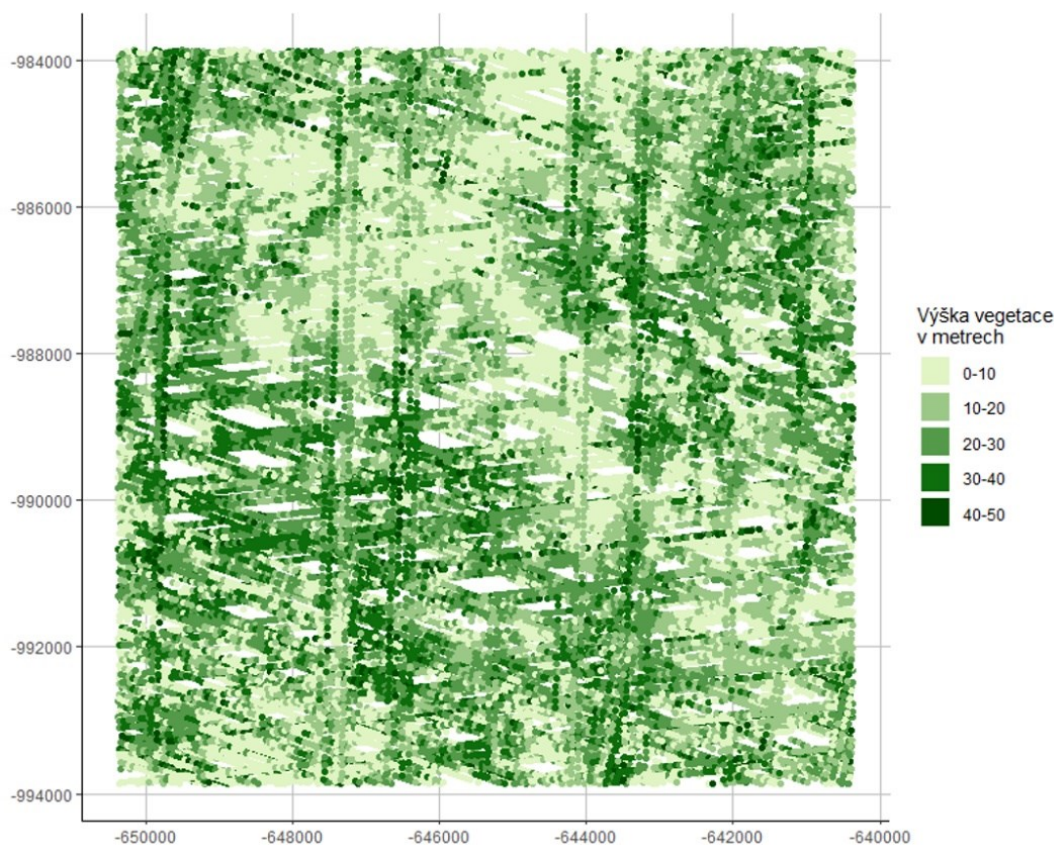


Obrázek 22: Vybraný vzorek území o rozměrech 10x10 km s body měření GEDI (modře) a ICESat-2 (červeně)

Na vybraném území o rozměrech 10x10 km bylo zaznamenáno celkem 81 441 měření z mise GEDI a 3 390 měření z ICESat-2. Pro účely interpolace byl často používán malý přesah 100 m ve všech směrech, aby byl zmírněn okrajový efekt. V neposlední řadě bylo nutné provést selekci vhodné RH metriky, která by sloužila jako vysvětlovaná proměnná výšky vegetace v modelu. Kvůli zajištění porovnatelnosti s globálními modely Lang et al. (2022a) a Potapov et al. (2020) byly uvažovány metriky RH95 nebo RH98. Nižší percentil výšky vegetace by navíc zajistil menší nadhodnocení a model by se tím přiblížil k validačním datům výšky vegetace z ALS. Pro jistotu byl u jednotlivých RH metrik prozkoumán lineární vztah s tímto validačním CHM z ALS pro rok 2012. Nejvyšší koeficient determinace R^2 , při porovnání metrik RH95, RH98 a RH100, vyšel pro RH95, $R^2 = 0,49$. Dále byly spočítány chyby ME (střední chyba), MAE (střední absolutní chyba) a RMSE (střední kvadratická chyba), podle kterých mělo RH95 opět nejlepší výsledek. Výběr 95. percentilu byl navíc v souladu s některými existujícími články (Lang et al., 2022a; Morin et al., 2022).

Tabulka 3: Porovnání validačního CHM z ALS (2012) s jednotlivými RH metrikami

Metrika	ME	MAE	RMSE	R^2
RH95	8,17	9,04	11,07	0,49
RH98	9,69	10,43	12,40	0,48
RH100	11,44	12,07	13,93	0,46



Obrázek 23: Výška vegetace z dat GEDI (RH95) a ICESat-2 (h_canopy) na vybraném území 10x10 km

3.3.2 Metoda filtrace a korekce dat

Dalším krokem předzpracování dat z GEDI a ICESat-2 byla filtrace kvalitních měření, která probíhala v několika fázích s cílem zajistit co nejmenší redukci počtu bodů a zároveň co nejvíce snížit odchylky od referenčního digitálního modelu terénu z roku 2012 (DTM 2012), pořízeného pomocí ALS (viz podrobnější popis v kapitole 3.5). Při geostatistické strukturní analýze (variografii) mohou chyby v měření zvýšit mikrovariabilitu a tím i rozptyl krigingu, který vyjadřuje nejistotu odhadů.

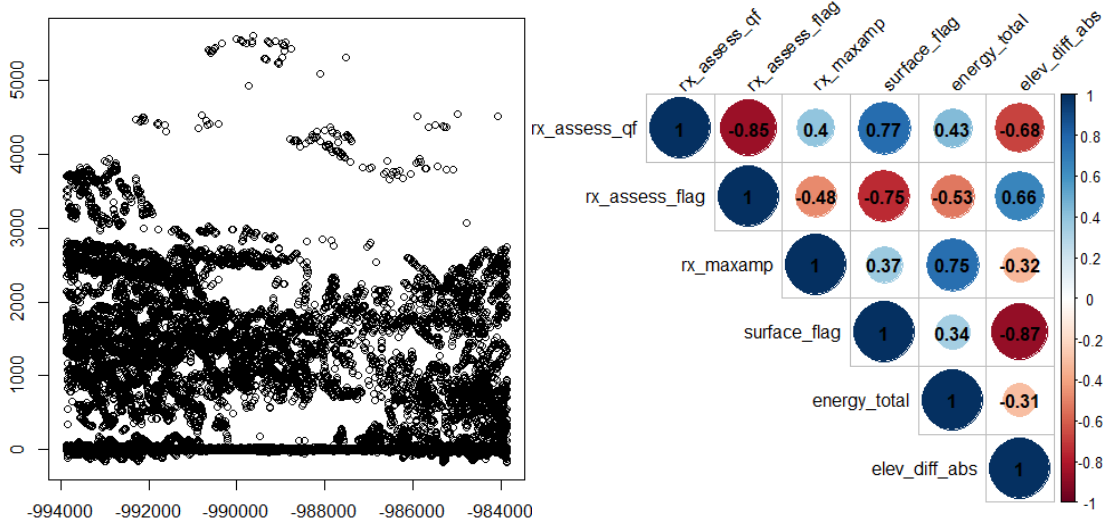
Pro obě datové sady bylo možné provést základní filtr, který k měřením přistupoval minimálně agresivně a současně eliminoval většinu prokazatelně chybných měření. U dat GEDI bylo cílem základního filtru odstranění všech měření s prázdným návratovým signálem. Počet lokálních maxim (modů), reprezentujících odrazy na návratové vlně, byl uložen v atributu *num_detectedmodes*. Měření, na jehož návratové vlně se nevyskytoval ani jediný odraz a hodnota tohoto atributu byla nula, bylo identifikováno jako neplatné a následně odstraněno. Tato filtrace spolehlivě odfiltrovala neúspěšná měření a nedošlo přitom k příliš výrazné redukci dat. V případě dat ICESat-2 bylo v atributu výšky vegetace *h_canopy* identifikováno 195 výrazně vysokých hodnot 3,402823³⁸. Tyto hodnoty byly odstraněny, jelikož označovaly chybu při odvozování výšky vegetace z fotonů.

V procesu filtrace obou datových sad bylo usilováno o minimalizaci chyb v určení terénu, což často zahrnovalo přímou filtraci odlehlých hodnot. Základem bylo vytvořit sloupec s odchylkami od referenčního DTM 2012. Měření nadmořské výšky terénu z GEDI (*elev_lowestmode*) a ICESat-2 (*h_te_bestfit*) bylo však určeno na elipsoidu WGS84, proto bylo pro účely porovnatelnosti s DTM 2012 z ALS nezbytné přetransformovat jejich výškový systém na baltský – po vyrovnání (Bpv). Pro tyto transformace byla použita aplikace Geoprohlížeč od ČÚZK (dostupná na: <https://ags.cuzk.cz/geoprohlizec>). Pokud by byl pro validační účely použit jiný referenční model terénu, například SRTM, bylo by tento krok možné vynechat. Po sjednocení výškového systému byl dále, jak pro GEDI, tak pro ICESat-2, vytvořen sloupec *elev_diff*, označující odchylku jednotlivých měření od DTM 2012.

Po každé fázi filtrace byla kvalita dat hodnocena s využitím kvantifikace chyb (ME, MAE, RMSE) a měření lineárního vztahu (koeficient determinace) vůči DTM 2012, spolu s analýzou procentuální redukce datové sady. Jako doplnění uvedených

kvantitativních metod byly využity 3D vizualizace výšky vegetace, které umožňovaly detailnější posouzení variability, jež bylo klíčové pro prostorovou interpolaci.

Pro datové sady získané z misí GEDI a ICESat-2 byla filtrace provedena samostatně. Na začátku byla provedena filtrace hustších dat z mise GEDI pomocí dvou různých metod filtrace. První, naivní metoda, vycházela z využití atributu *quality_flag*, který je v produktu dostupný pro označení kvalitních měření, přičemž cílem bylo zachovat pouze měření s hodnotou *quality_flag* = 1. Tato metoda poskytovala základ pro srovnání s pokročilejším druhým přístupem, který se snažil o minimalizaci datové redukce a zajištění vyšší kvality dat. Druhý přístup kombinoval filtraci založenou na dostupných attributech ve výstupním produktu GEDI L2A a využití algoritmů pro detekci odlehlých hodnot. Pro odstranění odlehlých hodnot odchylek *elev_diff* bylo použito pravidlo 1,5 IQR (mezikvartilové rozpětí). Tzv. Tukeyovo pravidlo říká, že odlehlá hodnota je ta, která je od prvního, resp. třetího kvartilu vzdálená víc než o $1.5 \cdot \text{IQR}$. Dále bylo pro filtraci využito atributů, které s odchylkami *elev_diff* v absolutní hodnotě nejvíce korelovaly. Největší negativní korelace byla zaznamenána s atributem *surface_flag* a *rx_assess_qf* (viz Obrázek 24). Atribut *surface_flag* vykazoval největší korelaci také s atributem *quality_flag*, avšak oproti němu obsahoval výrazně menší počet měření, označených jako nekvalitní. Zachovány byly data s hodnotami *surface_flag*=1 a *rx_assess_qf*=1.



Obrázek 24: Vizualizace odchylek měření výšky terénu od referenčního DTM 2012 (vlevo) a graf korelační matice s proměnnou *elev_diff* v absolutní hodnotě (vpravo)

Dalším krokem bylo odstranit velké odchylky každého měření výšky vegetace *RH95* pomocí algoritmu k-nejbližších sousedů (k-NN). Iterativně se tak pro každý datový bod v datové sadě vypočítal medián z hodnot *RH95* jeho k-nejbližších sousedů

a hodnota v daném bodě byla s tímto mediánem porovnána, přičemž byl určen práh povolené odchylky. Určení počtu sousedů a maximální povolené odchylky bylo nutné průběžně optimalizovat na základě ověřování chyb od DTM 2012. S nejmenší chybou vyšla filtrace při $k = 6$ s maximální povolenou odchylkou od mediánu výšky vegetace sousedů 30 m. Posledním krokem bylo odstranění hodnot s výškou vegetace *RH95* nad 40 m. Jelikož měření GEDI představovalo agregaci výšek stromů v kruhu o průměru 25 m, bylo pokládáno za málo pravděpodobné, že by se v takovýchto průměrech na území KRNAPu vyskytovala výška vegetace nad 40 m. Výsledná filtrace zahrnující kombinaci zmíněných kroků byla porovnána s naivní filtrací pomocí *quality_flag*. Tím bylo zjištěno, že pokročilejší filtrace dat vykazovala menší chybu od DTM, jelikož data vyfiltrovaná pomocí *quality_flag* stále obsahovala určitý podíl chybných měření.

Podobně bylo postupováno i v případě filtrace dat z ICESat-2. Nejprve byla odstraněna měření s hodnotami odchylek *elev_diff* menšími než -5 a většími než 1,5 IQR nad třetím kvantilem. Největší korelace *elev_diff* byla zaznamenána s atributem *h_te_uncertainty*, který však obsahoval 973 chybných hodnot a z tohoto důvodu nebyl využit pro filtraci. Dalším krokem byla detekce odchylek *h_canopy* od mediánu *RH95* okolních vyfiltrovaných datových bodů z GEDI. Záměrem bylo snížit variabilitu ve výsledné datové sadě, která vznikla spojením dat z obou misí. Optimální hodnota počtu sousedů byla zvolena jako $k = 10$ s maximální povolenou odchylkou *h_canopy* od mediánu hodnot *RH95* sousedů z GEDI = 7 m. Tímto algoritmem se tak snižovaly rozdíly ve výšce vegetace, které do dat GEDI vnesl ICESat-2. Nakonec byla zachována pouze měření s *h_canopy* < 40 m. Posledním krokem bylo spojení pokročile vyfiltrovaných dat z GEDI a ICESat-2 do nové datové sady, která byla dále předzpracována pro analýzu prostorové struktury a korelace. Pro účely pozdějšího porovnání byla kromě této datové sady vyexportována ještě datová sada obsahující pouze data z GEDI s naivní filtrací podle *quality_flag*.

3.3.3 Předzpracování prediktorů

Pro zpřesnění výsledného modelu ordinary krigingu bylo dále využito metody univerzálního krigingu, který modeloval deterministickou složku výšky vegetace v datech pomocí externích prediktorů jako vysvětlujících proměnných. Tyto prediktory byly vybrány ze satelitu Sentinel-1 (S1), radaru se syntetickou aperturou (SAR), a z optických multispektrálních snímků satelitu Sentinel-2 (S2).

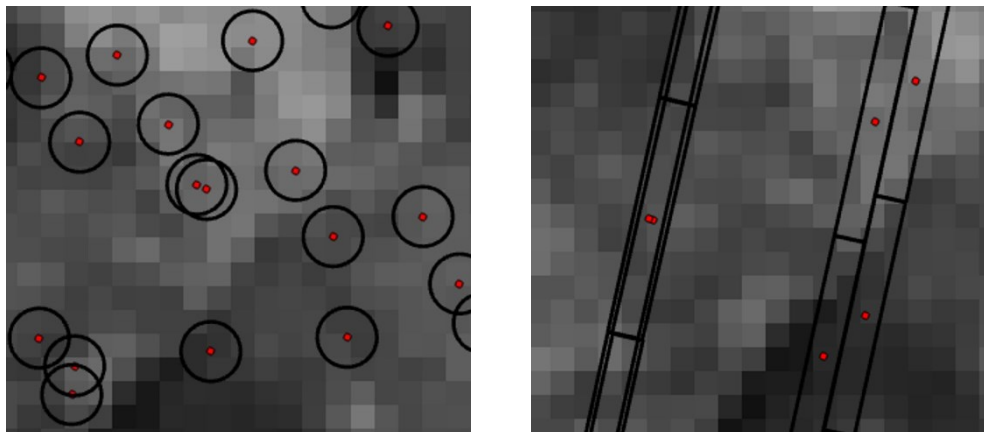
Data byla pořízena a předzpracována pomocí platformy GEE. Z optických dat Sentinel-2 Level 2A byla využita odrazivost v jednotlivých spektrálních pásmech a také odvozené základní vegetační indexy. Byla sestavena spojitá mozaika pokrývající celé území KRNAPu v období 2019 až 2022, přičemž bylo celkem využito 656 snímků. Data byla standartně předzpracována, lokálně byly odmaskovány pixely obsahující oblačnost, sníh apod. Radiometrické korekce byly ponechány v původním stavu ze Sentinelu-2 L2A. Data obsahovala medián a směrodatnou odchylku z vegetační sezóny pro jednotlivá spektrální pásma a základní vegetační indexy. Mezi tyto vegetační indexy patřilo NDVI (Normalized Difference Vegetation Index), NDMI (Normalized Difference Moisture Index) a EVI (Enhanced Vegetation Index). Směrodatná odchylka měla poskytnout odlišení zemědělsky využívaných ploch od stabilních lesních porostů. Ze Sentinelu-2 bylo tedy použito dohromady 26 prediktorů.

Prediktory ze Sentinelu-1 zahrnovaly vrstvy útlumu (v dB) pro polarizace VV (vertikální/vertikální) a VH (vertikální/horizontální), agregované zvláště pro letní období a pro celý rok. Jelikož byl pro každý prediktor opět spočítán medián a směrodatná odchylka, bylo tak celkem ze Sentinelu-1 použito 8 prediktorů. K prediktorům ze S1 a S2 byly přidány odchylky měření terénu z GEDI a ICESat-2 od referenčního DTM 2012 (*elev_diff*), takže bylo pro universal kriging dohromady použito 35 prediktorů.

Veškeré předzpracování rastrů probíhalo v R pomocí balíčku *terra*. Rastrové prediktory bylo potřeba sjednotit tak, aby každý z nich měl souřadnicový systém S-JTSK (EPSG:5514), rozsah území o čtverci 10x10 km a prostorové rozlišení 10 m. Dále bylo nutné zajistit, aby se jednotlivé buňky rastrů přesně překrývaly (parametr *snap* u funkce *align*). Prediktory *EVI_stdDev* a *elev_diff* bylo nutné dále prostorově interpolovat, jelikož obsahovaly prázdné pixely, na jejichž místech by při procesu krigingu nemohlo dojít k odhadu neznámých hodnot. Pro interpolaci těchto prázdných pixelů byla vybrána metoda IDW z balíčku *gstat*. Hodnota koeficientu mocniny α byla stanovena na 2, aby se zvýšil vliv okolních bodů.

Následně bylo nutné z rastrových prediktorů (viz jejich kolinearita v Příloha 11) extrahovat hodnoty pro již vyfiltrované datové body z GEDI a ICESat-2. Extrakce byla prováděna jako průměrová agregace hodnot pixelů spadajících na plochu původních tvarů stop (footprintů) obou satelitních lidarů, po vzoru studie (Liu et al., 2021). Pro GEDI byly extrahovány hodnoty z prediktorů v obalové zóně o průměru 25

m a pro ICESat-2 byly vytvořeny orientované obdélníky o rozměrech $\sim 13 \times 100$ m, reprezentující jednotlivé segmenty s měřicími body ICESat-2 umístěnými v jejich středu. Segmenty byly orientované vždy ve směru pozemní dráhy tak, aby bylo zajištěno co nejpřesnější pokrytí celé plochy, kterou jednotlivá měření reprezentují. Ze všech 35 rastrových prediktorů byly tak iterativně extrahovány hodnoty do každého z datových bodů GEDI a ICESat-2 s respektem k ploše, kterou pokrývaly jejich stopy (viz Obrázek 25).



Obrázek 25: Extrakce hodnot prediktorů v původním tvaru stop GEDI (vlevo) a segmentech ICESat-2 (vpravo)

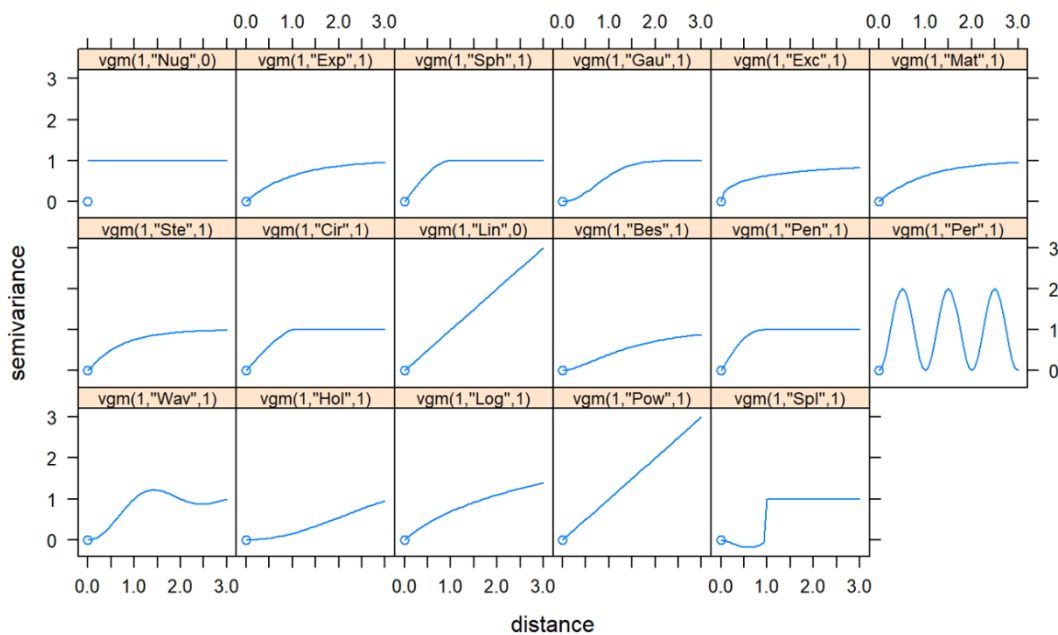
3.4 Geostatistické modelování

3.4.1 Analýza prostorové struktury a korelace

Po provedení filtrace a minimalizaci chyb v měření terénu od referenčního DTM 2012 z ALS byly datové sady z obou satelitních misí spojeny a následně použity pro analýzu prostorové autokorelace předcházející krigingu. Autokorelace v datech vyjadřuje vztah mezi hodnotami téže veličiny, v tomto případě výšky vegetace, což reflektuje jejich vzájemnou podobnost na základě vzdálenosti mezi nimi. Tento vztah na různých vzdálenostech byl analyzován pomocí variogramu. Analýza prostorové autokorelace a realizace krigingu byla provedena pomocí funkcí z balíčku *gstat* v R. Nejprve byl sestaven empirický variogram, který mohl být dále vizuálně posouzen. Tento variogram vykazoval významnou autokorelaci v datech, což bylo způsobeno výrazným stoupáním semivariance do vzdálenosti ~ 1 km. Tato vzdálenost tedy označovala přibližný dosah autokorelace. V datech se také vyskytovala určitá míra mikrovariability, která však nepřesahovala zbytkový rozsah semivariability autokorelované části variogramu. Variogram byl sestaven jak pro datovou sadu z GEDI s jednoduchou filtrací podle *quality_flag*, tak i pro data s pokročilou filtrací, kombinující data GEDI s ICESat-2. U dat s pokročilou filtrací byla pozorována

mnohem menší mikrovariabilita, protože byla efektivněji odstraněna chybná měření, což vedlo k vyšší kvalitě dat. Dále byla ověřena anizotropie v datech pomocí parametru α funkce *variogram()*, který umožňuje konstrukci variogramu do různých směrů. Anizotropie byla zjištěna pouze na větších vzdálenostech, kde docházelo ke změnám v průběhu variogramu. Otázkou bylo, do jaké míry se na těchto vzdálenostech projevuje zbytkový vliv autokorelace, nebo zda se jedná již o nekorelovanou náhodnou variabilitu. Pro snížení výpočetní náročnosti ordinary krigingu bylo dále rozhodnuto využít skutečnosti, že je vliv prostorové autokorelace prokazatelně patrný do vzdálenosti 1 km, k omezení oblasti bodů, zahrnutých do výpočtu vah krigingu (search neighborhood). Regionálním omezením krigingu také odpadla potřeba modelovat anizotropii, která se na krátkých vzdálenostech nevyskytovala.

Vnitřní stacionarita hodnot výšky vegetace byla později řešena volbou metody univerzálního krigingu, která zahrnovala modelování globálního trendu v datech. Empirický variogram, který byl sestaven z variogram mraku, byl poté aproximován teoretickým variogramem pomocí funkce *fit.variogram()*. Balíček *gstat* nabízí sadu variogramových modelů *vgm()* (viz Obrázek 26), které mají standardní parametry *range* (dosah autokorelace), *psill* (rozsah semivariance autokorelovaného náhodného procesu bez nugget efektu), *nugget* (rozsah mikrovariability).



Obrázek 26: Ukázka *vgm* modelů variogramu z balíčku *gstat* pomocí funkce *show.vgms()*

V této práci se nejvíce osvědčila efektivita modelů umožňujících rozšířenou parametrizaci pomocí vyhlazovacího parametru *kappa*. Optimalizace těchto parametrů

probíhala při aproximaci empirického variogramu metodou (obyčejných) nejmenších čtverců. Nejvíce se průběhu autokorelace přiblížil stabilní model exponenciální třídy s parametrem $\kappa = 0,5$. Tento model byl efektivní hlavně na kratších vzdálenostech do ~ 1 km, kde zahrnoval část mikrovariability, čímž se výsledný interpolovaný povrch při krigingu více blížil měřeným bodům (jevil se ostřejší). Dalším vhodným modelem byl Matérnův model, který obsahoval menší optimální hodnotu vyhlazovacího parametru $\kappa = 0,3$. Tento model měl na kratších vzdálenostech hladší průběh, proto byl spolu s ním modelován větší nugget efekt než u modelu exponenciální třídy s prudkým průběhem v počátku. Výsledný interpolovaný rastr CHM pak vznikl hladší s menším detailem, který by však mohl být pokládán za náhodnou chybu nebo šum.

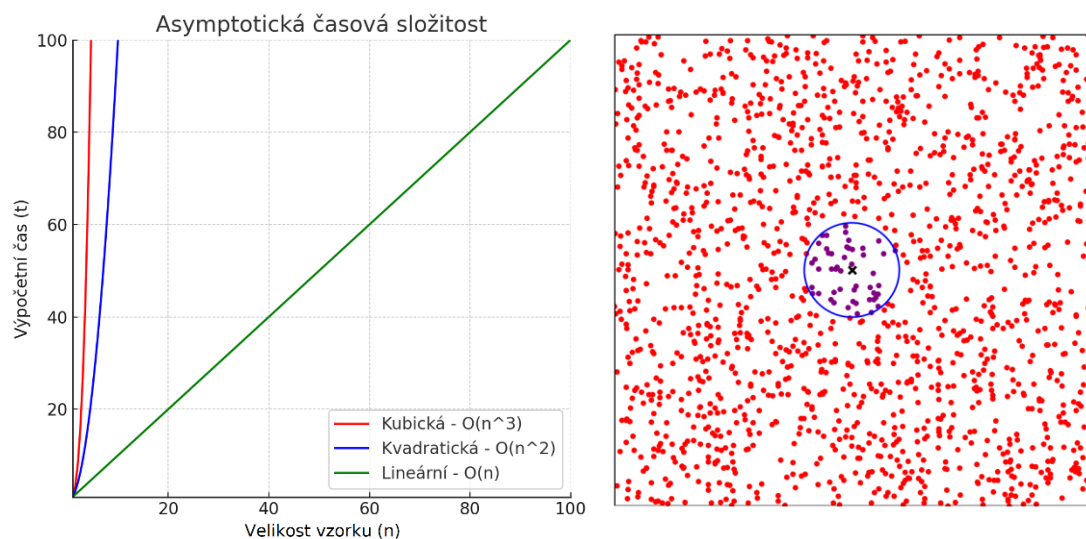
3.4.2 Ordinary kriging

Po analýze prostorové struktury a autokorelace měření výšky vegetace z dat GEDI a ICESat-2 byl jako nástroj prostorové interpolace pro tvorbu CHM první použit základní typ krigingu, tzv. ordinary kriging (OK). Tento typ krigingu předpokládá neznámou střední hodnotu konstantní v celém prostoru. Hlavním zdrojem informací o této metodě bylo dílo J. Ježka (2015) Geostatistika a prostorová interpolace.

Jelikož je kriging výpočetně náročnou metodou prostorové interpolace, prioritou bylo minimalizovat tuto výpočetní náročnost algoritmu krigingu tak, aby se snížil čas průběhu této operace. Funkce *krige()* v balíčku *gstat* pro jednotlivé kroky realizace krigingu využívá další funkce, převážně rozšiřujících knihoven DLL ve složce interpretačního programu jazyka R, a částečně běží na pozadí mimo relaci prostředí RStudio. Pro zrychlení algebraických výpočtů v rámci krigingu tedy mohlo být využito například změny základních knihoven pro lineární algebru, za knihovny podporující paralelní výpočet na více jádrech procesoru. S knihovnou Intel MKL (Math Kernel Library) obsaženou v R verzi Microsoft R Open 4.3.2 trval proces OK přibližně poloviční čas. Měření času výpočtu na menších vzorcích území a testování různých parametrů nebo balíčků pro kriging bylo v této fázi nezbytné pro ověření proveditelnosti tohoto typu prostorové interpolace pro celé zájmové území. Sérií testů byl zjištěn vztah výpočetního času krigingu s narůstajícím počtem vstupních bodů, tzv. asymptotická časová složitost algoritmu. Vstupní body pro kriging byly složeny z datových bodů se známou hodnotou, ze kterých se počítaly krigingové váhy, a pro jejichž počet výpočetní čas rostl kvadraticky a z bodů s neznámou hodnotou, pro

jejichž počet rostl výpočetní čas lineárně. Podle asymptotické časové složitosti tak výpočetní čas s počtem všech vstupních bodů rostl kubicky, tj. v řádu $O(n^3)$.

OK byl proveden zvláště pro datovou sadu naivně vyfiltrovaných měření pomocí *quality_flag* a dále pro datovou sadu využívající pokročilou filtraci (viz kapitola 3.4.1). V obou případech byla použit model krigingu, kde neznámé hodnoty představoval rastr třídy *stars* ze stejnojmenného balíčku v R. Tento rastr pokrýval zájmové území 10x10 km se 100 m přesahem na každé straně kvůli zmírnění okrajového efektu krigingu. Při prostorovém rozlišení 10 m se tak jednalo o 1 040 400 buněk s neznámými hodnotami určenými pro odhad. V datové sadě pokročile vyfiltrovaných měření se zároveň nacházelo 35 702 datových bodů známých hodnot. Podle asymptotické časové složitosti by výpočetní čas krigingu v této situaci trval přibližně 4,5 měsíce. Pro realizaci bylo tak nutné určit optimální prostorové omezení výpočtu krigingu. Při odhadu každé buňky byl využit kruhový výběrový region s průměrem 1 km, určeným dle dosahu prostorové autokorelace z variogramu (viz kapitola 3.4.1). Velikost tohoto regionu určoval parametr *maxdist* u funkce *krige()*, kde byl navíc výpočet krigingových vah omezen jen na 100 nejbližších sousedů pomocí parametru *nmax*. Optimální hodnota pro *nmax* byla volena na základě minimalizace chyb křížové validace a výpočetního času, který s počtem sousedů rostl kubicky. Po omezení na 100 nejbližších sousedů trvalo OK přibližně 12 hodin. Graf asymptotické časové složitosti a regionální omezení krigingu jsou vidět na Obrázek 27. V regionu omezení (vpravo) jsou tmavší body sousedy odhadovaného bodu, jejichž prostorová struktura je do výpočtu odhadu zahrnuta.



Obrázek 27: Znázornění asymptotické časové složitosti (vlevo) a ilustrace 1 km kruhového regionu pro výpočet vah krigingu (vpravo)

Stručně řečeno byly pro OK použity parametry funkce *krige()*: *formula* (označující předpis funkce korelace proměnné výšky vegetace *ch* se sebou samou: $ch \sim I$), *data* (objekt třídy SF obsahující data známých hodnot), *newdata* (rastr třídy stars s prázdnými buňkami pro interpolaci), *model* (vgm objekt teoretického variogramu), *maxdist*, *nmax* (pro omezení rozsahu hledání a kubické snížení výpočetní náročnosti) a *debug.level=-1* (pro zobrazení průběhu krigingu v %). Výsledkem byl vždy rastr odhadů a rastr rozptylu krigingu pro určení nejistoty spojené s odhady. OK model byl zároveň vždy podroben křížové validaci (viz kapitola 3.5), při které bylo například zjištěno, že mají odhady OK nejmenší RMSE, MAE a ME v případě použití modelu variogramu stabilní exponenciální třídy s parametrem $kappa = 0,5$, po omezení na okolí 100 nejbližších sousedů. Při zvýšení počtu sousedů na 200 se dokonce chyby zvětšily. OK také vyšlo lépe pro datovou sadu s pokročilou filtrací. Nakonec byla provedena validace modelu s CHM z ALS (2012), kde byly opět kvantifikovány chyby a koeficient determinace pro lineární vztah R^2 .

3.4.3 Universal kriging

Dalším krokem pro zpřesnění výsledného CHM bylo použití pokročilejšího typu krigingu, tzv. universal krigingu (UK). UK využíval deterministické modelování globálního trendu (driftu) v datech a nepředpokládal tak konstantní střední hodnotu výšky vegetace v prostoru (viz kapitola 2.4.4). Globální trend v datech může popisovat určitý environmentální prediktor, například rastr spojených hodnot, který má nějaký vztah s výškou vegetace a mohl by tak UK model podporovat a zpřesňovat na místech spojených s velkou nejistotou. Velká nejistota se vyskytovala především na místech, nepokrytých orbity přístroje GEDI nebo satelitu ICESat-2. UK model ve své podstatě spojuje modelování deterministické složky, jako trendu střední hodnoty výšky vegetace v prostoru, s modelováním zbytkové náhodné autokorelované složky pomocí variogramu. Při tvorbě variogramu pomocí funkce *variogram()* v *gstat* bylo tak potřeba použít stejný předpis modelu jako u krigingu ve funkci *krige()*. V tomto předpisu se definoval vztah zápisem ve formě: $Y \sim X_1 + \dots + X_n$, kde Y je vysvětlovaná proměnná (výška vegetace) a X_n jsou hodnoty externích prediktorů jako vysvětlující proměnné. Předpis vlastní autokorelace náhodné veličiny by byl značen jako $Y \sim I$ (viz kapitola 3.4.2). Globální trend výšky vegetace reprezentovaly v této práci především prediktory ze S1 a S2, rozebírané v kapitole 3.3.3. Jednalo se tak v podstatě o nezávislé proměnné vysvětlující závislou proměnnou výšky vegetace *ch*.

Nejprve však byla provedena jednodušší verze modelu UK, která jako jedinou vysvětlující proměnnou využívala odchylky výšky terénu od DTM 2012 (*elev_diff*). Parametrizace tohoto modelu ve funkci *krige()* probíhala obdobně jako u OK modelu s datovou sadou pokročilé filtrace. Pro parametr *formula* byl akorát použit předpis $ch \sim elev_diff$ a parametr *newdata* obsahoval příslušný rastr třídy *stars* s odchylkami *elev_diff* pro celou oblast. Chyby výsledného modelu při křížové validaci a validaci pomocí CHM z ALS byly obecně menší než u OK modelu. Menší byl také rozptyl krigingu, což potvrzovalo, že zavedení prediktorů odhady krigingu zpřesňovalo. Dalším krokem bylo tedy provést UK se všemi 34 prediktory ze S1 a S2.

Do bodů měření výšky vegetace, uvnitř datové sady pokročilé filtrace, byly extrahovány hodnoty prediktorů pomocí původních tvarů stop zdrojového laserového altimetru pomocí funkce *extract()* z balíčku *terra*. Tyto nově vzniklé proměnné u měřených bodů byly v parametru *formula* využity pro tvar předpisu UK modelu $Y \sim X_1 + \dots + X_n$ jako vysvětlující proměnné X_n výšky vegetace *ch*. Měřené body byly vloženy do parametru *data* jako SF objekt. Bylo nezbytné, aby u UK modelu objekt v parametru *newdata* zahrnoval hodnoty všech 34 použitých vysvětlujících proměnných ve formě spojitých povrchů, pokrývajících interpolované území. Pro zjednodušení procesu spojování byly tak všechny rastrové prediktory převedeny na bodový SF objekt. Jeden bod neznámé hodnoty výšky vegetace, určený pro interpolaci, obsahoval tedy 34 proměnných (sloupců), které reprezentovaly hodnoty jednotlivých rastrových prediktorů ze S1 a S2. V tomto případě se jednalo o bodový kriging, jehož výstupem byly odhady a nejistoty umístěné v bodovém poli. V parametru *model* se nacházel příslušný předem vypočtený teoretický variogram *vgm()*, vytvořený z empirického variogramu se stejným předpisem, jaký byl použit v parametru *formula*. Vymezení kruhového regionu pro výpočet vah krigingu okolo interpolovaného bodu bylo dále provedeno pomocí parametrů *maxdist* a *nmax*, stejně jako u předešlých modelů. Tímto omezením však v tomto modelu využívajícím všech 34 prediktorů z S1 a S2 došlo k značnému vychýlení odhadovaných výšek vegetace u výsledného CHM.

Modelování nugget efektu akceptuje existenci mikrovariability v datech, což má za následek, že interpolovaný povrch metodou kriging nemusí přesně odpovídat měřeným bodům, ale může se kolem nich pouze pohybovat. V případě deterministického modelování polynomického trendu pro střední hodnotu výšky vegetace s využitím externích prediktorů, jako je tomu například u UK modelu, může

dojít k předpovídání hodnot mimo rozsah hodnot měřených bodů. Aplikace krigingu na rezidua tohoto modelu může pak vést k negativním odhadům výšky vegetace. Omezení výpočetního rozsahu krigingu na region s průměrem 1 km tedy zvýraznilo tento jev a výsledný CHM obsahoval hodnoty od negativních až po hodnoty přesahující 50 m. Příčinou bylo pravděpodobně nerovnoměrné rozdělení nebo výskyt extrémních hodnot v prediktorech S1 a S2. Jedním z možných řešení byla logaritmická nebo Box-Coxova transformace těchto prediktorů, aby se dosáhlo normálního rozdělení, případně odstranění extrémních hodnot nebo jiná úprava škál prediktorů. Další možností bylo provádět odhady krigingu s využitím všech známých bodů v datovém souboru, čímž by se však kubicky zvýšila doba výpočtu. Nejjednodušším řešením bylo tedy vytvoření CHM s prostorovým rozlišením 30 m místo původních 10 m, což výrazně snížilo počet vstupních bodů potřebných pro interpolaci. Rovněž bylo rozhodnuto zmenšit interpolovanou oblast na původní zájmové území 10x10 km bez přesahu 100 m, kterým byla tato oblast původně rozšířena, aby se předešlo okrajovému efektu. Společně s využitím výkonného počítače bylo nakonec dosaženo výrazného snížení výpočetního času, což UK učinilo proveditelným bez nutnosti omezení okolí.

3.5 Postup validace modelu

Tato kapitola se věnuje metodám validace, které byly použity při vytváření modelů výšky vegetace v rámci této diplomové práce. Součástí je rovněž příprava na porovnání výsledků s globálními CHM (Lang et al., 2022a; Potapov et al., 2021).

Výstupem funkce *krige()* z balíčku *gstat*, používané pro OK a UK, byla nejen vrstva odhadů pro všechny buňky nebo body v zájmovém území, ale také vrstva nejistoty spojené s těmito odhady. To bylo také jednou z hlavních motivací pro výběr krigingu jako statistické metody prostorové interpolace. Tato nejistota, vyjádřená rozptylem krigingu, poskytovala základní přehled o spolehlivosti odhadů v modelu. Rozptyl mohl být následně odmocněn pro zjištění směrodatné chyby v jednotkách měřené veličiny. V podstatě tato spojitá rastrová vrstva poskytovala informace o rozsahu intervalu spolehlivosti, ve kterém se podle modelu mohly nacházet reálné hodnoty výšky vegetace. Tyto aspekty jsou úzce spojeny s principy Gaussových procesů. Minimalizace krigingové nejistoty mohla být dosažena buď zvýšením hustoty dat, nebo snížením nekorelované náhodné variability, tj. mikrovariability, například odstraněním chybných měření pomocí vhodné metody filtrace.

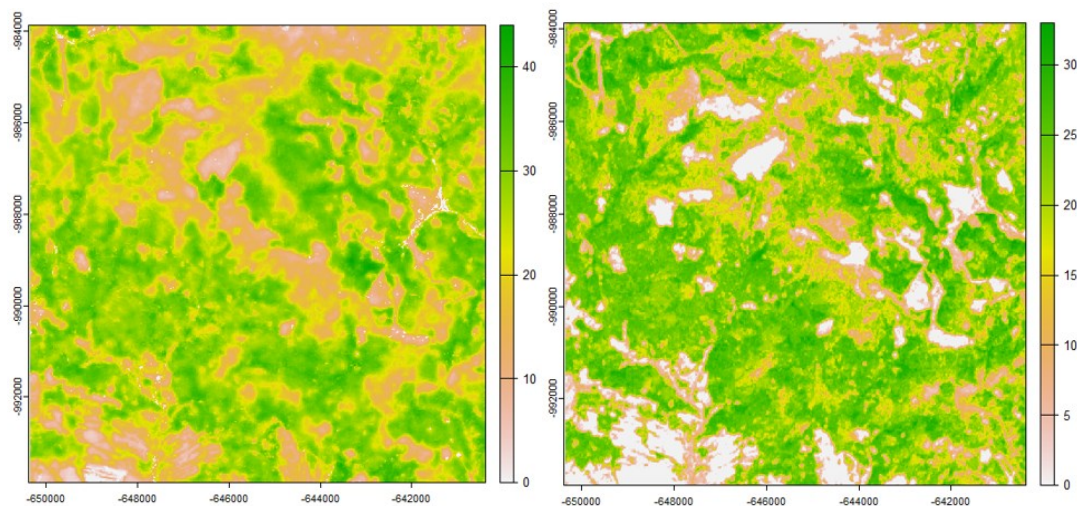
Paralelně se zpracováním krigingových modelů byla prováděna křížová validace pomocí funkce *krige.cv()*, která je rovněž součástí balíčku *gstat*. Tato funkce implementuje princip leave-one-out křížové validace, přičemž byly aplikovány identické parametry jako u funkce *krige()*, s výjimkou absence parametru *newdata*. Při této validaci byl krigingový model trénován na souboru měřených bodů s vynecháním jednoho bodu, který byl následně využit jako testovací data. Odhadnutá hodnota pro vynechaný bod byla porovnána s jeho původní hodnotou, čímž byla určena odchylka. Tento postup se opakoval pro každý bod v datové sadě, což u velkých souborů dat představovalo příliš výpočetně náročnou metodu validace. Nicméně, u většiny krigingových modelů byla realizace možná díky omezení oblasti pro výpočet vah k odhadu pro každý bod. Došlo také k lineárnímu snížení výpočetního času, neboť se nevyžadoval odhad hodnot pro celý rastr nebo bodové pole s milionem bodů pro území 10x10 km. Odhad se prováděl vždy n krát pro *jeden* neznámý bod z $(n - 1)$ známých bodů v datovém souboru. V datové sadě s pokročilou filtrací dat z GEDI a ICESat-2 se například nacházelo 35 702 známých bodů a díky omezení krigingu na 100 nejbližších sousedů v oblasti o poloměru 1 km byl každý odhad proveden rychle. Funkce *krige.cv()* mohla být dále omezena parametrem *nfold* na menší počet bodů než kolik obsahuje celá datová sada. Dojde tím však ke snížení přesnosti odhadu chyby. V případě UK modelu s využitím všech prediktorů ze satelitů S1 a S2, kde nebylo možné aplikovat omezení krigingového okolí, musel být *nfold* nastaven na 10 000 bodů pro zajištění výpočetní efektivity operace. Z dat křížové validace byly dále odvozeny chybové ukazatele RMSE, MAE a ME, které posoudily přesnost odhadů modelu.

Po vytvoření interpolovaného povrchu s využitím jedné z metod krigingu byla dále provedena validace pomocí CHM 2012 z ALS. Tento model výšky vegetace, jehož tvorba je popsána v článku Gábor et al. (2019), byl pořízen laserovým leteckým skenováním na území KRNAPu v roce 2012. Společně s parametry lidarového přístroje RIEGL LMS Q-680i bylo ve výsledném bodovém mračnu dosaženo hustoty ~ 6 bodů/m². Po automatické klasifikaci bodového mračnu v programu LAStools byl z bodů terénu vygenerován digitální model terénu (DTM 2012) a společně s tím byl i standartními postupy odvozen model výšky vegetace (CHM 2012). Původně byly oba modely vytvořeny pro KRNAP v prostorovém rozlišení 0,5 m. (Gábor et al., 2019).

Pro účely této práce byly validační modely převedeny na rozlišení 10 m nebo 30 m, v případě potřeby. DTM 2012 sloužilo pro validaci terénu při procesu filtrace a

CHM 2012 pro validaci výsledných modelů krigingu. Pokud se jednalo o bodový kriging, převedlo se výsledné bodové pole na rastr kvůli snadnější vizualizaci vrstvy odhadů (výsledný CHM) a nejistot. V rastrové podobě byl pak výsledný CHM oříznut o přesah 100 m, který byl do území zaveden kvůli zmenšení vlivu okrajového efektu, a dále byl posunut tak, aby se buňky rastrů přesně překrývaly s validačním CHM 2012. Po přesné kalibraci a sjednocení obou modelů mohl být výsledný CHM převeden na body, do kterých byly extrahována hodnoty z rastru validačního CHM 2012. Vlastní hodnoty výšky vegetace modelovaného CHM a hodnoty validačního CHM bylo pak možné vynést do grafu lineárního vztahu pomocí balíčku *ggplot2*. Dále byl vypočítán koeficient determinace R^2 a míry chyb ME, MAE, RMSE od validačního CHM.

Do vzniklého bodového objektu mohly být dále extrahovány hodnoty z globálních modelů výšky vegetace (Lang et al., 2022a; Potapov et al., 2021). Všechny rastry musely být nejprve sjednoceny na prostorové rozlišení a souřadnicový systém výsledného CHM. Dále musel být opět zajištěn přesný překryv buněk rastrů, a nakonec byla funkcí *extract()* provedena extrakce hodnot. Skrze kvantifikaci chyb a vizualizaci grafu *ggplot* bylo s modely Lang et al. (2022a) a Potapov et al. (2021) provedeno porovnání, aby se zjistilo, zda jsou výsledné modely v této práci přesnější. Následující Obrázek 28 zobrazuje oba globální modely, nacházející se na vybraném čtvercovém území o rozměrech 10x10 km v oblasti KRNAPu.



Obrázek 28: Globální CHM vlevo podle studie Lang et. al (2022) a vpravo podle studie Potapov et. al (2021).

4. Výsledky

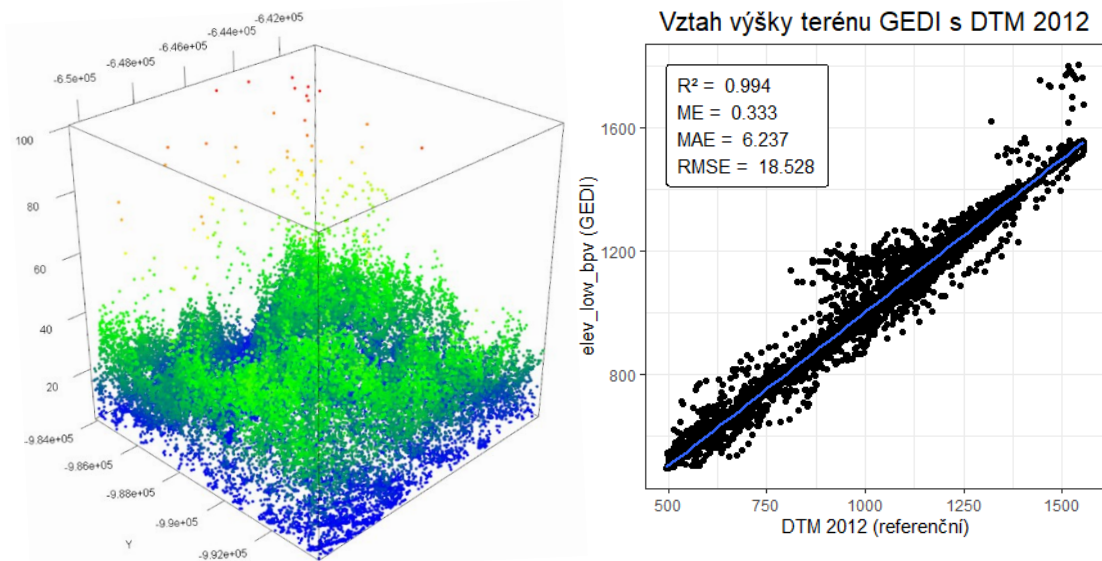
4.1 Výsledky filtrace

Přesnost měření u GEDI a ICESat-2 byla ovlivněna mnoha externími faktory, jako byly podmínky na oběžné dráze, sluneční záření nebo oblačnost v atmosféře. Použité postupy filtrace jsou v této práci podrobně popsány v kapitole 3.3.2. V průběhu byla snaha udržet v datovém souboru co nejvíce měření a minimalizovat přitom odchylky od referenčního DTM 2012. Hodnoty výšky vegetace byly totiž závislé na správném a přesném určení nadmořské výšky terénu (viz kapitola 2.2.1). Při kontrole kvality provedené filtrace byly tak vždy z odchylek od DTM 2012 kvantifikovány chyby ME, MAE, RMSE a určeno bylo R^2 , vyjadřující míru vysvětlené variability. Dále bylo ověřeno, kolik procent z původních dat zbylo.

Nejprve filtrace probíhala pro data z mise GEDI, do kterých byla později přidána vyfiltrovaná data z mise ICESat-2 pro docílení větší hustoty ve výsledné datové sadě. Původně se na vybraném území 10x10 km v KRNAPu za období 2018-2013 vyskytovalo **81 441 měření z GEDI L2A** a **3 390 měření z ICESat-2 ATL08**. Prvním krokem bylo odstranění všech měření, která prokazatelně neobsahovala hodnoty výšky vegetace. Minimálním filtrem pro data GEDI bylo odstranění hodnot splňujících *num_detectedmodes* = 0. Měření s tímto atributem nezaznamenaly žádný odraz. Po tomto filtru **zbylo 59,27 %** z původních dat. Pomocí odchylek od terénu bylo zjištěno, že měřený terén DTM 2012 výrazně nadhodnocuje (302 m). Chyby od DTM 2012 tedy vyšly: **ME = 302; MAE = 306; RMSE = 788 m** a **$R^2 = 0,02$** .

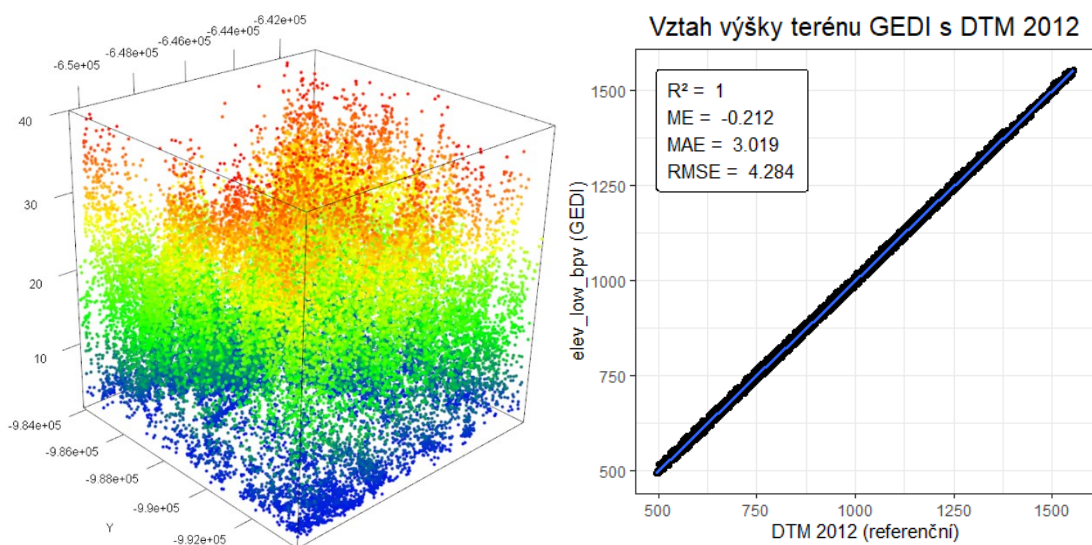
Dalším krokem bylo zvolit přístup, který by snižoval vysoce nepřesná měření a tím i vysokou variabilitu v hodnotách výšky vegetace RH95 u GEDI. Pro tyto účely bylo první volbou odstranit všechna data s *quality_flag*=0. Atribut *quality_flag* označoval nekvalitní měření na základě komplexního posouzení s ním spojených nejistot distributorem. Cílem bylo výsledek této filtrace, dále označované jako naivní, porovnat s pokročilejší formou filtrace. Při naivní filtraci **zbylo přibližně 41,64 %** původních dat GEDI a chyby určení terénu, reprezentovaného validačním DTM 2012, byly: **ME = 0,333; MAE = 6,237; RMSE = 18,528** a **$R^2 = 0,994$** .

Na následujícím Obrázek 29 je vidět 3D model ilustrující variabilitu ve výsledné datové sadě, kde tato filtrace zanechala poměrně odlehle hodnoty a vztah s referenčním terénem.



Obrázek 29: Naivní filtrace GEDI (quality_flag = 1). 3D model výšky vegetace RH95 (vlevo), vztah výšky terénu GEDI s DTM 2012 (vpravo)

Pokročilejší filtrace již byla zaměřena na to, že by měla být výsledkem datová sada, kombinující data z GEDI a ICESat-2, určená pro následnou geostatistickou analýzu a kriging. Základem byla minimálně vyfiltrovaná data z GEDI, která byla neustále podrobována různým filtračním algoritmům, které jsou popsány v kapitole 3.3.2. Filtrační algoritmy byly neustále optimalizovány tak, aby byla výsledná data co nejméně odchýlena od DTM 2012 a zároveň obsahovala co nejvíce měření výšky vegetace. Pro nejvíce optimální verzi filtrace dat GEDI (varianta 7) byly výsledky validace pomocí DTM 2012: **ME = -0,212; MAE = 3,019; RMSE = 4,284** a **R² = 1**. Došlo k **zachování 42 %** původních bodů GEDI, výsledek je vidět na Obrázek 30.



Obrázek 30: Pokročilá filtrace GEDI (varianta 7). 3D model výšky vegetace RH95 (vlevo), vztah výšky terénu GEDI s DTM 2012 (vpravo)

V alternativní variantě pokročilé filtrace dat GEDI byla například zachována pouze měření, která se od referenčního DTM 2012 v absolutní hodnotě neodchylovala více než 20 m. Chyby této varianty filtrace byly při validaci: **ME = -0,367; MAE = 3,324; RMSE = 4,897** a **R² = 1**; přičemž došlo k **zachování 45,85 %** dat z celku.

Následující Tabulka 4 zobrazuje jednotlivé varianty filtrace dat GEDI, které obsahují procento zbytku z původních dat společně s chybou RMSE od referenčních modelů DTM 2012 a CHM 2012 z ALS. Potvrdilo se, že sedmá varianta pokročilé filtrace vyšla nejlépe, proto byla následně použita pro kombinaci s daty ICESat-2.

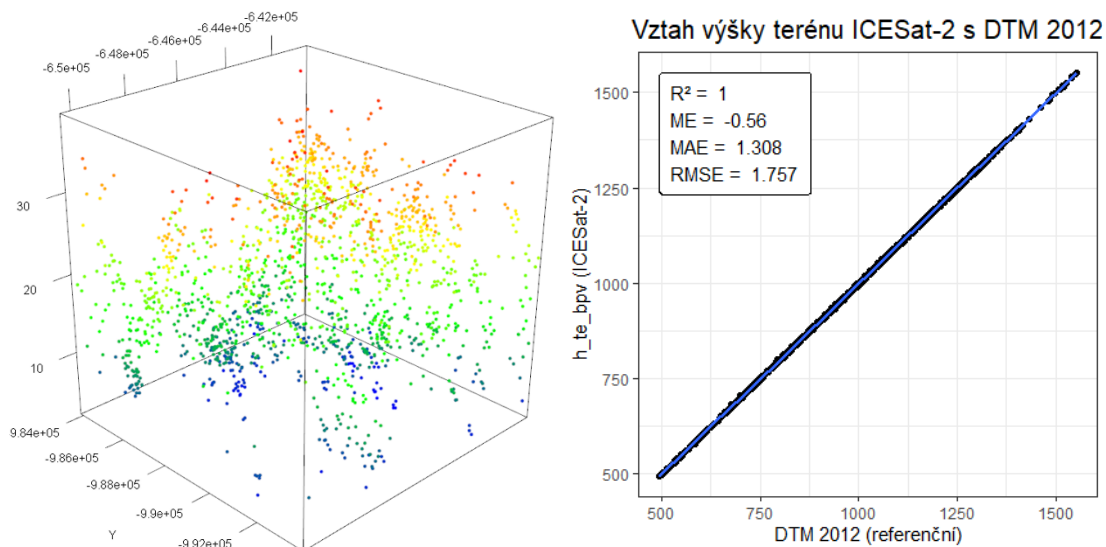
Tabulka 4: Výsledky všech variant filtrací dat z GEDI

Typ filtru	Zbytek (%)	Zbytek od mf (%)	RMSE (DTM 2012)	RMSE (CHM 2012)
Minimální filtr (mf) (num_detectedmodes > 0)	59,27	100	787,93	14,28
Naivní filtrace (qf) (quality_flag = 1)	41,64	70,26	18,52	11,07
Minimální odchylky (dif) abs(odchylky) < 20	45,85	77,35	4,92	11,1
Pokročilá filtrace (f7) (sedmá varianta)	42	70,66	4,28	9,71

Po dokončení pokročilé filtrace dat GEDI bylo potřeba nejprve zajistit správnou minimální filtraci dat ICESat-2. Z celkových 3 390 měřených bodů bylo nakonec **odstraněno 195 bodů** s prokazatelně chybnými hodnotami (**17,38 %**). ICESat-2 terén naopak mírně podhodnocoval, protože chyby při validaci pomocí DTM 2012 vyšly: **ME = -2,65; MAE = 3,46; RMSE = 6,28** a **R² = 0,999**. Přesto však bylo určení terénu pomocí ICESat-2 oproti GEDI velmi přesné.

Dále bylo využito stejné sady algoritmů, jako při filtraci dat GEDI. Jediným rozdílem bylo zaměření na minimalizaci rozdílu výšky vegetace *h_canopy* z ICESat-2 od RH95 z GEDI. Po zkušebním sloučení dat z obou satelitních misí bylo zjištěno, že se odchylka od DTM 2012 s přidáním dat ICESat-2 do GEDI zmenší. Pokud však validace proběhla pomocí CHM 2012 z ALS, chyba výšky vegetace se s přidáním ICESat-2 zvětšila. ICESat-2 na rozdíl od GEDI určoval výšku vegetace méně přesně a chyba **RMSE od CHM 2012** byla **18,75 m**. Zmenšováním rozdílů mezi daty z obou misí společně s validací pomocí dat z ALS bylo tedy docíleno optimálních parametrů všech algoritmů (viz kapitola 3.3.2). Výsledná vyfiltrovaná data z ICESat-2 měla

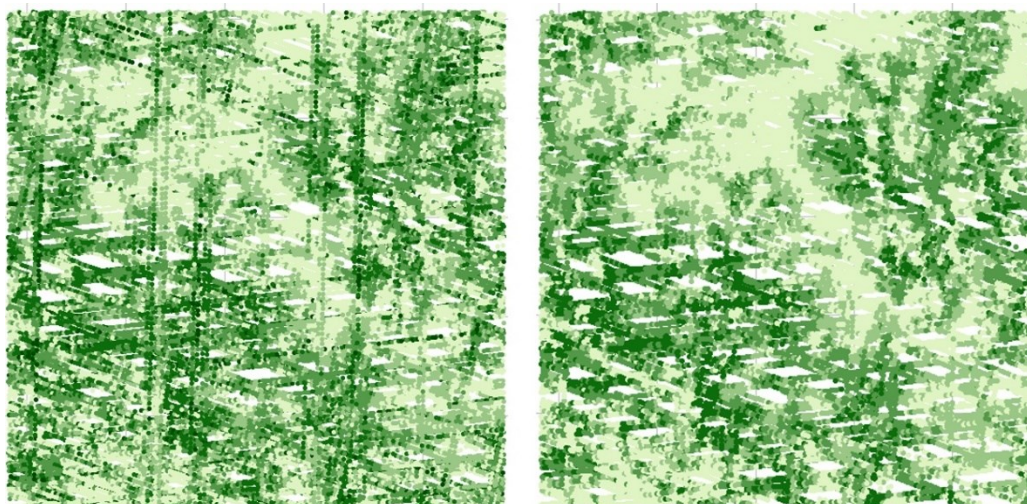
chyby od DTM 2012: **ME = -0,56; MAE = 1,308; RMSE = 1,757** a **R² = 1**. Z celkových dat ICESat-2 **zbylo 47,14 %** a výsledek je vidět na Obrázek 31.



Obrázek 31: Pokročilá filtrace ICESat-2. 3D model výšky vegetace h_{canopy} (vlevo), vztah výšky terénu ICESat-2 s DTM 2012 (vpravo)

Po vytvoření výsledné datové sady kombinací pokročile vyfiltrovaná data z GEDI a ICESat-2, byl chyby od DTM 2012: **ME = -0,228; MAE = 2,942; RMSE = 4,204** a **R² = 1**. Při validaci pomocí CHM 2012 z ALS bylo dosaženo **RMSE = 9,799**, což bylo o ~0,09 m více než u samotných vyfiltrovaných dat z GEDI bez ICESat-2. Kromě této malé nejistoty však ICESat-2 velmi dobře zhustil výsledná data, protože jeho orbity byly na území KRNAPu téměř kolmé k orbitům GEDI na ISS. Vztah výsledné datové sady s referenčním DTM a CHM je umístěn v Příloha 10.

Následující Obrázek 32 porovnává kombinaci výšek vegetace $RH95$ z GEDI a h_{canopy} z ICESat-2 **před** provedením pokročilé filtrace s výslednou datovou sadou **po** aplikování pokročilých filtrací na data z obou misí.



Obrázek 32: Vyška vegetace z GEDI a ICESat-2 před (vlevo) a po pokročilé filtraci (vpravo)

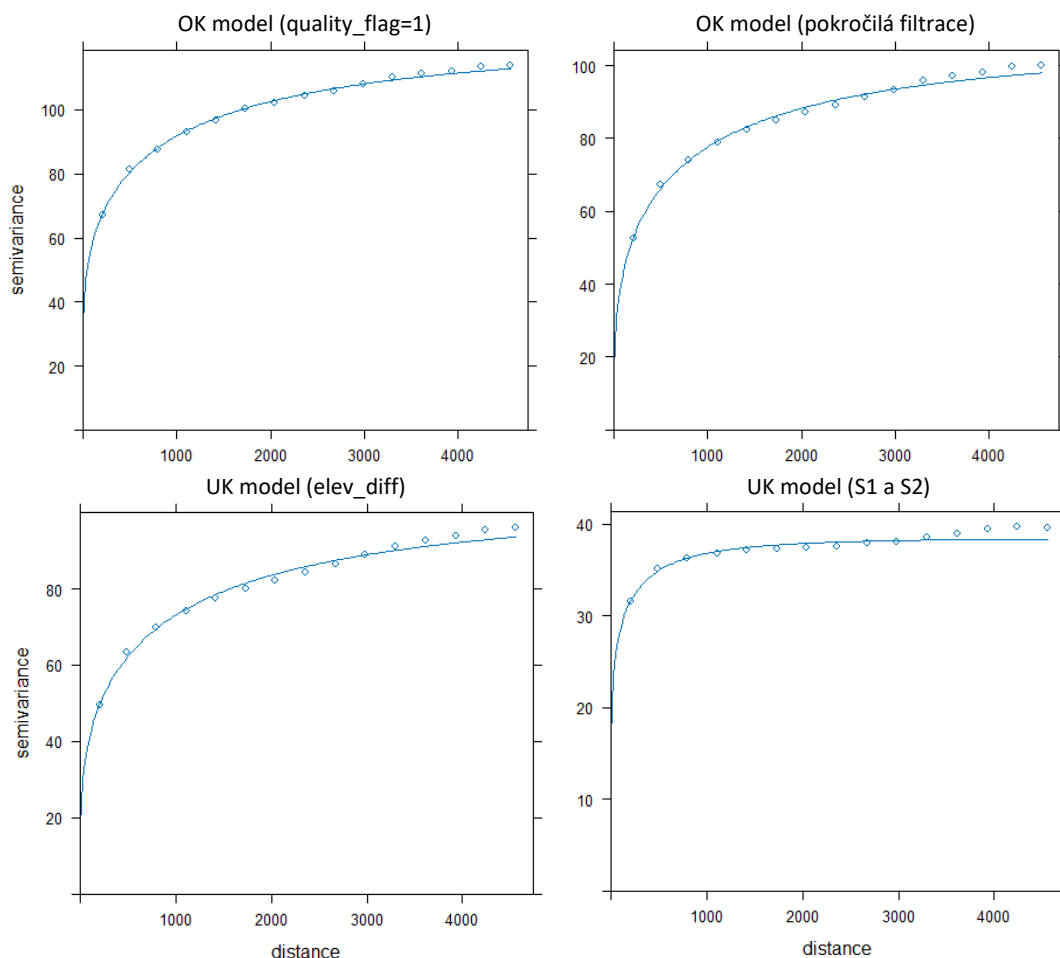
4.2 Prostorová struktura a korelace

Součástí analýzy prostorové struktury a korelace byla především tvorba empirického a teoretického variogramu (variogram modelu) pro popis autokorelované náhodné veličiny před provedením krigingu. Výstupem z této práce byly čtyři CHM, vytvořené pomocí různých typů krigingu (OK, UK) v kombinaci s různými datovými sadami, především pro účely porovnání. Pro všechny čtyři modely krigingu byly tak postupně sestaveny variogramy zobrazené na Obrázek 33. Postupu tvorby variogramu a analýze prostorové struktury se věnuje kapitola 3.4.1. Nejlepší aproximaci u všech empirických variogramů provedl **model stabilní exponenciální třídy** s parametrem shlazení **kappa = 0,5**.

Pro Ordinary Kriging (OK) s datovou sadou naivní filtrace (*quality_flag=1*) obsahoval tento model variogramu parametry: **nugget = 33,03; partial sill = 88,43 a range = 838,16**. Jednalo se tedy o poměrně velkou míru autokorelace, která však obsahovala výrazný podíl mikrovariability. OK byl dále proveden pro data s pokročilou filtrací, která měla parametry modelu variogramu po aproximaci: **nugget = 16,20; partial sill = 89,13 a range = 732,18**. V pokročile filtrovaných datech bylo snížení mikrovariability patrné i vizuálně. Menší nugget a větší částečný práh (partial sill) naznačoval, že se na kratších vzdálenostech vyskytovala větší míra autokorelace, kterou variogram dobře popisoval. Na kratších vzdálenostech tak mohlo dojít k přesnějším odhadům krigingu. Menší parametr dosahu (range) znamenal zároveň pokles schopnosti autokorelace vysvětlit variabilitu bodů na větších vzdálenostech.

Pro datovou sadu s pokročilou filtrací byl tedy následně proveden Universal Kriging (UK) s prediktorem odchylek terénu od referenčního DTM 2012 (*elev_diff*). Ačkoliv probíhala tvorba empirického variogramu bez globálního trendu, určeného atributem *elev_diff*, nejvhodnější byl opět model exponenciální třídy s parametrem $\text{kappa} = 0,5$. Další jeho parametry byly: **nugget = 17,28; partial sill = 84,94 a range = 872,80**. Část dat vysvětlená prediktorem již nemůže přispívat k autokorelaci, proto například partial sill zmenšil. Větší range však znamená lepší schopnost popisovat vzdálenější data. V případě druhé varianty UK byly použity všechny prediktory ze satelitů Sentinel-1 a Sentinel-2 a variogram model vyšel s parametry: **nugget = 14,42; partial sill = 22,03 a range 148,44**. Tato varianta byla tedy příkladem, kdy prediktory

vysvětlily většinu variability v datech, stále však byla přítomná jistá míra autokorelace na krátkých vzdálenostech.



Obrázek 33: Empirické a teoretické variogramy pro jednotlivé varianty tvorby CHM

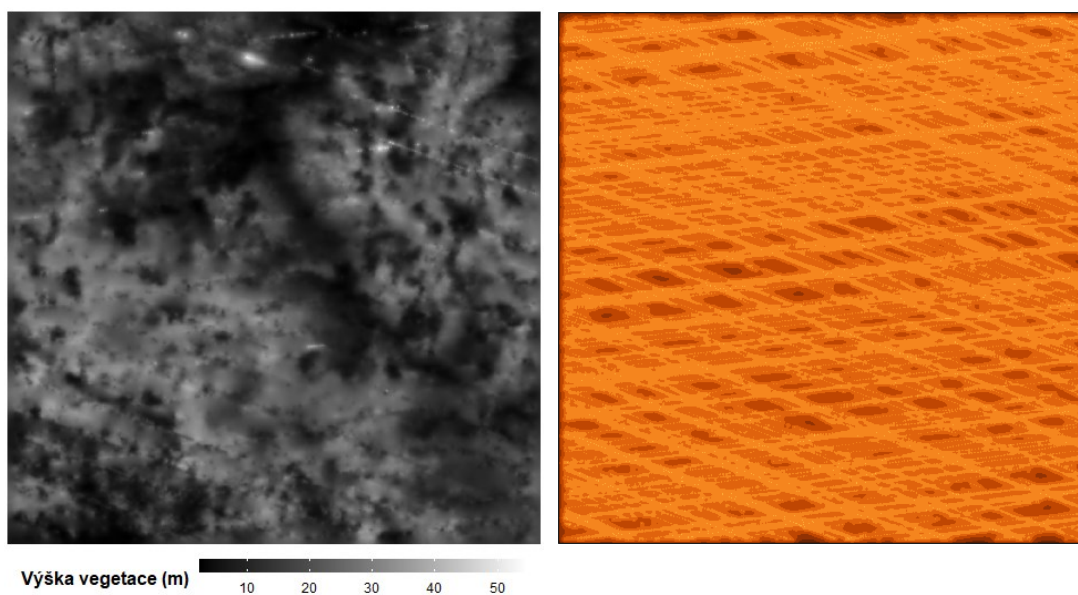
U datasetu s pokročilou filtrací byla dále provedena analýza hustoty bodů ve 100 čtvercích 1x1 km pomocí balíčku funkcí balíčku *spatstat*. Čtverec s minimální hustotou obsahoval 183 měření a čtverec s maximální hustotou 541 měření (viz Příloha 1). Pro celou plochu 10x10 km byla průměrná hustota měření 0,036 bodu/m². Maximální hustota se pohybovala okolo 3 bodů/m², což také značilo určitou míru shlukování. Aby na jednotku plochy připadl průměrně 1 měřený bod, muselo by být rozlišení ~52,92 m.

4.3 Modely výšky vegetace

V této práci byly vytvořeny celkem čtyři modely výšky vegetace, využívající výsledné datové sady v kombinaci s ordinary (OK) nebo universal krigingem (UK). Výsledné chyby určené při validaci těchto modelů jsou dále uvedeny v tabulkách Tabulka 5 a Tabulka 6. Graf vztahu s CHM z ALS je pro OK modely k dispozici

v Příloha 2 a pro UK modely v Příloha 3. Pro další grafické porovnání byl z odchylek OK modelů od CHM z ALS vytvořen rastr rozdílů výšek vegetace (viz Příloha 7). Zobrazeny byly také rozdíly profilů výšky vegetace pro UK modely s daty z ALS (Příloha 8 a Příloha 9).

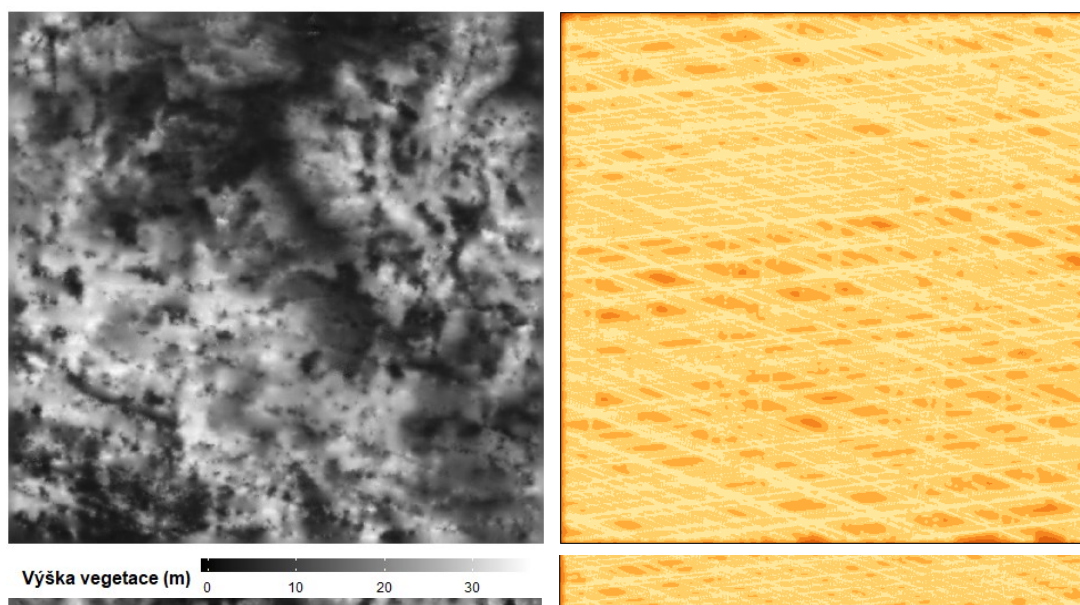
První výsledné CHM (10 m) bylo vytvořeno pomocí metody OK z datové sady s naivní filtrací *quality_flag* = 1. Obrázek 34 zobrazuje výsledný OK model (vlevo) spolu s nejistotou (vpravo), která je ve formě standardní chyby krigingu v rozsahu **6,6 – 9,3 m**. V CHM jsou velké odchylky výšky vegetace, kopírující oběžnou dráhu GEDI.



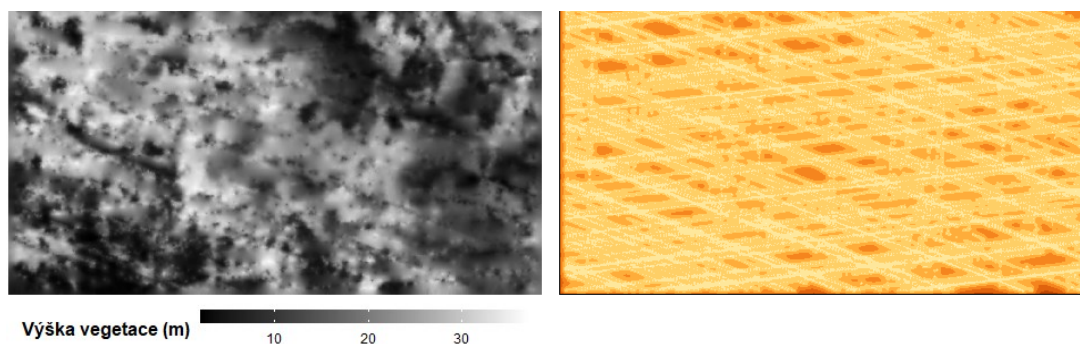
Obrázek 34: Ordinary kriging CHM (10 m) – *quality_flag* = 1 (vlevo) s nejistotou v rozsahu 6,6 – 9,3 m (vpravo)

Další možností bylo použít OK na pokročile vyfiltrovanou datovou sadu. Na Obrázek 35 je vidět, že přitom došlo ke snížení nejistot, v rozsahu **4,9 – 8,5 m**.

Pro datovou sadu s pokročilou filtrací byl dále použit universal kriging. Třetí CHM pro UK využívalo prediktoru odchylek od DTM 2012 (*elev_diff*). Tyto odchylky byly předtím interpolovány pomocí IDW, aby tvořily souvislý povrch. Výsledný model je na Obrázek 36, s rozsahem nejistot ve formě standartní chyby **5,0 – 8,3 m**.



Obrázek 36: Universal kriging CHM (10 m) – elev_diff (vlevo) s nejistotou v rozsahu 5,0 – 8,3 m (vpravo)

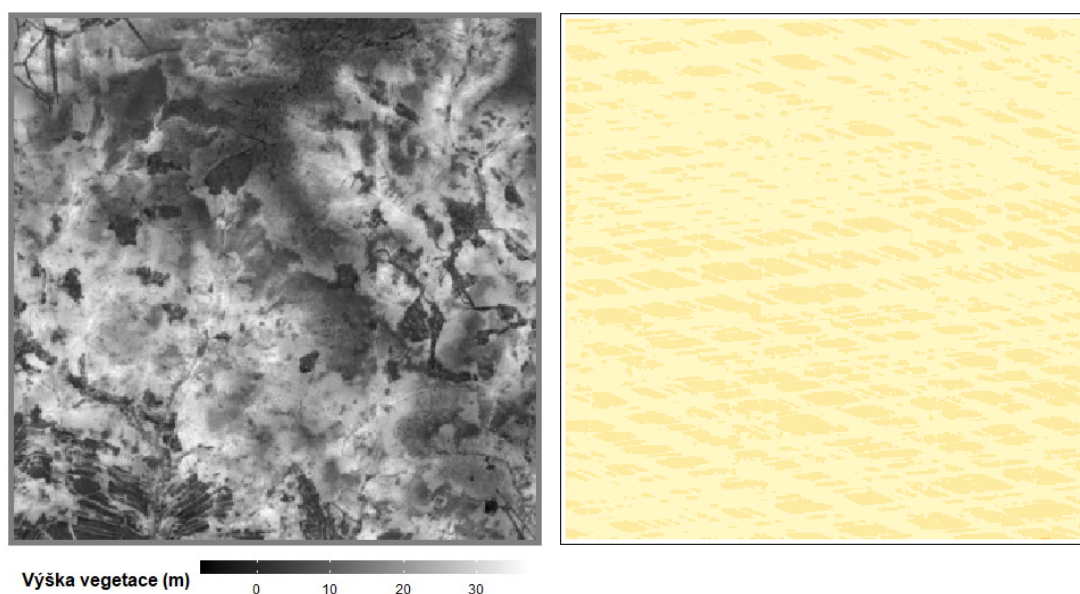


Obrázek 35: Ordinary kriging CHM (10 m) – pokročilá filtrace (vlevo) s nejistotou v rozsahu 4,9 – 8,5 m (vpravo)

Poslední CHM využívalo UK se všemi prediktory pocházejícími ze satelitů Sentinel-1 a Sentinel-2 (viz kapitola 3.3.3) v prostorovém rozlišení 30 m. Jednalo se o nejlepší model výšky vegetace, co se týče výsledků křížové validace a nejistot, které se pohybovaly v rozsahu **4,7 – 5,8 m** (viz následující Obrázek 37).

Pro všechny modely byla provedena křížová validace, kvantifikující chyby ME, MAE a RMSE, uvedené v Tabulka 5. Obecně se dá říci, že UK modely vykazovaly lepší výsledky a zahrnutí globálního trendu tak snižovalo chybu odhadu. Nejlépe vyšel druhý UK model s prediktory ze S1 a S2, kde vyšla střední kvadratická chyba **RMSE ~4,4 m**. První UK model přitom vyšel s **RMSE = ~5,5 m**. Chyby se s každým dalším modelem snižovaly, což je vidět na sloupcovém grafu v Příloha 5.

Tabulka 5: Chyby při křížové validaci u jednotlivých modelů



Obrázek 37: Universal kriging CHM (30 m) – S1 a S2 (vlevo) s nejistotou v rozsahu 4,7 – 5,8 m (vpravo)

Model	ME	MAE	RMSE
Ordinary kriging CHM (quality_flag = 1)	-0,0002	4,729	6,927
Ordinary kriging CHM (pokročilá filtrace)	0,006	4,075	5,702
Universal kriging CHM	0,015	3,890	5,490

(elev_diff)			
Universal kriging CHM (S1 a S2)	-0,0024	3,117	4,422

Dále proběhla validace pomocí CHM z ALS (2012), ve které s nejmenší chybou vyšel naopak první UK model s **RMSE = ~9,14 m**. Druhý UK model obsahoval větší chybu **RMSE = 9,38 m**, ačkoliv z CHM z ALS vysvětloval největší míru variability, **R² = 0,602** (viz Tabulka 6).

Tabulka 6: Chyby při validaci pomocí CHM z ALS (2012) s lineárním vztahem R²

Model	R ²	ME	MAE	RMSE
Ordinary kriging CHM (quality_flag = 1)	0,519	8,152	8,725	10,113
Ordinary kriging CHM (pokročilá filtrace)	0,546	7,502	8,185	9,500
Universal kriging CHM (elev_diff)	0,560	7,141	7,863	9,139
Universal kriging CHM (S1 a S2)	0,602	7,58	8,108	9,376

Větší míry chyb u druhého UK modelu mohly být způsobeny poměrně starým validačním CHM z roku 2012. Na rastru rozdílů od tohoto validačního CHM (viz Příloha 7) bylo vidět, že mezitím v období šesti let (2012–2019) došlo patrně na mnoha místech k odlesnění nebo naopak k přírůstu nových stromů. Tato skutečnost především zvyšovala chybu RMSE, citlivější na výrazné odchylky.

4.4 Porovnání modelů s globálními CHM

Při procesu porovnávání výsledných modelů s globálními CHM (Lang et al., 2022a; Potapov et al., 2021) byly nejprve pro oba globální modely určeny chyby od validačního CHM z ALS (2012) včetně koeficientu determinace R². Graf vztahu obou modelů s validačním CHM je dále dostupný v Příloha 4.

Model Potapov et al. (2021) vyšel s chybami a R²:

ME = 9,277; MAE = 10,323; RMSE = 12,308 a R² = 0,285.

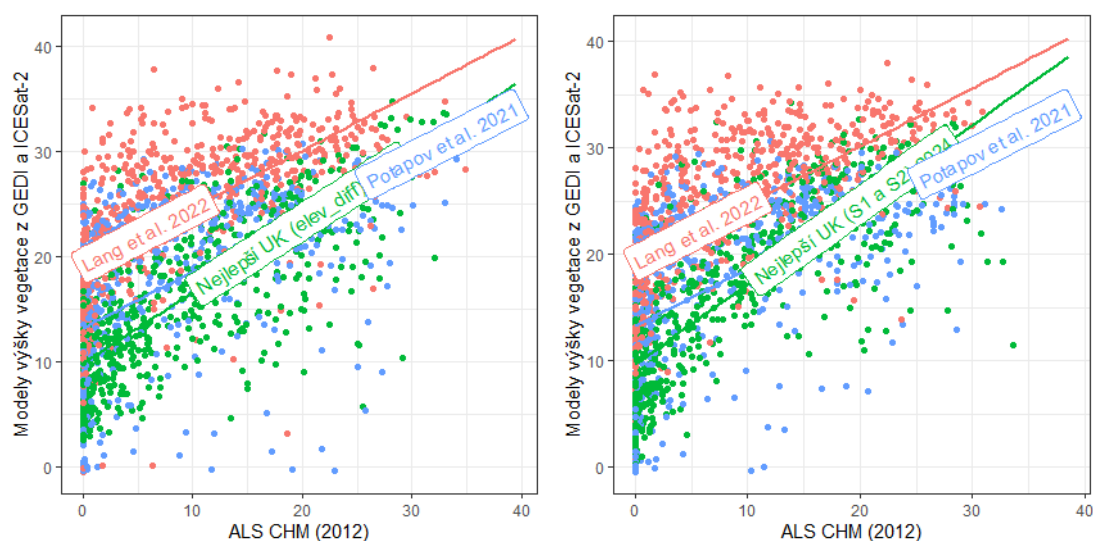
Model Lang et al. (2022a) vyšel s chybami a R²:

ME = 15,298; MAE = 15,481; RMSE = 16,818 a R² = 0,378.

V porovnání s UK modelem (elev_diff) s nejmenší chybou při validaci v této práci:

ME = 7,141; MAE = 7,863; RMSE = 9,139 a R² = 0,56.

Celkově byly chyby globálních CHM daleko větší a také nepopisovaly tolik variability validačního CHM z ALS (2012) jako UK model (elev_diff). Oba globální modely navíc výšku vegetace daleko více nadhodnocovaly, především pak model Lang et al. (2022a) se střední chybou výšky ~15,3 m. RMSE pro CHM ze studie Potapov et al. (2021) vyšlo zhruba o **3,2 m** větší než u UK modelu (elev_diff). U modelu ze studie Lang et al. (2022a) byl tento rozdíl v RMSE daleko větší, zhruba o **7,7 m**. Oba globální modely výšky vegetace jsou dále porovnány s UK modely v této práci na následujícím Obrázek 38.



Obrázek 38: Porovnání globálních CHM ze studií Lang et al. (2022) a Potapov et al. (2021), vlevo s UK modelem (elev_diff) a vpravo s UK modelem (S1 a S2)

Podrobné grafy validací všech OK a UK modelů, včetně globálních CHM jsou k dispozici v přílohách na straně 103.

5. Diskuse

Modelováním globálních modelů výšek vegetace pomocí dat z misí GEDI a ICESat-2 se zabývají výzkumy (Lang et al., 2022a; Potapov et al., 2021). V těchto a dalších studiích, zabývajících se tvorbou lokálních CHM (Schwartz et al., 2022; Sothe et al., 2022; Li et al., 2020), se autoři na rozdíl od této práce téměř nevěnují analýze prostorové struktury a neuvádějí zde např. hustotu nebo redukci výsledných vyfiltrovaných dat z misí GEDI nebo ICESat-2. Jelikož se v každé ze studií provádí jiná metoda filtrace, ať už kombinací algoritmů nebo v datech dostupných atributů, není tak zřejmé, která z nich přinese nejhustší pokrytí dat pro možnost použití geostatistického přístupu modelování CHM. Zejména použitím krigingu pro tvorbu CHM se dále nevěnuje skoro žádná studie. Tato metoda prostorové interpolace byla například využita jen ve studii Narin a Gullu (2023) pro tvorbu DEM.

Této diplomové práci se výběrem prostorového rozlišení 10 m a metriky RH95 z produktu GEDI L2A v002 v kombinaci s daty ze Sentinel-1 a Sentinel-2 nejvíce podobá studie (Schwartz et al., 2022). Použití kombinace dat S2 s GEDI diskutuje přitom již Lang et al. (2022a). Ve studii Schwartz et al. (2022) je výsledkem modelu hlubokého učení U-Net lokální CHM v lesní oblasti Landes ve Francii. Autoři tohoto článku porovnávají výsledný model s globálními modely výšky vegetace Lang et al. (2022a) a Potapov et al. (2021). Validace tohoto modelu probíhala zároveň pomocí více validačních datových sad, především z lesní inventarizace. Podle inventarizačních dat GLORIE například model Schwartz et al. (2022) vyšel s RMSE = 2,84 m, přičemž model Lang et al. (2022a) měl chybu RMSE = 7,06 m a model Potapov et al. (2021) RMSE = 7,37 m. Na rozdíl od výsledků v této diplomové práci (viz kapitola 4.4) zde chyby RMSE obou globálních CHM od dat lesní inventarizace GLORIE vyšly zhruba poloviční. Z toho lze vyvodit, že by modely prezentované v této práci pravděpodobně nevyšly lépe než model Schwartz et al. (2022).

Studie Schwartz et al. (2022) a Lang et al. (2022a) dále například pro filtraci využívají mj. atributu *quality_flag*. Ačkoliv tuto metodu naivní filtrace kombinují i s jinými, pokročilejšími metodami filtrace, dojde tím zbytečně k úbytku dat. V této diplomové práci se totiž potvrdilo, že *quality_flag* zachová v datové sadě GEDI chybná měření a odstraní přitom naopak i některá kvalitní.

Metody prostorové interpolace, založené na strojovém učení nebo hlubokém učení, mohou být také, oproti geostatistickým metodám, náchylné k problému nazývanému jako saturace (Lang et al., 2022a). Saturace nastává, když má model tendenci předpovídat hodnoty, které se vyskytují ve velkém počtu, na úkor hodnot, které jsou méně časté nebo nejsou v datech příliš časté. To může vést k podhodnocení těchto méně častých hodnot a k nepřesným predikcím v těchto oblastech. Tento problém může být zvláště významný v prostředích s nerovnoměrným rozložením dat nebo v datech s vysokou variabilitou. Tyto metody také mnohdy neposkytují přímou kvantifikaci nejistot spojenou s odhady, kterou má například kriging jako součást svého výstupu. Jedná se tedy o určitou výhodu a motivaci využití krigingu jako metody prostorové interpolace.

Kriging má však i své limitace, a to především ve výpočetní náročnosti. Tato práce se tak neustále zabývala otázkou, jak proces výpočtu urychlit, tak aby byl výpočet odhadů z hlediska výpočetního času proveditelný. Pro interpolaci bylo tedy vybráno pouze čtvercové území 10x10 km na území KRNAPu. Bez výkonného hardwaru se dá proces dále zrychlit např. omezením vzdálenosti výpočtu. Tuto vzdálenost je přitom třeba volit opatrně a neustále provádět křížovou validaci. Optimální volbu maximální vzdálenosti výpočtu nebo velikosti regionu omezení je však náročné najít s dostatečnou přesností. S nepřesností je také spojen proces modelování empirického variogramu. Nejčastěji se pro odhad parametrů teoretického variogramu využívá metody nejmenších čtverců, která má však svá omezení a nepřesnosti. Výsledný výběr modelu je přitom ponechán na výzkumníkovi. S těmito nejistotami jsou tedy spojeny i výsledky této práce a jistě by šlo zvolit např. jiný model než model stabilní exponenciální třídy. Dále vznikla otázka, zda je potřeba modelovat anizotropii, protože na větších vzdálenostech se u filtrovaných dat GEDI a ICESat-2 vyskytovala. Ačkoliv byl její vliv po omezení vzdálenosti zanedbatelný, mohlo by její modelování opět přinést odlišné výsledky.

Modelování globálního trendu výšky vegetace pomocí 34 prediktorů ze satelitů Sentinel-1 a Sentinel-2 pomocí universal krigingu bylo v této práci dále spojeno s problémem výrazného rozšíření škály výsledných odhadů, mnohdy i do záporných hodnot. Tento problém byl částečně vyřešen snížením prostorového rozlišení výsledného CHM na 30 m, přesto však nějaký vliv této chyby přetrvával (viz kapitola 3.4.3). Křížovou validaci u tohoto modelu – UK model (S1 a S2) – navíc nešlo provést

pro celý datový soubor, nýbrž pro $nfold = 10000$ (deset tisíc náhodných bodů, což bylo ~25 % z celkového počtu), a proto jsou výsledné míry chyb ME, MAE a RMSE spojeny s určitou nejistotou.

Pro extrakci dat z prediktorů byl zvolen vždy původní tvar stopy GEDI a ICESat-2 podle Liu et al. (2021). Tato stopa však nemusela být např. podle studie Roy et al. (2021) správně geolokována. Schwartz et al. (2022) uvádí například u produktu GEDI v002 geolokační chybu ~10 m. Provedena mohla být také extrakce hodnot prediktorů přímo z bodu měření, což by ale pravděpodobně přesnost hodnoty snížilo. Prediktor odchylek od terénu DTM 2012 (*elev_diff*) pro první UK model musel být interpolován, aby tvořil souvislý rastr pro celé vybrané území 10x10 km. Interpolace byla tedy provedena pomocí IDW s parametrem $\alpha = 2$. Hodnota tohoto parametru byla zvolena tak, aby se zvýšil vliv okolních bodů. Optimální hodnota se však určit nedala. Tento prediktor byl navíc v podstatě vstupem interpolace dat GEDI a ICESat-2 do jiné interpolace dat GEDI a ICESat-2, proto mohlo dojít k velkým nepřesnostem. UK model byl však s tímto prediktorem stále přesnější než OK modely.

Velké nepřesnosti mohly vzniknout také při samotné validaci pomocí CHM 2012, jelikož se jednalo o starší data z ALS. Mezi datem pořízení tohoto CHM (2012) a datem pořízení dat GEDI a ICESat-2 (2018) uběhlo zhruba 6 let, při kterých mohlo dojít k velkým změnám ve struktuře vegetace. Tímto faktem tak mohly být negativně ovlivněny všechny chyby RMSE. Na druhou stranu mohly vzniknout rastry rozdílů výšek vegetace, které tak porovnávají přírůst a úbytek mezi lety 2012–2018 (viz Příloha 7). Na rozdíl od modelů ve studiích Lang et al. (2022a) a Potapov et al. (2021) také nebyla odstraněna z výsledných CHM zástavba, ačkoliv se na vybraném území 10x10 km v KRNAPu vyskytovala především vegetace.

Z výše uvedených limitů a nejistot vyplývá i doporučení pro další výzkum této problematiky. Pro validaci by mohla být především vybrána novější data z ALS a krigingové CHM by pak mohlo být vytvořeno pro celé území KRNAPu. Po vzoru studie Schwartz et al. (2022) by dále šlo ověřit přesnost (provést validaci) zvláště pro listnaté či jehličnaté lesy. Jelikož například ve studii Lang et al. (2022a) řešili problém podhodnocení výšek vegetace CNN modelu strojového učení, zajímavé by bylo také porovnat model krigingu s tímto globálním CHM na různých výškových úrovních.

Další možností by bylo zahrnutí anizotropie do modelu nebo snaha o výběr přesnější metody popisu prostorové autokorelace. Pro modelování po vzoru UK by šlo také využít smíšené modely například v balíčku *spaMM*. Tyto modely kombinují jak deterministický přístup modelování globálního trendu střední hodnoty, tak statistický přístup modelování náhodného efektu pomocí Matérnovy korelační funkce. Z těchto modelů lze provádět odhad neznámých hodnot a zjistit lze také signifikanci jednotlivých prediktorů v modelu. Alternativou by mohl být Regression Kriging (RK), který modelování provádí zvlášť ve dvou krocích. Zjištění signifikance jednotlivých prediktorů v modelu by mohlo vést k odstranění nesignifikantních prediktorů pro potenciální zrychlení výpočtu odhadů. Další možností, jak zkrátit výpočetní čas, by mohla být redukce počtu prediktorů (dimenzí) pomocí analýzy hlavních komponent, kde by jako nezávislé proměnné do UK modelu vstupovaly pouze komponenty vysvětlující např. 90 % původních dat. Alternativním způsobem redukce by také mohl být výběr prediktorů s minimální vzájemnou korelací např. $< 0,7$. Byly by tak zvoleny co nejvíce nezávislé proměnné (viz Příloha 12).

6. Závěr

V rámci této diplomové práce byly posouzeny možnosti využití krigingu jako metody prostorové interpolace výšky vegetace z dat satelitní laserové altimetrie misí GEDI a ICESat-2. Prověřeny a testovány byly různé metody filtrace dat z těchto misí, které byly obecně spojeny s velkou nejistotou a chybami měření. Pro vyfiltrovaná data byla následně provedena analýza prostorové autokorelace, která potvrdila její přítomnost ve velké míře. Na základě posouzení a popsání prostorové struktury teoretickým variogramem byly vytvořeny čtyři varianty modelů výšky vegetace (CHM), které kombinovaly různé přístupy filtrací pro Ordinary Kriging (OK) a různé prediktory pro Universal Kriging (UK). Pro druhý z UK modelů byly zároveň použity prediktory ze satelitu Sentinel-1 a Sentinel-2. Výsledné modely byly vždy podrobeny křížové validaci a validaci pomocí CHM z leteckého laserového skenování (ALS) z roku 2012. Pro celkové ověření použité metodiky a přesnosti modelů vytvořených pomocí krigingu bylo provedeno porovnání těchto modelů s existujícími globálními modely výšky vegetace ze studií Lang et al. (2022a) a Potapov et al. (2021). Tyto globální CHM byly vytvářeny pomocí metod strojového učení. Tento přístup umožnil specifikaci optimálních parametrů pro zpracování, analýzu a modelování dat, včetně ověření limitů a nejistot spojených s použitými metodami. Všechny Výsledky byly podrobně shrnuty v jednotlivých podkapitolách kapitoly č. 4.

Pro snížení času výpočtu bylo vybráno území (pilotní lokalita) 10x10 km v areálu Krkonošského národního parku. Tato lokalita sloužila zároveň pro ověření výkonnosti modelu na jejím členitém povrchu. Jako reprezentativní metrika výšky vegetace z dat GEDI byla kvůli nejmenšímu nadhodnocování CHM z ALS (2012) vybrána metrika RH95, tj. 95. percentil výšky vegetace.

Filtrace měla za cíl co nejvíce minimalizovat chybu od referenčního digitálního modelu výšky terénu (DTM) z dat ALS pro rok 2012. Při filtraci bylo zároveň kritériem zachování co největšího počtu dat. Pro data GEDI přinesla nejlepší výsledky sedmá varianta pokročilé filtrace s chybou RMSE = 4,284 m od DTM 2012. Toto zjištění potvrzuje, že je tato metoda filtrace dat GEDI vhodnější než tradiční přístupy, jako je například (naivní) filtrace podle atributu *quality_flag*. Pokročilá filtrace navíc zachovala 42 % původních dat, oproti naivní filtraci *quality_flag*, která zachovala pouze 41,64 %. Filtrace pomocí kombinace GEDI atributů *sensitivity*, *degrade_flag*

nebo *solar_elev* přitom zachovávaly mnohokrát méně bodů než naivní filtrace *quality_flag*. Většina studií pracujících s daty výšky vegetace GEDI a ICESat-2 pro tvorbu CHM však tyto neefektivní metody filtrace používala (Schwartz et al., 2022). Filtracemi pomocí kombinací zmíněných atributů z GEDI, s velkou datovou redukcí, a dalším poměrně agresivním přístupům filtrace se přitom věnovaly i studie globálních CHM Lang et al. (2022a) a Potapov et al. (2021).

V dalším krocích analýza prostorové autokorelace (variografie) odhalila vzájemnou signifikantní prostorovou závislost měření výšek vegetace, což bylo klíčové pro úspěšnou aplikaci metody krigingu. Prostorová autokorelace byla potvrzena nejen empirickými variogramy, ale také teoretickými variogramy, což přispělo hlubšímu porozumění prostorové struktury výšky vegetace z dat GEDI a ICESat-2 v závislosti na metodě filtrace nebo modelování s využitím prediktorů. Obecně byla velká část této práce věnována analýzám prostorové struktury, včetně měření hustoty dat, která pro pokročile vyfiltrovaná data vyšla průměrně 0,036 bodů/m². Tato analýza zároveň na vyfiltrovaných datech odhalila potenciální přítomnost shluků a přítomnost anizotropie na větších vzdálenostech.

Pomocí Ordinary Krigingu (OK) byl 1. vytvořen model z datové sady GEDI s naivní filtrací, který byl porovnán 2. s OK modelem z datové sady GEDI a ICESat-2 po pokročilé filtraci. Zde bylo ověřeno, že filtrace pomocí *quality_flag* zachovává v datech chybná měření a zvyšuje tak nejistotu v OK modelu. Druhý OK model vyšel daleko lépe, RMSE od CHM z ALS (2012) vyšlo při validaci 9,5 m. Pro datovou sadu kombinující data z GEDI a ICESat-2 byly vytvořeny dvě varianty Universal Kriging (UK) modelu, který za 1. využíval prediktory odchylek terénu GEDI a ICESat-2 od DTM 2012 (*elev_diff*) a za 2. využíval všech 34 vybraných prediktorů ze satelitů Sentinel-1 (S1) a Sentinel-2 (S2) (viz Příloha 11). Zatímco všechny modely byly pro účely porovnatelnosti s globálními CHM vytvářeny v prostorovém rozlišení 10 m, druhý UK model musel být, kvůli výpočetní náročnosti, vytvářen v rozlišení 30 m. Ačkoliv druhý UK model (S1 a S2) vysvětloval nejvíce variability z validačních dat CHM 2012 ($R^2 = 0,602$), nejmenší chyby při validaci měl první UK model (*elev_diff*), RMSE = 9,139.

Porovnání výsledných modelů s globálními modely CHM odhalilo, že modely v této práci dosáhly výrazně lepších výsledků. Zatímco nejnižší chyba RMSE globálních modelů byla 10,323 m (Potapov et al., 2021), modely vytvořené v této práci

prezentovaly mnohem nižší chyby s nejmenší chybou RMSE = 9,139 m pro první UK model (elev_diff). Tento úspěch je přičítán pokročilejším metodám filtrace a efektivní parametrizaci krigingu, včetně poměrně přesného popisu prostorové autokorelace pomocí modelu stabilní exponenciální třídy.

Tato diplomová práce tak celkově přináší novou, alternativní metodu tvorby CHM pomocí krigingu, ověřenou na pilotní lokalitě 10x10 km v areálu Krkonošského národního parku. Dále pak zdůrazňuje potřebu dalšího výzkumu v této oblasti, pro který podává základ kapitolou č. 4 s výsledky a diskusí v kapitole č. 5. Bylo prokázáno, že použitím pokročilé metody filtrace při minimální redukci dat (viz kapitola 3.3.2) v kombinaci s vhodnými parametry krigingu, je možné spolehlivě vytvářet modely výšky vegetace z dat GEDI a ICESat-2, které jsou přesnější, než existující globální modely výšky vegetace Lang et al. (2022a) a Potapov et al. (2021). Tyto modely výšky vegetace mají velký potenciál ve studiu dynamiky struktury vegetace, monitorování přírůstu a odlesňování či v jiných environmentálních či lesnických aplikacích. CHM ze satelitních lidarů mohou být také efektivním nástrojem pro aktualizace či doplnění přesnějších modelů výšky vegetace z ALS. Jejich hlavními výhodami jsou tedy časté aktualizace a volná dostupnost dat z misí GEDI a ICESat-2.

7. Seznam použité literatury

ADAM, Markus; URBAZAEV, Mikhail; DUBOIS, Clémence a SCHMULLIUS, Christiane, 2020. Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters: Influence of Environmental and Acquisition Parameters. *Remote Sensing*. Roč. 12, č. 23. ISSN 2072-4292. Dostupné z: <https://doi.org/10.3390/rs12233948>.

BAKX, Tristan; KOMA, Zsófia; SEIJMONSBERGEN, Arie a KISSLING, W., 2019. Use and categorization of Light Detection and Ranging vegetation metrics in avian diversity and species distribution research. *Diversity and Distributions*. Roč. 25. Dostupné z: <https://doi.org/10.1111/ddi.12915>.

BECK, Jared; WIRT, Bradford; ARMSTON, John; HOFTON, Michelle; LUTHCKE, Scott et al., 2021. *GLOBAL Ecosystem Dynamics Investigation (GEDI) Level 2 User Guide: For SDPS PGEVersion 3 (P003) of GEDI L2A Data and SDPS PGEVersion 3 (P003) of GEDI L2B Data*. PDF. Version 2. Dostupné z: https://lpdaac.usgs.gov/documents/986/GEDI02_UserGuide_V2.pdf. [cit. 2024-03-08].

BERGEN, K. M.; GOETZ, S. J.; DUBAYAH, R. O.; HENEGBRY, G. M.; HUNSAKER, C. T. et al., 2009. Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions: Review and implications for lidar and radar spaceborne missions. *Journal of Geophysical Research: Biogeosciences*. Roč. 114, č. G2. ISSN 0148-0227. Dostupné z: <https://doi.org/https://doi.org/10.1029/2008JG000883>.

CALDERS, Kim; ADAMS, Jennifer; ARMSTON, John; BARTHOLOMEUS, Harm; BAUWENS, Sebastien et al., 2020. Terrestrial laser scanning in forest ecology: Expanding the horizon: Expanding the horizon. *Remote Sensing of Environment*. Roč. 251, s. 112102. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2020.112102>.

DOBROVOLNÝ, Petr, 2005a. Metody prostorové interpolace: Z8102 Geostatistika. Online. In: . Masarykova univerzita, s. 41-45. Dostupné z: https://is.muni.cz/el/sci/jaro2005/Z8102/um/Geostatistika_6.pdf. [cit. 2024-03-10].

DOBROVOLNÝ, Petr, 2005b. Geostatistické metody interpolace: Z8102 Geostatistika. Online. In: . Masarykova univerzita, s. 59-72. Dostupné z: https://is.muni.cz/el/sci/jaro2005/Z8102/um/Geostatistika_7.pdf. [cit. 2024-03-10].

DOBROVOLNÝ, Petr, 2005c. KRIGING – geostatistické metody interpolace: Z8102 Geostatistika. Online. In: . Masarykova univerzita, s. 73-86. Dostupné z: https://is.muni.cz/el/sci/jaro2005/Z8102/um/Geostatistika_6.pdf. [cit. 2024-03-10].

DRAKE, Jason B; DUBAYAH, Ralph O; KNOX, Robert G; CLARK, David B a BLAIR, J.B, 2002. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*. Roč. 81, č. 2, s. 378-392. ISSN 0034-4257. Dostupné z: [https://doi.org/https://doi.org/10.1016/S0034-4257\(02\)00013-5](https://doi.org/https://doi.org/10.1016/S0034-4257(02)00013-5).

DUBAYAH, Ralph; ARMSTON, John; HEALEY, Sean P.; BRUENING, J. M.; PATTERSON, Paul L. et al., 2022. GEDI Launches a New Era of Biomass Inference From Space. *Environmental Research Letters*. Dostupné z: <https://doi.org/10.1088/1748-9326/ac8694>.

DUBAYAH, Ralph; BLAIR, James Bryan; GOETZ, Scott; FATOYINBO, Lola; HANSEN, Matthew et al., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*. Roč. 1, s. 100002. ISSN 2666-0172. Dostupné z: <https://doi.org/10.1016/j.srs.2020.100002>.

DUBAYAH, Ralph; BLAIR, James Bryan; GOETZ, Scott; FATOYINBO, Lola; HANSEN, Matthew et al., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*. Roč. 1, s. 100002. ISSN 2666-0172. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.srs.2020.100002>.

DUNCANSON, Laura; KELLNER, James R.; ARMSTON, John; DUBAYAH, Ralph; MINOR, David M. et al., 2022. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission. *Remote Sensing of Environment*. Roč. 270, s. 112845. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2021.112845>.

FAYAD, Ibrahim; BAGHDADI, Nicolas a RIEDI, Jérôme, 2021a. Quality Assessment of Acquired GEDI Waveforms: Case Study over France, Tunisia and French Guiana: Case Study over France, Tunisia and French Guiana. *Remote Sensing*. Roč. 13, č. 16. ISSN 2072-4292. Dostupné z: <https://doi.org/10.3390/rs13163144>.

FAYAD, Ibrahim; BAGHDADI, Nicolas; ÁLVARES, Clayton Alcarde; STAPE, José Luiz; BAILLY, Jean Stéphane et al., 2021b. Assessment of GEDI's LiDAR Data for the Estimation of Canopy Heights and Wood Volume of Eucalyptus Plantations in Brazil. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Dostupné z: <https://doi.org/10.1109/jstars.2021.3092836>.

GÁBOR, Lukáš; MOUDRÝ, Vítězslav; LECOURS, Vincent; MALAVASI, Marco; BARTÁK, Vojtěch et al., 2019. The Effect of Positional Error on Fine Scale Species Distribution Models Increases for Specialist Species. *Ecography*. Dostupné z: <https://doi.org/10.1111/ecog.04687>.

GEDI L2A Product Data Dictionary. Online. In: NASA EOSDIS LP DAAC, USGS EROS. *LP DAAC*. Dostupné z: https://lpdaac.usgs.gov/documents/982/gedi_l2a_dictionary_P003_v2.html. [cit. 2024-03-22].

GOETZ, Scott a DUBAYAH, Ralph, 2011. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management*. Roč. 2, č. 3, s. 231-244. ISSN 1758-3004. Dostupné z: <https://doi.org/10.4155/cmt.11.18>.

HANCOCK, Steven; ARMSTON, John; HOFTON, M. A.; SUN, Xiaoli; TANG, Hao et al., 2019. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth and Space Science*. Dostupné z: <https://doi.org/10.1029/2018ea000506>.

HEIN, Jonas; GUARIN, Alejandro; FROMMÉ, Ezra a PAUW, Pieter, 2018. Deforestation and the Paris climate agreement: An assessment of REDD + in the national climate action plans. *Forest Policy and Economics*. Roč. 90, s. 7-11. ISSN 13899341. Dostupné z: <https://doi.org/10.1016/j.forpol.2018.01.005>.

HEROLD, Martin; CARTER, Sarah; AVITABILE, Valerio; ESPEJO, Andrés B.; JONCKHEERE, Inge et al., 2019. The Role and Need for Space-Based Forest

Biomass-Related Measurements in Environmental Management and Policy. *Surveys in Geophysics*. Roč. 40, č. 4, s. 757-778. ISSN 1573-0956. Dostupné z: <https://doi.org/10.1007/s10712-019-09510-6>.

HOFTON, Michelle; BLAIR, Brian J; STORY, Sarah a YI, Donghui, 2019. *Algorithm Theoretical Basis Document (ATBD): for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products*. PDF. Dostupné z: https://lpdaac.usgs.gov/documents/581/GEDI_WF_ATBD_v1.0.pdf. [cit. 2024-03-08].

HUETTERMANN, Sven; JONES, Simon; SOTO-BERELOV, Mariela a HISLOP, Samuel, 2022. Intercomparison of Real and Simulated GEDI Observations across Sclerophyll Forests. *Remote Sensing*. Roč. 14, č. 9. ISSN 2072-4292. Dostupné z: <https://doi.org/10.3390/rs14092096>.

ISENBURG, Martin, 2014. *Rasterizing Perfect Canopy Height Models from LiDAR*. Online. In: Rapidlasso GmbH. Dostupné z: <https://rapidlasso.de/rasterizing-perfect-canopy-height-models-from-lidar/>. [cit. 2024-03-05].

JEŽEK, Josef, 2015. *Geostatistika a prostorová interpolace*. V Praze: Univerzita Karlova, nakladatelství Karolinum. ISBN 978-80-246-3076-2.

JIN, Li a HEAP, A. D., 2008. *A review of spatial interpolation methods for environmental scientists*. Canberra: Geoscience Australia.

KELLNER, James R.; ARMSTON, John a DUNCANSON, Laura, 2023. Algorithm Theoretical Basis Document for GEDI Footprint Aboveground Biomass Density. *Earth and Space Science*. Roč. 10, č. 4, s. e2022EA002516. ISSN 2333-5084. Dostupné z: <https://doi.org/https://doi.org/10.1029/2022EA002516>.

KHALSA, S.J.S.; BORSA, A. a NANDIGAM, V., 2020. OpenAltimetry - rapid analysis and visualization of Spaceborne altimeter data. *Earth Science Informatics*. Dostupné z: <https://doi.org/10.1007/s12145-020-00520-2>.

KHOSRAVIPOUR, A.; SKIDMORE, A.K.; ISENBURG, M.; WANG, Tiejun a HUSSIN, Y.A., 2014. Generating pit - free canopy height models from airborne lidar. *Photogrammetric engineering and remote sensing*. Roč. 80, č. 9, s. 863-872. ISSN 0099-1112. Dostupné z: <https://doi.org/10.14358/PERS.80.9.863>.

LANG, Nico; JETZ, Walter; SCHINDLER, Konrad a WEGNER, Jan Dirk, 2022a. *A high-resolution canopy height model of the Earth*. Dostupné z: <https://doi.org/10.48550/arXiv.2204.08322>.

LANG, Nico; JETZ, Walter; SCHINDLER, Konrad a WEGNER, Jan Dirk, 2023. A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*. Roč. 7, č. 11, s. 1778-1789. ISSN 2397-334X. Dostupné z: <https://doi.org/10.1038/s41559-023-02206-6>.

LANG, Nico; KALISCHEK, Nikolai; ARMSTON, John; SCHINDLER, Konrad; DUBAYAH, Ralph et al., 2022b. Global Canopy Height Regression and Uncertainty Estimation From GEDI LIDAR Waveforms With Deep Ensembles. *Remote Sensing of Environment*. Dostupné z: <https://doi.org/10.1016/j.rse.2021.112760>.

LANG, Nico; SCHINDLER, Konrad a WEGNER, Jan Dirk, 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sensing of Environment*. Roč. 233, s. 111347. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2019.111347>.

LEFSKY, M. A., 2010. A Global Forest Canopy Height Map From the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*. Dostupné z: <https://doi.org/10.1029/2010gl043622>.

LEWIS, William; CHANDLER, Richard; DELANCEY, Clayton; RUSHTON, Emily; WANN, Gregory et al., 2022. Abundance and distribution of ruffed grouse *Bonasa umbellus* at the southern periphery of the range. *Wildlife Biology*. Roč. 2022. Dostupné z: <https://doi.org/10.1002/wlb3.01017>.

LI, Wang; NIU, Zheng; SHANG, Rong; QIN, Yuchu; WANG, Li et al., 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*. Roč. 92, s. 102163. ISSN 1569-8432. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.jag.2020.102163>.

LIU, Aobo; CHENG, Xiao a CHEN, Zhuoqi, 2021. Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals.

Remote Sensing of Environment. Roč. 264, s. 112571. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2021.112571>.

LIU, Haijian a DONG, Pinliang, 2014. A new method for generating canopy height models from discrete-return LiDAR point clouds. *Remote Sensing Letters*. Roč. 5, č. 6, s. 575-582. ISSN 2150-704X. Dostupné z: <https://doi.org/10.1080/2150704X.2014.938180>.

LIU, Xin; HUI, Guihua; GUO, Jinyun; ZHANG, Tinghui a SONG, Menghao, 2022. Inversion of Deflection of the Vertical in the South China Sea Using ICESat-2 Sea Surface Height Data. *Remote Sensing*. Roč. 15, s. 30. Dostupné z: <https://doi.org/10.3390/rs15010030>.

LOH, Ho Yan; JAMES, Daniel; IOKI, Keiko; WONG, Wilson Vun Chiong; TSUYUKI, Satoshi et al., 2020. Aboveground Biomass Changes in Tropical Montane Forest of Northern Borneo Estimated Using Spaceborne and Airborne Digital Elevation Data. *Remote Sensing*. Dostupné z: <https://doi.org/10.3390/rs12223677>.

LP DAAC, 2021. *GEDI02_A v002*. Online. THE LAND PROCESSES DISTRIBUTED ACTIVE ARCHIVE CENTER (LP DAAC), USGS, NASA. GEDI02_A. Dostupné z: https://lpdaac.usgs.gov/products/gedi02_av002/. [cit. 2024-03-09].

LUTHCKE, Scott B; REBOLD, Tim; THOMAS, Taylor a PENNINGTON, Teresa, 2019. *Algorithm Theoretical Basis Document (ATBD): for GEDI Waveform Geolocation for L1 and L2 Products*. PDF. Dostupné z: https://lpdaac.usgs.gov/documents/579/GEDI_WFGEO_ATBD_v1.0.pdf. [cit. 2024-03-08].

MARKUS, Thorsten; NEUMANN, Tom; MARTINO, Anthony; ABDALATI, Waleed; BRUNT, Kelly et al., 2017. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation: Science requirements, concept, and implementation. *Remote Sensing of Environment*. Roč. 190, s. 260-273. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2016.12.029>.

MARSELIS, Suzanne M; KEIL, Petr; CHASE, Jonathan M a DUBAYAH, Ralph, 2022. The use of GEDI canopy structure for explaining variation in tree species

richness in natural forests. *Environmental Research Letters*. Roč. 17, č. 4, s. 045003. ISSN 1748-9326. Dostupné z: <https://doi.org/10.1088/1748-9326/ac583f>.

MEYER, Hanna a PEBESMA, Edzer, 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*. Roč. 13, č. 1, s. 2208. ISSN 2041-1723. Dostupné z: <https://doi.org/10.1038/s41467-022-29838-9>.

MILENKOVIĆ, Milutin; REICHE, Johannes; ARMSTON, John; NEUENSCHWANDER, Amy; DE KEERSMAECKER, Wanda et al., 2022. Assessing Amazon rainforest regrowth with GEDI and ICESat-2 data. *Science of Remote Sensing*. Roč. 5, s. 100051. ISSN 2666-0172. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.srs.2022.100051>.

MORAGA, Paula, 2023. *Spatial Statistics for Data Science: Theory and Practice with R*. 1. Chapman & Hall/CRC Data Science Series.

MORIN, David; PLANELLS, Milena; BAGHDADI, Nicolas; BOUVET, Alexandre; FAYAD, Ibrahim et al., 2022. Improving Heterogeneous Forest Height Maps by Integrating GEDI-Based Forest Height Information in a Multi-Sensor Mapping Process. *Remote Sensing*. Dostupné z: <https://doi.org/10.3390/rs14092079>.

MOUDRÝ, Vítězslav, 2021. *LiDAR data processing: Lesson 7 (Generating Canopy Height Models, LAStools/ArcGIS)*. Online. In: MOUDRÝ, Vítězslav. *Remote Sensing and GIS in Ecology*. Dostupné z: <http://www.gisdata.cz/lidar-data-processing/>. [cit. 2024-03-06].

MOUDRÝ, Vítězslav; CORD, Anna F.; GÁBOR, Lukáš; LAURIN, Gaia Vaglio; BARTÁK, Vojtěch et al., 2023. Vegetation structure derived from airborne laser scanning to assess species distribution and habitat suitability The way forward. *Diversity and Distributions*. Roč. 29, č. 1, s. 39. Dostupné také z: <https://www.jstor.org/stable/48706373>.

MOUDRÝ, Vítězslav; GDULOVÁ, Kateřina; GÁBOR, Lukáš; ŠÁROVCOVÁ, Eliška; BARTÁK, Vojtěch et al., 2022. Effects of environmental conditions on ICESat-2 terrain and canopy heights retrievals in Central European mountains. *Remote Sensing of Environment*. Roč. 279, s. 113112. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2022.113112>.

MULVERHILL, Christopher; COOPS, Nicholas; HERMOSILLA, Txomin; WHITE, Joanne a WULDER, Michael, 2022. Evaluating ICESat-2 for monitoring, modeling, and update of large area forest canopy height products. *Remote Sensing of Environment*. Roč. 271, s. 112919. Dostupné z: <https://doi.org/10.1016/j.rse.2022.112919>.

NARIN, Omer Gokberk a GULLU, Mevlut, 2023. A comparison of vertical accuracy of global DEMs and DEMs produced by GEDI, ICESat-2. *Earth Science Informatics*. Roč. 16, č. 3, s. 2693-2707. ISSN 1865-0481. Dostupné z: <https://doi.org/10.1007/s12145-023-01060-1>.

NEUENSCHWANDER, A.; PITTS, K.; JELLEY, B.; ROBBINS, J.; MARKEL, J. et al., 2022. *Ice, Cloud, and Land Elevation Satellite (ICESat-2) Project Algorithm Theoretical Basis Document (ATBD) for Land - Vegetation Along-Track Products (ATL08), Version 6*. ICESat-2 Project. Dostupné z: <https://doi.org/10.5067/8ANPSL1NN7YS>.

NEUMANN, Thomas A.; MARTINO, Anthony J.; MARKUS, Thorsten; BAE, Sungkoo; BOCK, Megan R. et al., 2019. The Ice, Cloud, and Land Elevation Satellite – 2 mission: A global geolocated photon product derived from the Advanced Topographic Laser Altimeter System: A global geolocated photon product derived from the Advanced Topographic Laser Altimeter System. *Remote Sensing of Environment*. Roč. 233, s. 111325. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2019.111325>.

PAN, Yude; BIRDSEY, Richard A.; FANG, Jingyun; HOUGHTON, Richard; KAUPPI, Pekka E. et al., 2011. A Large and Persistent Carbon Sink in the World's Forests. Online. *Science*. Roč. 333, č. 6045, s. 988-993. ISSN 0036-8075. Dostupné z: <https://doi.org/10.1126/science.1201609>. [cit. 2024-03-06].

POTAPOV, P.; HANSEN, M.C.; KOMMAREDDY, I.; KOMMAREDDY, A.; TURUBANOVA, S. et al., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sens*. Roč. 12, s. 426. Dostupné z: <https://doi.org/10.3390/rs12030426>.

POTAPOV, Peter; LI, Xinyuan; HERNANDEZ-SERNA, Andres; TYUKAVINA, Alexandra; HANSEN, Matthew C. et al., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of*

Environment. Roč. 253, s. 112165. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2020.112165>.

Products Overview - GEDI, 2019. Online. NASA & THE UNIVERSITY OF MARYLAND. GEDI: Home Page. Dostupné z: <https://gedi.umd.edu/data/products/>. [cit. 2024-03-08].

RANNDAL, Heidi; SIGAARD CHRISTIANSEN, Philip; KLIVING, Pernille; BALTAZAR ANDERSEN, Ole a NIELSEN, Karina, 2021. Evaluation of a Statistical Approach for Extracting Shallow Water Bathymetry Signals from ICESat-2 ATL03 Photon Data. *Remote Sensing*. Roč. 13, č. 17. ISSN 2072-4292. Dostupné z: <https://doi.org/10.3390/rs13173548>.

ROY, David P.; KASHONGWE, Herve B. a ARMSTON, John, 2021. The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring. *Science of Remote Sensing*. Roč. 4, s. 100024. ISSN 2666-0172. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.srs.2021.100024>.

SCHUTZ, B.; ZWALLY, H.; SHUMAN, C.A. a HANCOCK, D., 2005. Overview of the ICESat mission. *Geophysical Research Letters*. Roč. 32. Dostupné z: <https://doi.org/10.1029/2005GL02400>.

SCHWARTZ, Martin; CIAIS, Philippe; OTTLÉ, Catherine; DE TRUCHIS, Aurelien; VEGA, Cedric et al., 2022. High-resolution canopy height map in the Landes forest (France) based on GEDI, Sentinel-1, and Sentinel-2 data with a deep learning approach. Dostupné z: <https://doi.org/https://doi.org/10.48550/arXiv.2212.10265>.

SIMARD, Marc; PINTO, N.; FISHER, Joshua B. a BACCINI, Alessandro, 2011. Mapping Forest Canopy Height Globally With Spaceborne Lidar. *Journal of Geophysical Research Atmospheres*. Dostupné z: <https://doi.org/10.1029/2011jg001708>.

SOTHE, Camile; GONSAMO, Alemu; LOURENÇO, Ricardo B.; KURZ, Werner A. a SNIDER, James, 2022. Spatially Continuous Mapping of Forest Canopy Height in Canada by Combining GEDI and ICESat-2 with PALSAR and Sentinel. *Remote Sensing*. Roč. 14, č. 20. ISSN 2072-4292. Dostupné z: <https://doi.org/10.3390/rs14205158>.

SPRACKLEN, Benedict D. a SPRACKLEN, Dominick V., 2021. Determination of Structural Characteristics of Old-Growth Forest in Ukraine Using Spaceborne LiDAR. *Remote Sensing*. Dostupné z: <https://doi.org/10.3390/rs13071233>.

STEREŃCZAK, Krzysztof; CIESIELSKI, Mariusz; BALAZY, Radomir a ZAWIŁA-NIEDŹWIECKI, Tomasz, 2016. Comparison of various algorithms for DTM interpolation from LIDAR data in dense mountain forests. *European Journal of Remote Sensing*. Roč. 49, s. 599-621. Dostupné z: <https://doi.org/10.5721/EuJRS20164932>.

TANG, Hao; DUBAYAH, Ralph; SWATANTRAN, Anu; HOFTON, Michelle; SHELDON, Sage et al., 2012. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sensing of Environment*. Roč. 124, s. 242-250. ISSN 0034-4257. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.rse.2012.05.005>.

VALBUENA, R.; O'CONNOR, B.; ZELLWEGER, F.; SIMONSON, W.; VIHERRVAARA, P. et al., 2020. Standardizing Ecosystem Morphological Traits from 3D Information Sources. *Trends in Ecology & Evolution*. Roč. 35, č. 8, s. 656-667. ISSN 0169-5347. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.tree.2020.03.006>.

VANDENDAELE, Bastien; FOURNIER, Richard; VEPAKOMMA, Udayalakshmi; PELLETIER, Gaétan; LEJEUNE, Philippe et al., 2021. Estimation of Northern Hardwood Forest Inventory Attributes Using UAV Laser Scanning (ULS): Transferability of Laser Scanning Methods and Comparison of Automated Approaches at the Tree- And Stand-Level: Transferability of Laser Scanning Methods and Comparison of Automated Approaches at the Tree- And Stand-Level. *Remote Sensing*. Dostupné z: <https://doi.org/10.3390/rs13142796>.

WANG, X.; ZHAO, Qingyun; HAN, Feng; ZHANG, Jianxin a JIANG, Ping, 2021. Canopy Extraction and Height Estimation of Trees in a Shelter Forest Based on Fusion of an Airborne Multispectral Image and Photogrammetric Point Cloud. *Journal of Sensors*. Dostupné z: <https://doi.org/10.1155/2021/5519629>.

ZHANG, Zhiyu; LIU, Xinyuan; MA, Yue; XU, Nan; ZHANG, Wenhao et al., 2021. Signal Photon Extraction Method for Weak Beam Data of ICESat-2 Using Information

Provided by Strong Beam Data in Mountainous Areas. *Remote Sensing*. Roč. 13, č. 5.
ISSN 2072-4292. Dostupné z: <https://doi.org/10.3390/rs13050863>.

Seznam obrázků

Obrázek 1: Ukázka normalizovaného bodového mračna obsahujícího strukturu vegetace v absolutních výškách nad terénem (vlevo) a odvozeného rastru CHM (vpravo). Zdroj: Isenburg, 2014	14
Obrázek 2: Schematické znázornění rastrů, které zahrnují nadmořskou výšku terénu (DSM, DTM) a absolutní výšku nad terénem (nDSM, CHM), odvozených z lidarových bodových mračen.	15
Obrázek 3: Sběr a zpracování dat LiDAR: a) Srovnání satelitního a leteckého LiDARu, zahrnující rozdíly v návratových signálech a hustotě bodů. b) Metody zpracování dat: plošné jako ABA (např. CHM), voxelové a objektové segmentace. Zdroj: Bakx et al., 2019.....	16
Obrázek 4: Vlevo – GEDI orbity dohromady za dva týdny, každý obsahuje 8 paprsků. Vpravo – simulované rozložení pozemních drah GEDI po dvou letech (na rovníku, kde bylo 50 % drah odstraněno jako simulace vlivu mraků). Nepravidelnost pozemních drah souvisí s nepravidelností dráhy letu ISS. Zdroj: Dubayah et al., 2020	19
Obrázek 5: Vzor pozemních drah GEDI s parametry rozmístění. Zdroj: Dubayah et al., 2020.....	19
Obrázek 6: Návratová vlna GEDI v NIR spektru, se odrazí od vegetace v oblasti kruhu o průměru ~25 m (vpravo) a následně je z křivky odvozena výška terénu, vegetace (korun stromů) a metriky relativní výšky (RH metriky). Zdroj: Dubayah et al., 2020	20
Obrázek 7: Přehled způsobů LiDAR detekce. Zdroj: Neuenschwander et al., 2022 .	25
Obrázek 8: Schéma rozložení paprsků ICESat-2 při dopředné orientaci přístroje ATLAS. Zdroj: Zhang et al., 2021	27
Obrázek 9: Přehled produktů mise ICESat-2 v jednotlivých úrovních zpracování. Zdroj: NASA, NSIDC DAAC	28
Obrázek 10: Odhad hodnoty bodu z_0 jako váženého průměru z hodnot v jeho okolí Z_1 , Z_2 , Z_3 . Zdroj: (Ježek, 2015).....	34
Obrázek 11: Ukázka vlivu mocniny α na podobě interpolované křivky pomocí IDW (ukázka v 1D).....	35
Obrázek 12: Variogram mrak. Zdroj: Ježek, 2015.....	38
Obrázek 13: Schématické vyjádření výpočtu empirického variogramu. Zdroj: Ježek, 2015.....	39

Obrázek 14: Teoretický variogram s označením základních parametrů. Zdroj: Ježek, 2015.....	40
Obrázek 15: Princip ordinary krigingu (OK). Zdroj: Dobrovolný, 2005c.....	44
Obrázek 16: Princip simple krigingu (SK). Zdroj: Dobrovolný, 2005c	45
Obrázek 17: Princip universal krigingu (UK). Zdroj: Dobrovolný, 2005c.....	45
Obrázek 18: Vybrané území pro stahování dat (minimum bounding box), území pro realizaci cílů práce (KRNAP) a vybrané území 10x10 km pro tvorbu modelu (červený čtverec).....	48
Obrázek 19: Vybrané území pro stahování dat s orbity GEDI. Zdroj: NASA, search.earthdata.nasa.gov.....	50
Obrázek 20: Ukázka agregace rastrového produktu GEDI L2A v Google Earth Engine v porovnání s vektorovým.....	52
Obrázek 21: Vennův diagram zobrazující počty a shodu měření v každém z datasetů	53
Obrázek 22: Vybraný vzorek území o rozměrech 10x10 km s body měření GEDI (modře) a ICESat-2 (červeně)	55
Obrázek 23: Výška vegetace z dat GEDI (RH95) a ICESat-2 (h_canopy) na vybraném území 10x10 km.....	56
Obrázek 24: Vizualizace odchylek měření výšky terénu od referenčního DTM 2012 (vlevo) a graf korelační matice s proměnnou elev_diff v absolutní hodnotě (vpravo)	58
Obrázek 25: Extrakce hodnot prediktorů v původním tvaru stop GEDI (vlevo) a segmentech ICESat-2 (vpravo)	61
Obrázek 26: Ukázka vgm modelů variogramu z balíčku gstat pomocí funkce show.vgms().....	62
Obrázek 27: Znázornění asymptotické časové složitosti (vlevo) a ilustrace 1 km kruhového regionu pro výpočet vah krigingu (vpravo)	64
Obrázek 28: Globální CHM vlevo podle studie Lang et. al (2022) a vpravo podle studie Potapov et. al (2021).	69
Obrázek 29: Naivní filtrace GEDI (quality_flag = 1). 3D model výšky vegetace RH95 (vlevo), vztah výšky terénu GEDI s DTM 2012 (vpravo)	71
Obrázek 30: Pokročilá filtrace GEDI (varianta 7). 3D model výšky vegetace RH95 (vlevo), vztah výšky terénu GEDI s DTM 2012 (vpravo)	71

Obrázek 31: Pokročilá filtrace ICESat-2. 3D model výšky vegetace h_canopy (vlevo), vztah výšky terénu ICESat-2 s DTM 2012 (vpravo)	73
Obrázek 32: Vyška vegetace z GEDI a ICESat-2 před (vlevo) a po pokročilé filtraci (vpravo).....	73
Obrázek 33: Empirické a teoretické variogramy pro jednotlivé varianty tvorby CHM	75
Obrázek 34: Ordinary kriging CHM (10 m) – quality_flag = 1 (vlevo) s nejistotou v rozsahu 6,6 – 9,3 m (vpravo)	76
Obrázek 35: Ordinary kriging CHM (10 m) – pokročilá filtrace (vlevo) s nejistotou v rozsahu 4,9 – 8,5 m (vpravo)	76
Obrázek 36: Universal kriging CHM (10 m) – elev_diff (vlevo) s nejistotou v rozsahu 5,0 – 8,3 m (vpravo).....	77
Obrázek 37: Universal kriging CHM (30 m) – S1 a S2 (vlevo) s nejistotou v rozsahu 4,7 – 5,8 m (vpravo).....	77
Obrázek 38: Porovnání globálních CHM ze studií Lang et al. (2022) a Potapov et al. (2021), vlevo s UK modelem (elev_diff) a vpravo s UK modelem (S1 a S2).....	79

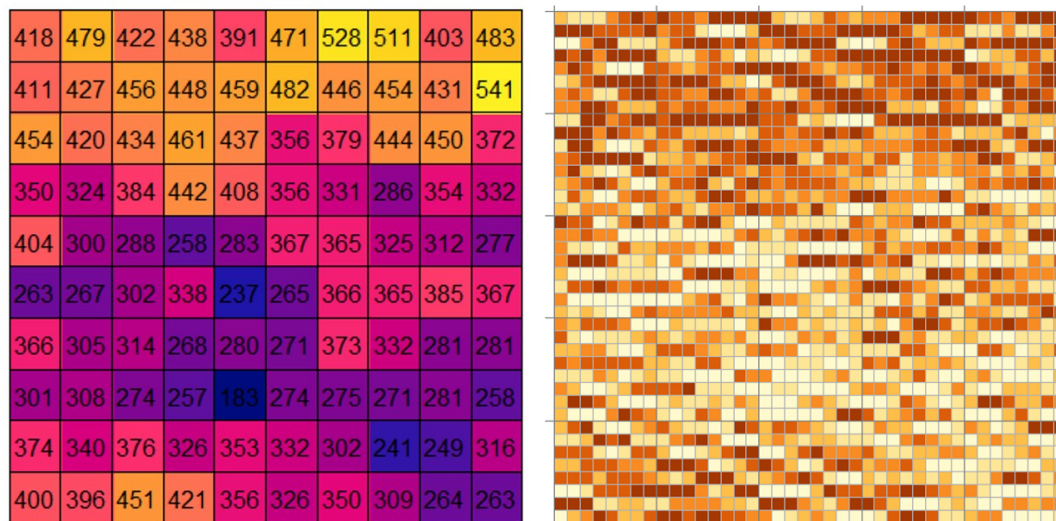
Seznam tabulek

Tabulka 1: Přehled datových produktů mise GEDI. Zdroj: Hofton et al., 2019; Products Overview - GEDI, 2019	21
Tabulka 2: Základní typy krigingu. Upraveno podle: Ježek, 2015	44
Tabulka 3: Porovnání validačního CHM z ALS (2012) s jednotlivými RH metrikami	56
Tabulka 4: Výsledky všech variant filtrací dat z GEDI	72
Tabulka 5: Chyby při křížové validaci u jednotlivých modelů	78
Tabulka 6: Chyby při validaci pomocí CHM z ALS (2012) s lineárním vztahem R^2	78

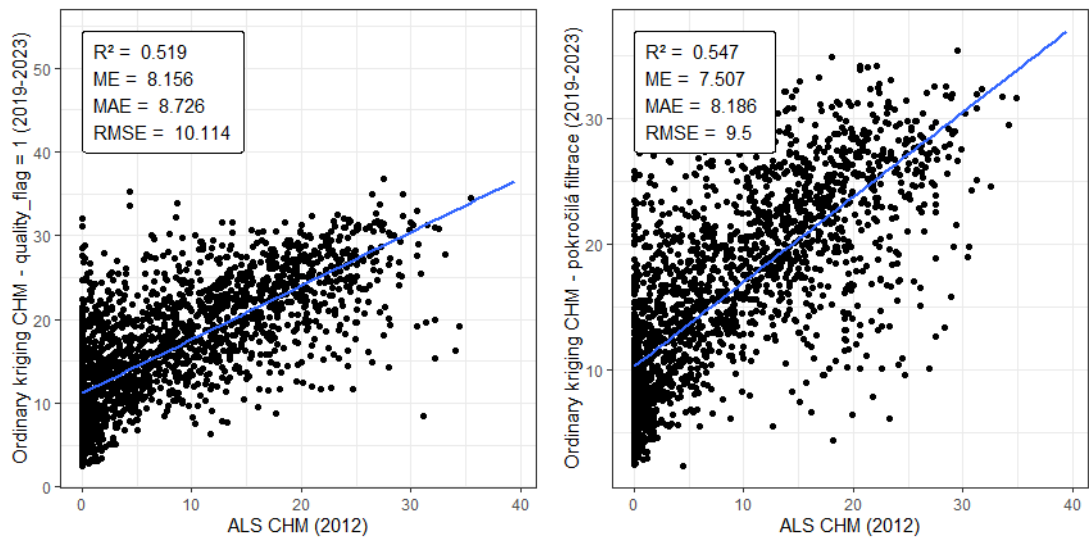
Přílohy

Příloha 1: Výsledná hustota bodů v datové sadě po filtraci v čtvercové mřížce o rozměrech 1x1 km s počty bodů (vlevo), a rozměrech 250x250 km pro porovnání (vpravo).....	102
Příloha 2: Validace OK modelů výšky vegetace s CHM z ALS (2012).....	103
Příloha 3: Validace UK modelů výšky vegetace s CHM z ALS (2012).....	103
Příloha 4: Validace globálních CHM s CHM z ALS (2012)	103
Příloha 5: Sloupcový graf chyb křížové validace u výsledných modelů	104
Příloha 6: Sloupcový graf chyb validace a koeficientu determinace u výsledných modelů.....	104
Příloha 7: Rastry rozdílů výšky vegetace výsledných modelů od CHM z ALS (2012)	105
Příloha 8: Porovnání profilu UK modelu výšky vegetace (elev_diff) s CHM z ALS (2012).....	105
Příloha 9: Porovnání profilu UK modelu výšky vegetace (S1 a S2) s CHM z ALS (2012).....	106
Příloha 10: Vztah kombinované datové sady po pokročilé filtraci s referenčními modely.....	106
Příloha 11: Kolinearita vybraných prediktorů ze satelitů Sentinel-1 a Sentinel-2...	107
Příloha 12: Vybrané prediktory ze satelitu Sentinel-1 a Sentinel-2 se vzájemnou korelací menší než 0,7.....	107

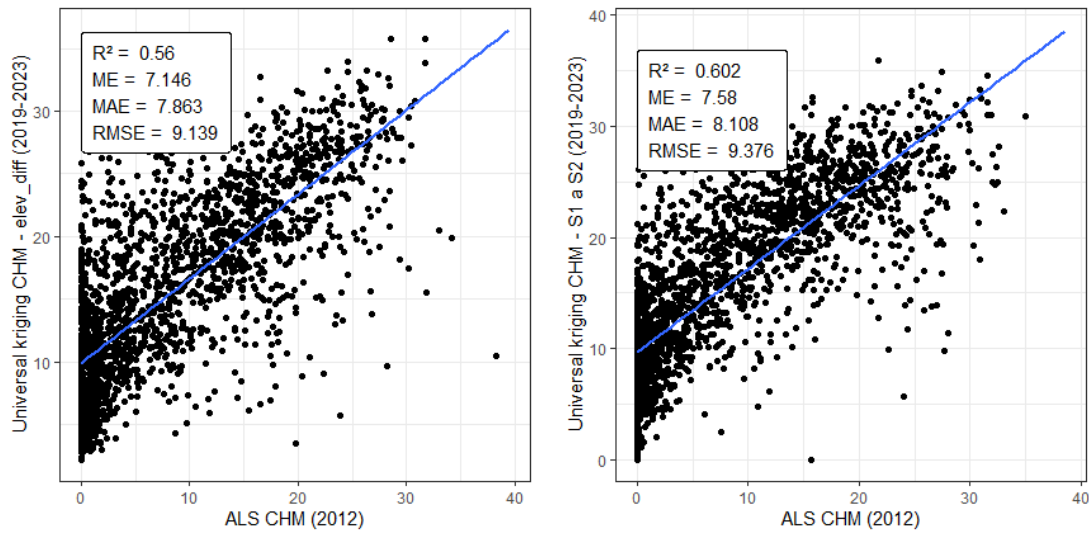
Příloha 1: Výsledná hustota bodů v datové sadě po filtraci v čtvercové mřížce o rozměrech 1x1 km s počty bodů (vlevo), a rozměrech 250x250 km pro porovnání (vpravo)



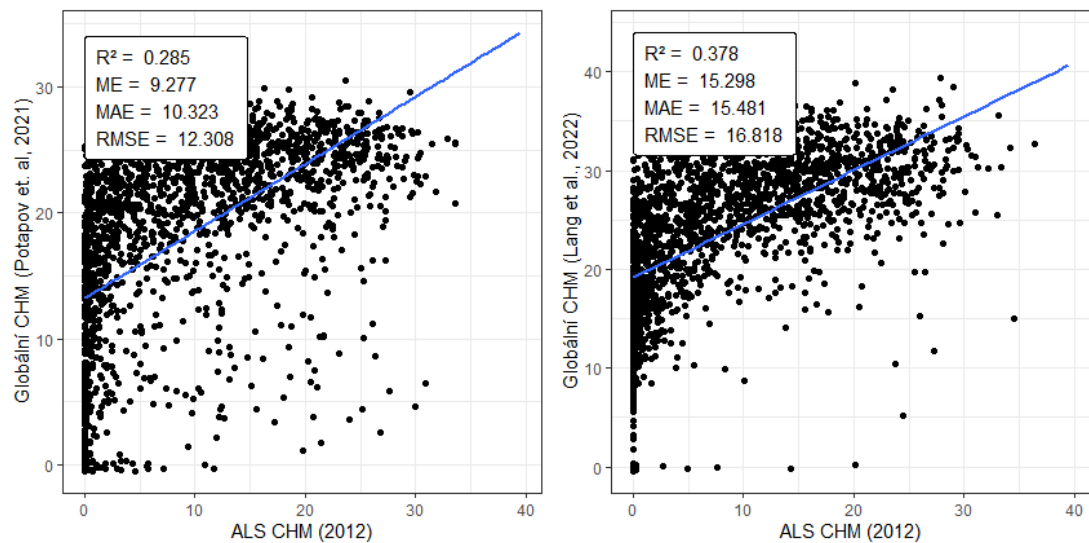
Příloha 2: Validace OK modelů výšky vegetace s CHM z ALS (2012)



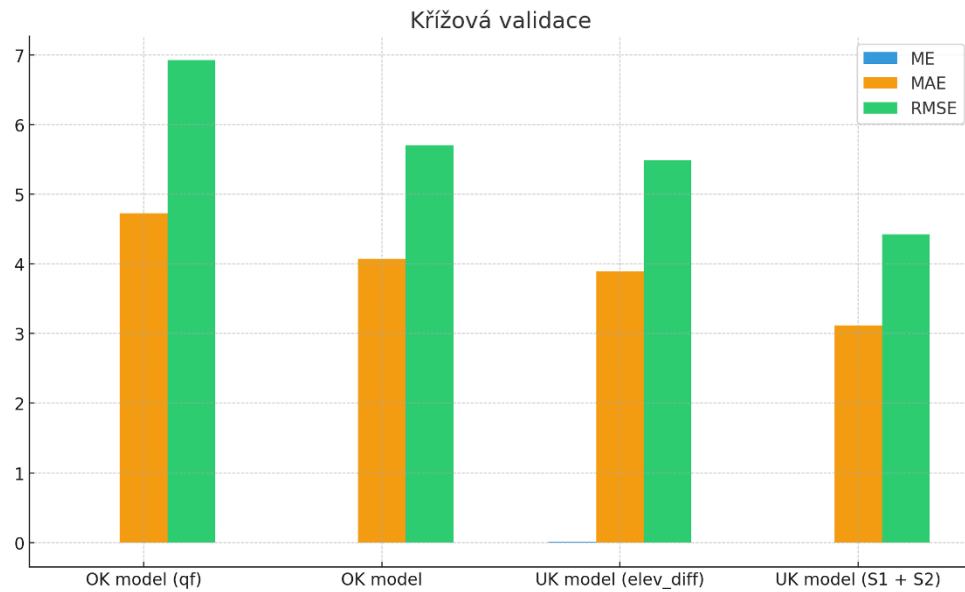
Příloha 3: Validace UK modelů výšky vegetace s CHM z ALS (2012)



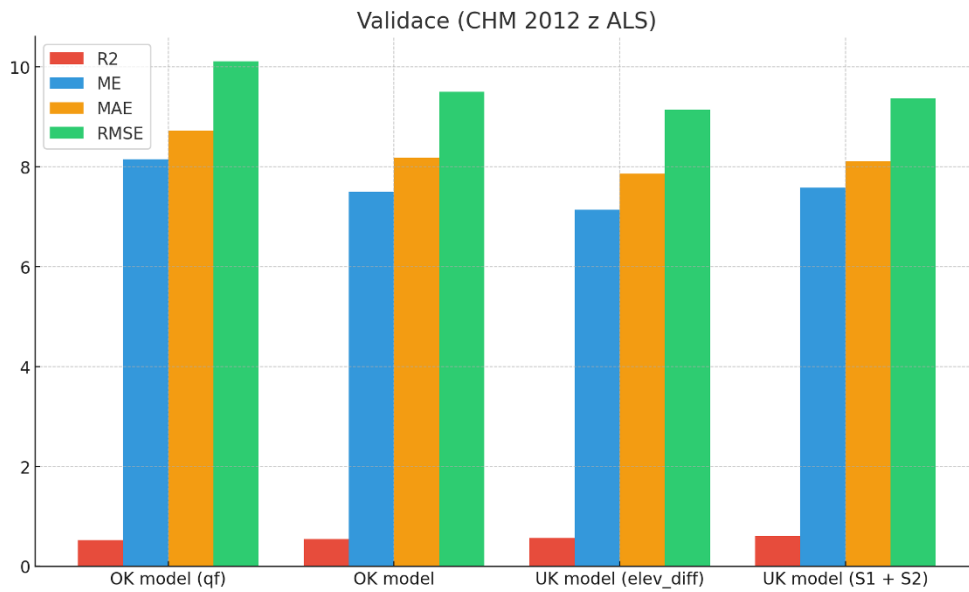
Příloha 4: Validace globálních CHM s CHM z ALS (2012)



Příloha 5: Sloupcový graf chyb křížové validace u výsledných modelů



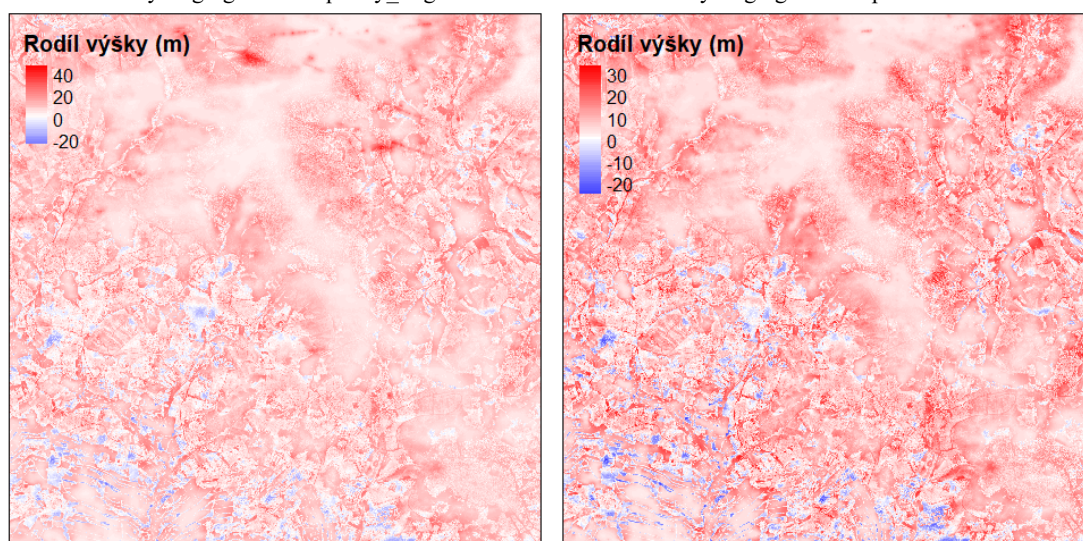
Příloha 6: Sloupcový graf chyb validace a koeficientu determinace u výsledných modelů



Příloha 7: Rastry rozdílů výšky vegetace výsledných modelů od CHM z ALS (2012)

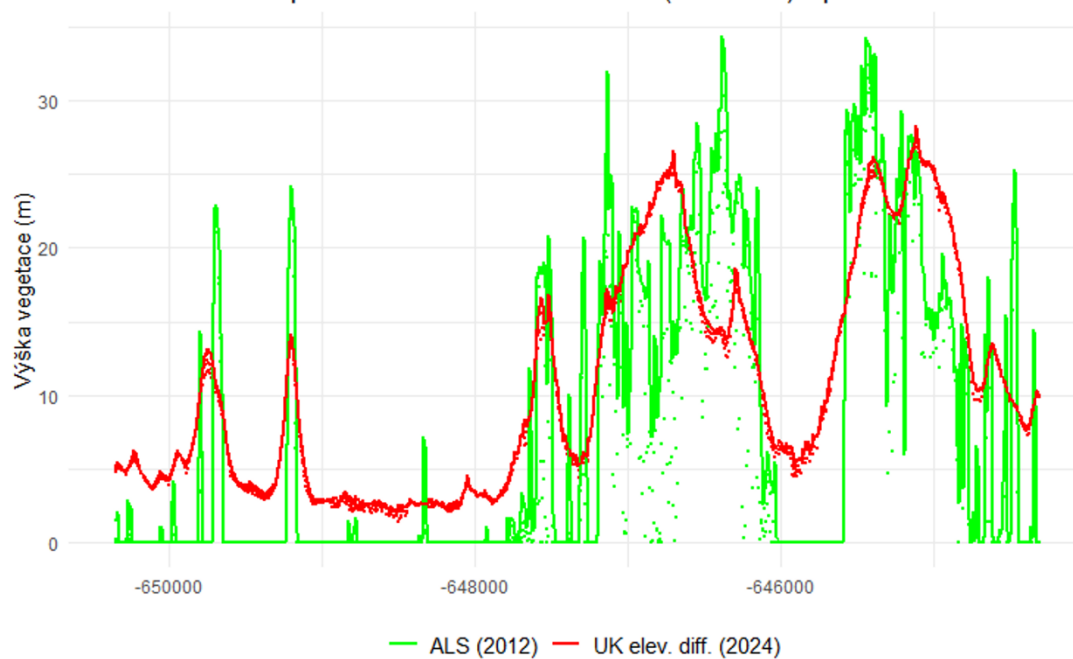
Ordinary kriging CHM – quality_flag = 1

Ordinary kriging CHM – pokročilá filtrace

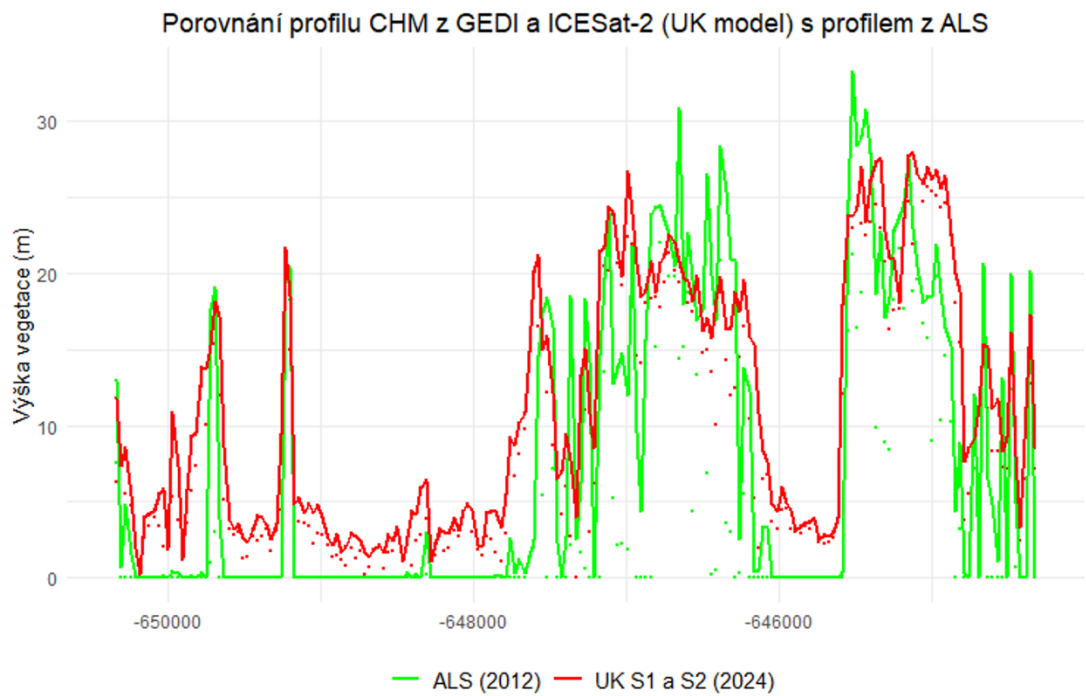


Příloha 8: Porovnání profilu UK modelu výšky vegetace (elev_diff) s CHM z ALS (2012)

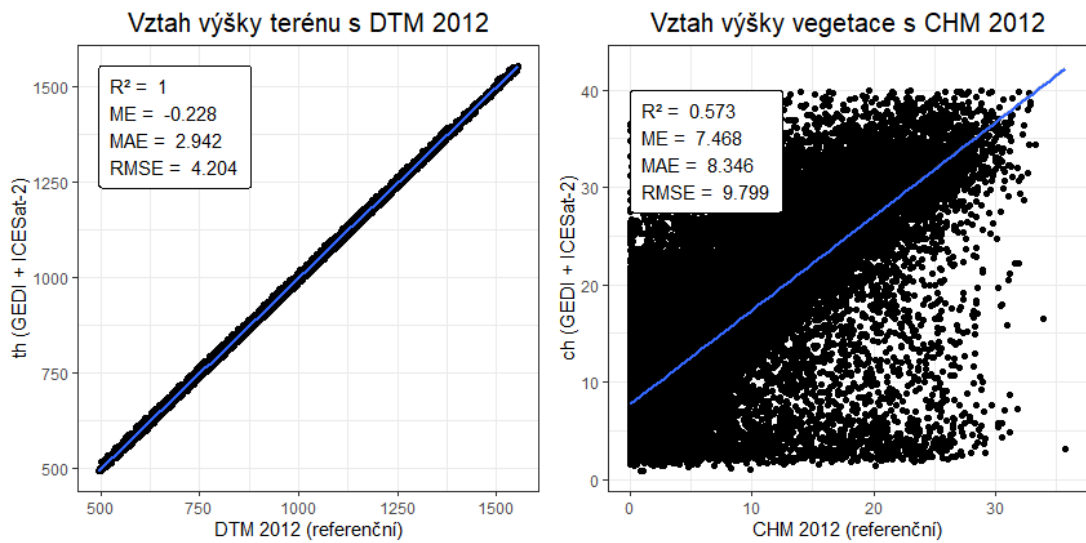
Porovnání profilu CHM z GEDI a ICESat-2 (UK model) s profilem z ALS



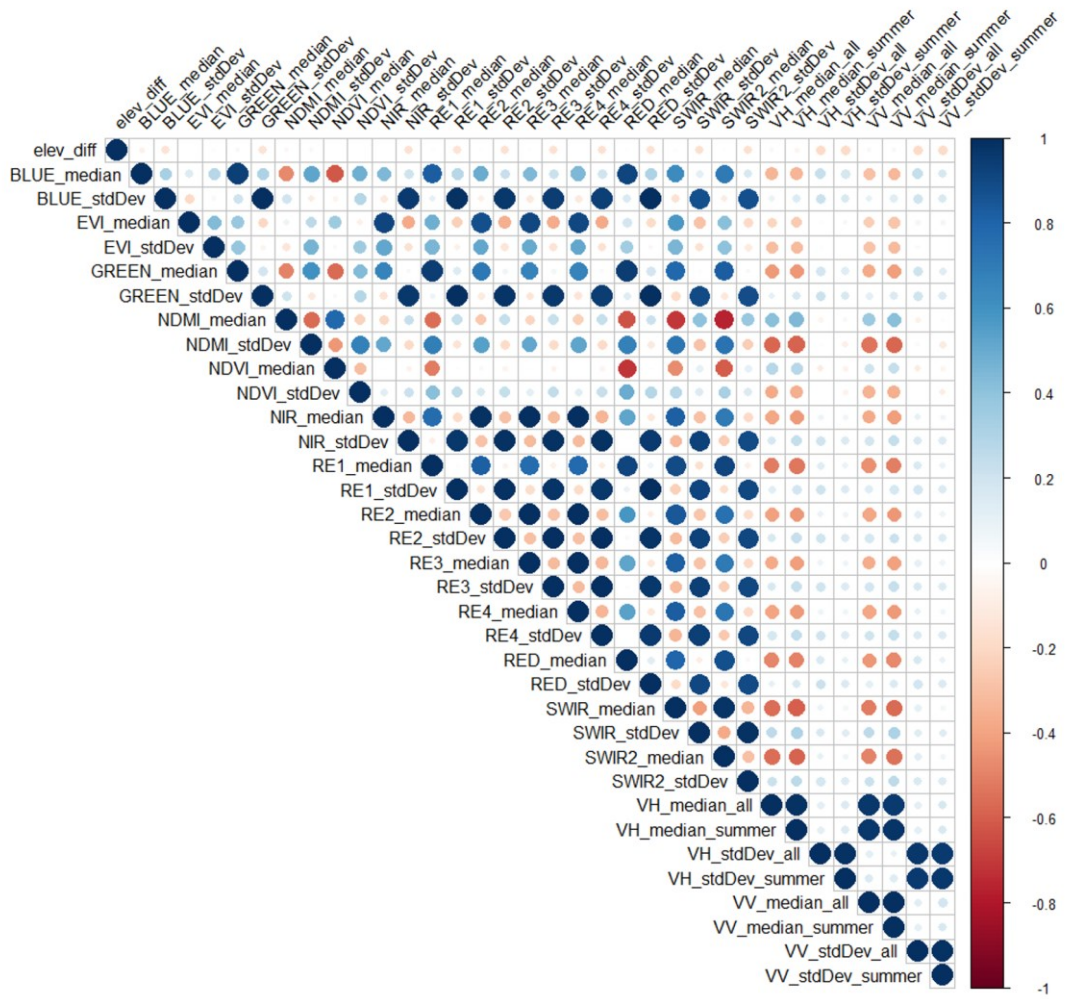
Příloha 9: Porovnání profilu UK modelu výšky vegetace (S1 a S2) s CHM z ALS (2012)



Příloha 10: Vztah kombinované datové sady po pokročilé filtraci s referenčními modely



Příloha 11: Kolinearita vybraných prediktorů ze satelitů Sentinel-1 a Sentinel-2



Příloha 12: Vybrané prediktory ze satelitu Sentinel-1 a Sentinel-2 se vzájemnou korelací menší než 0,7

