

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA PROSTOROVÝCH VĚD

**Vliv chyb ve vstupních datech na kvalitu modelů
druhové distribuce**

DIPLOMOVÁ PRÁCE

Vedoucí práce: doc. Ing. Vítězslav Moudrý, Ph.D.

Diplomant: Bc. Jakub Voska

2024

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Jakub Voska

Prostorové vědy v životním prostředí

Název práce

Vliv chyb ve vstupních datech na kvalitu modelů druhové distribuce

Název anglicky

Effect of input data errors on the performance of species distribution models

Cíle práce

Cílem práce je posoudit vliv chyb ve vstupních datech na kvalitu modelů druhové distribuce.

- Generování virtuálního druhu
- Vytvoření modelů druhové distribuce s daty různé kvality
- Porovnání vytvořených modelů z hlediska predikované presence/absence a pravděpodobnosti výskytu

Metodika

Modelování druhové distribuce (Species Distribution Modelling, SDM) patří mezi oblíbené nástroje v ekologii a biogeografii. Jejich cílem je obvykle najít environmentální proměnné, které mohou mít významný vliv na rozšíření druhu nebo rozšíření druhu predikovat. Nicméně data použitá pro kalibraci těchto modelů mohou mít zásadní vliv na jejich přesnost. Zatímco s využitím reálně sbíraných dat je obtížné tyto vlivy kvantifikovat, generováním tzv. virtuálního druhu lze snadno určit, jaký vliv mají vstupní data na výsledný model. Úkolem autora je zpracovat literární rešerši na téma kvality vstupních dat v modelech druhové distribuce a vytvořit virtuální druh, na jehož příkladu otestuje vliv chyb ve vstupních datech na kvalitu modelu.

Doporučený rozsah práce

30 – 40 stran

Klíčová slova

Formulace klíčových slov je úkolem autora.

Doporučené zdroje informací

- Gábor, L., Jetz, W., Lu, M., Rocchini, D., Cord, A., Malavasi, M., ... & Moudrý, V. (2022). Positional errors in species distribution modelling are not overcome by the coarser grains of analysis. *Methods in Ecology and Evolution*, 13(10), 2289-2302.
- Gábor, L., Jetz, W., Zarzo-Arias, A., Winner, K., Yanco, S., Pinkert, S., ... & Moudrý, V. (2023). Species distribution models affected by positional uncertainty in species occurrences can still be ecologically interpretable. *Ecography*, e06358.
- Gábor, L., Moudrý, V., Lecours, V., Malavasi, M., Barták, V., Fogl, M., ... & Václavík, T. (2020). The effect of positional error on fine scale species distribution models increases for specialist species. *Ecography*, 43(2), 256-269.
- Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F. (2016). virtualspecies, an R package to generate virtual species distributions. *Ecography*, 39(6), 599-607.
- Moudrý, V. (2015). Modelling species distributions with simulated virtual species. *Journal of Biogeography*, 42(8), 1365-1366.
-

Předběžný termín obhajoby

2023/24 LS – FŽP

Vedoucí práce

doc. Ing. Vítězslav Moudrý, Ph.D.

Garantující pracoviště

Katedra prostorových věd

Elektronicky schváleno dne 6. 3. 2024

prof. Ing. Petra Šimová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 7. 3. 2024

prof. RNDr. Michael Komárek, Ph.D.

Děkan

V Praze dne 09. 03. 2024

Čestné prohlášení

Prohlašuji, že jsem diplomovou práci na téma Vliv chyb ve vstupních datech na kvalitu modelů druhové distribuce 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 Praze dne 28.3.2024

Podpis:

Poděkování

Velké poděkování patří doc. Ing. Vítězslavu Moudrému, PhD. za jeho vstřícnost, ochotu a cenné rady a připomínky při vedení diplomové práce. Děkuji také rodině za podporu při studiu.

Abstrakt

Modely druhové distribuce jsou oblíbeným nástrojem používaným v ekologii a ochraně přírody. Jako každé modely jsou i tyto zatížené chybami, které jsou součástí vstupních dat. Konkrétně se jedná o data o druhu a data o prostředí, také označovány jako prediktory. V této práci bylo testováno, jak polohová chyba dat o výskytu druhů ovlivňuje kvalitu modelů druhové distribuce a zda míra ovlivnění modelů touto chybou souvisí s kvalitou prediktorů. Těmi byly 3 modely výšky vegetace, 2 byly vytvořeny z dat leteckého laserového skenování s 10letým rozdílem mezi nimi a jeden byl predikovaný globální model z dat GEDI. V neposlední řadě byl také posouzen vliv odlišné metody simulování polohové chyby na modely druhové distribuce. Testování probíhalo s využitím tzv. virtuálního druhu, jehož pravděpodobnost výskytu a z ní odvozená prezence/absence závisela na aktuální výšce vegetace z roku 2022. Z vytvořeného druhu byly dále získány sady nálezových dat, u kterých byly dvěma způsoby simulovány různé úrovně polohové chyby. Pro každou úroveň polohové chyby a model výšky vegetace byl vytvořen model predikující pravděpodobnost výskytu druhu. U těchto modelů byla následně hodnocena jejich predikční schopnost, a to běžně používanými metrikami, jako jsou AUC, Warrenův index I a Schoenerův index D. Z těch vyplývá, že polohová chyba v datech má na přesnost modelů zásadní vliv. Ten ale závisí na kvalitě prediktorů, které do modelu vstupují. Prediktory vysoké kvality produkují přesné modely, které jsou však polohovou chybou ovlivněny výrazněji než modely založené na prediktorech s vyšší mírou neurčitosti. U nich narůstající chyba nezpůsobuje natolik zásadní snížení přesnosti. Vyloučeno tak bylo synergické působení obou chyb současně. To otevírá diskuzi o tom, zda je nutné vybírat pouze polohově přesná data, pokud víme, že dostupné prediktory jsou omezené kvality. Nakonec bylo zjištěno, že odlišná metoda simulování polohové chyby nevede k rozdílným výsledkům.

Klíčová slova: modely druhové distribuce, polohová chyba, virtuální druh, vstupní data, AUC

Abstract

Species distribution models are a popular tool used in ecology and nature conservation. Like all models, these are also burdened with errors that are part of the input data. Specifically, these include species and environmental data, also known as predictors. This study aimed to test how positional error in species occurrence data affects the quality of species distribution models and whether the degree of influence of this error on models is related to the quality of predictors. Three canopy height models were used, two created from airborne laser scanning data with a 10-year difference between them, and one predicted global model from GEDI data. Additionally, the impact of different methods of simulating positional error on species distribution models was assessed. Testing was conducted using a so-called virtual species, whose probability of occurrence and derived presence/absence depended on the current vegetation height from the year 2022. Data sets of occurrences were then obtained from this created species, with positional error simulated in two ways. For each level of positional error and canopy height model, a model predicting the probability of species occurrence was created. The predictive ability of these models was subsequently evaluated using commonly used metrics such as AUC, Warren's I, and Schoener's D. It was found that positional error in the data significantly impacts the accuracy of models. However, this impact depends on the quality of predictors entering the model. High-quality predictors produce precise models, but they are significantly more affected by positional error than models based on predictors with higher uncertainty. For the latter, increasing error does not cause such a significant decrease in accuracy. Thus, the synergistic effect of both errors simultaneously was ruled out. This opens up a discussion about whether it is necessary to select only spatially accurate data if we know that the available predictors are of limited quality. Finally, it was found that a different method of simulating positional error does not lead to different results.

Keywords: species distribution models, positional error, virtual species, input data, AUC

Obsah

1	Úvod a cíle práce.....	1
2	Literární rešerše.....	3
2.1	Modely druhové distribuce	3
2.2	Problémy vstupních dat.....	5
2.2.1	Data o druhu.....	6
2.2.2	Data o prostředí	8
2.3	Virtuální druhy	9
3	Metodika	11
3.1	Použitá data	11
3.2	Generování virtuálního druhu	13
3.3	Simulace polohové chyby	17
3.4	Tvorba modelů	20
4	Výsledky	22
5	Diskuze.....	28
6	Závěr	31
7	Přehled literatury a použitých zdrojů	32
8	Přílohy	40

1 Úvod a cíle práce

Mezi hlavní úlohy ekologie patří zkoumání toho, kde se organismy vyskytují a jaké podmínky ke svému výskytu potřebují. Tyto informace jsou velmi důležité, aby bylo možné například chránit habitaty vzácných druhů. Vznikly tak příbuzné obory, jako je biogeografie, jejichž úkolem je tyto vztahy mezi druhem a prostředím odhalovat (Lomolino 2004). Zatímco jejich popisování probíhalo v počátcích výhradně kvalitativně, později byly vyvinuty numerické modely, a právě ty jsou dnes široce používané. Označují se jako druhové distribuční modely, které se snaží na základě omezených pozorování výskytu druhů organismů a enviromentálních proměnných predikovat jejich výskyt v prostoru (Elith a Leathwick 2009). Tato vstupní data modelů jsou však velice často zatížena chybami, které snižují jejich přesnost a ty tak nemusí výskyt druhů predikovat správně. Jednou z takových chyb je polohová chyba, neboli nepřesné určení souřadnic při zaznamenání dat o výskytu druhu, které může vzniknout lidskou chybou nebo například špatným příjmem signálu navigačních přístrojů (Gábor et al. 2020b).

Právě polohová chyba a její vliv na kvalitu modelů druhové distribuce jsou v dnešní době intenzivně zkoumány (Moudrý a Šimová 2012; Naimi et al. 2014; Smith et al. 2023; Tulowiecki et al. 2015). Vznikají také studie, které zjišťují, jak i jiné vlastnosti vstupních dat mění efekt polohové chyby. Těmi jsou například velikost vzorku nálezových dat (Mitchell et al. 2017), šířka niky zkoumaného druhu (Gábor et al. 2020b) nebo rozlišení enviromentálních proměnných (Gábor et al. 2022; Hanberry 2013). Velké množství těchto prací při tom nevyužívá skutečné záznamy o výskytu druhu, které se k hodnocení vlivu polohové chyby využívají obtížně, ale generují tzv. virtuální druh, který umožňuje simulovat výskyt podle jeho vztahu s enviromentálními faktory a kontrolovat další vlastnosti druhu (Meynard a Kaplan 2013).

Dalším aspektem ovlivňujícím přesnost modelů druhové distribuce je kvalita prediktorů. V praxi se stává, že i při výběru nesprávných prediktorů, které nemusí mít na distribuci druhu žádný vliv, může být výsledkem model s vysokou přesností. Takový model ale rozhodně nebude reflektovat realitu (Fourcade et al. 2018). Aby bylo možné modelovat distribuci druhů kdekoli na světě, vznikla potřeba mít prediktory s globálním pokrytím. Ty však mají svá úskalí, často například nemají potřebné rozlišení, kvůli čemuž v některých případech nemusí být efektivní (Williams

et al. 2012). Také mohou být velmi nepřesné, což bývá někdy opomíjeno. Bylo ale dokázáno, že přesnost stejné enviromentální proměnné se může z různých zdrojů velmi odlišovat (Moudrý et al. 2018).

Tato práce navazuje na předchozí studie zabývající se vlivem polohové chyby a vlastnostmi vstupních dat na přesnost modelů druhové distribuce. Cílem práce je zhodnotit vliv (i) polohové chyby a (ii) prediktorů omezené přesnosti na kvalitu modelů a zejména jejich vzájemnou (iii) interakci. Jinými slovy jde o vyhodnocení, zda dochází k synergickému efektu a obě chyby se sčítají nebo při existenci jedné (např. nízká kvalita prediktorů), již nedochází k dalšímu zhoršení modelu bez ohledu na velikost chyby druhé (nízká polohová přesnost pozorování). Cíle bylo dosaženo generováním virtuálního druhu, následným simulováním polohové chyby v datech, vytvořením modelů druhové distribuce se vstupními daty odlišné kvality a porovnáním vzniklých modelů z hlediska predikované prezence/absence a pravděpodobnosti výskytu.

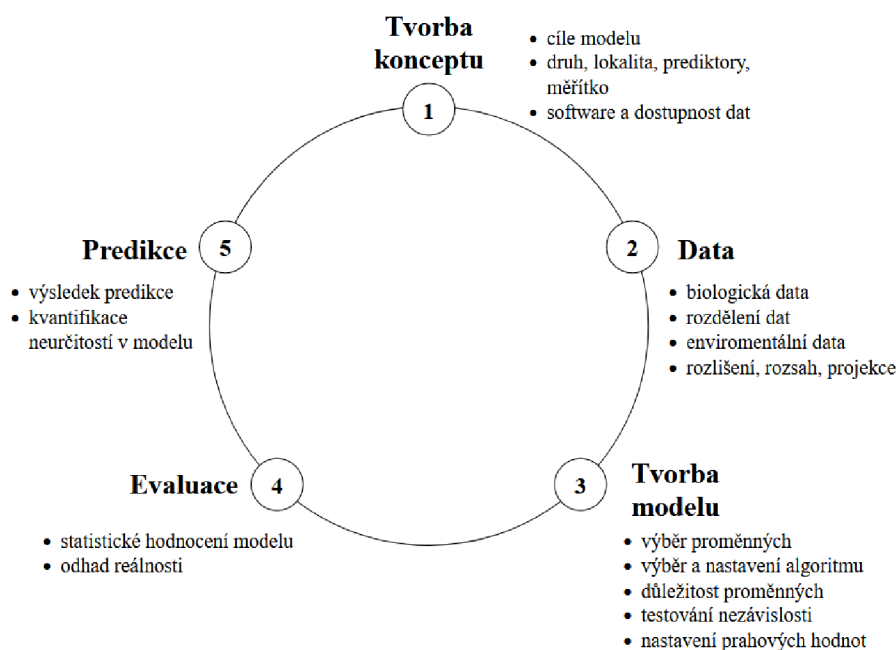
2 Literární rešerše

2.1 Modely druhové distribuce

Modely druhové distribuce (Species distribution models – SDMs) jsou populárním nástrojem zejména v ekologii a ochraně biodiverzity. Využívají vztahu mezi enviromentálními proměnnými a distribucí druhů organismů v prostoru. Tento vztah kvantifikují a vyjadřují jím důležitost určitých faktorů pro výskyt druhů a také pomocí něj predikují distribuci druhů v oblastech, ve kterých data o výskytu nemáme (Miller 2010). Dnešní možnosti ve smyslu dostupnosti dat, software a návodů mají za důsledek, že míra popularity distribučních modelů stále roste (Arenas-Castro et al. 2022). Využívají se například také k detekci nových populací vzácných druhů (Claerhout et al. 2023), zkoumání evolučních procesů (Kozak et al. 2008) nebo identifikaci kritických habitatů (Volis a Tojibaev 2021).

Tyto modely typicky vyžadují 2 druhy vstupních dat, a to biologická, která popisují známou distribuci druhů a enviromentální, také označované jako prediktory (Miller 2010). Modelování druhové distribuce je komplexní proces, který se skládá z řady kroků. Podle Elith a Leathwick (2009) začíná ve výběru relevantních dat a rozhodnutí o jejich zakomponování do modelu. Také se určí jejich přesnost a úplnost. Na straně prediktorů se může stát, že mezi některými proměnnými bude vysoká míra korelace, a je tedy důležité se rozhodnout, jak s takovými daty naložit. Dále se vybere vhodná metoda k tvorbě modelu. Existuje velké množství algoritmů a jejich dělení se často provádí podle struktury vstupních dat na a) pouze prezenční a b) prezenční/absenční algoritmy. Další dělení se odvíjí od samotné charakteristiky modelu, například na regresní (zobecněný lineární model, zobecněný aditivní model), distanční (BIOCLIM, ENFA) nebo algoritmy založené na strojovém učení (Maximum Entropy). Je dokázáno, že univerzální metoda, která by byla vhodná pro všechny modely, neexistuje. Proto je výběru vhodného algoritmu přikládána velká důležitost a dobře zvolený algoritmus může zmírnit neurčitosti ve výsledném modelu (Pecchi et al. 2019). Elith a Leathwick (2009) dále popisují, že po volbě modelovacího algoritmu následuje jeho aplikace na trénovací data, složená ze vzorku biologických a enviromentálních dat. V této fázi dochází k nastavení parametrů algoritmu a definování vztahu mezi druhem a prostředím. Poté se získaný model použije k predikci distribuce druhu a jeho zanesení do prostoru. Posledním krokem je evaluace modelu z hlediska jeho schopnosti predikovat druhovou distribuci a rozlišovat relevantní

enviromentální proměnné. Bylo navrženo vícero protokolů, které všechny tyto kroky zahrnují a podle nichž by se tvorba modelů druhové distribuce měla řídit. Jedním z nich je i protokol ODMAP (Overview, Data, Model fitting, Assessment, Predictions) (*Obrázek 1*), navržený jako návod, ale i jako formát pro zveřejňování výsledků a diskuzi nad nimi, snadnější kontrolu a reprodukovatelnost (Zurell et al. 2020).



Obrázek 1: Protokol ODMAP (Zurell et al. 2020)

Ačkoliv jsou modely druhové distribuce mocným nástrojem, při nesprávném použití mohou vést ke zkresleným, či zcela nepravdivým výsledkům. Bylo dokázáno, že predikční schopnost se mění v závislosti na požadovaném výstupu modelu. Tím je často výskyt sledovaného druhu, ale modely se používají také k predikci abundance, populačního růstu nebo genetické diverzity. Zatímco predikce výskytu bývá většinou úspěšná, u ostatních metrik kvalita modelů postupně klesá (A. Lee-Yaw et al. 2022). Je také důležité nepřehlížet mnohá doporučení a vyvarovat se porušení klíčových předpokladů tvorby metodicky správných modelů. Často chyby vznikají již ve výběru vstupních dat, kdy se například používají nesprávně georeferencovaná data nebo nehomogenně sesbíraná data napříč územím (sampling bias). Chybná ale může být i evaluace a interpretace modelu (Soley-Guardia et al. 2024).

2.2 Problémy vstupních dat

Jak již bylo zmíněno, vstupní data druhových distribučních modelů lze rozdělit na data o druhu (biologická) a data o prostředí (enviromentální). Data o druhu jsou většinou ve formě prezenčních, či prezenčně/absenčních záznamů o výskytu druhu. Data o prostředí se dále rozlišují na klimatická, topografická a ostatní. Nejčastěji mají podobu prostorových rastrových vrstev vytvořených nástroji geografických informačních systémů (GIS) (Miller 2010). Obě tyto skupiny vstupních dat mají určité vlastnosti a chyby (*Tabulka 1*), které ovlivňují výslednou kvalitu modelu. Určit dopady jejich použití na model není snadné a takřka všechny modely budou těmito chybami ovlivněny.

Data o druhu		Data o prostředí	
Chyba/Vlastnost	Reference	Chyba/Vlastnost	Reference
Polohová chyba	(Gábor et al. 2023; 2022; 2020b; Naimi et al. 2014; Tulowiecki et al. 2015)	Neurčitost v datech	(Fernandez et al. 2013; Stoklosa et al. 2015)
Sampling bias	(Boria et al. 2014; Kramer-Schadt et al. 2013; Soley-Guardia et al. 2024; Monsarrat et al. 2019)	Měřítko	(Connor et al. 2019; Franklin et al. 2013; Moudrý a Šimová 2012)
Velikost výběru	(Hernandez et al. 2006; Gábor et al. 2020a; Moudrý a Šimová 2012)		

Tabulka 1: Přehled vlastností a chyb dat o druhu a dat o prostředí

2.2.1 Data o druhu

Nálezová data o druzích mohou přímo obsahovat několik možných chyb. Mezi ně patří polohová chyba (Gábor et al. 2020b) nebo sampling bias (Kramer-Schadt et al. 2013). Nikoliv již chybou, ale vlastností těchto dat je velikost výběru, tedy počet záznamů v nálezových datech. Ta výslednou kvalitu modelů také ovlivňuje (Moudrý a Šimová 2012).

Polohová chyba

Polohová chyba je nepřesnost v určení geografické polohy lokality, kde byl druh pozorován. Jedná se o odchylku mezi zaznamenanou a skutečnou polohou druhu. Dříve se předpokládalo, že vliv takové chyby na model je pouze velice malý nebo že se vyskytuje jen u starších záznamů, jejichž poloha byla odvozena z historických písemných popisů výskytu. Ale polohová nepřesnost se vyskytuje i u nových záznamů pořízených moderními navigačními přístroji a technikami. Zde může být chyba způsobena použitím neaktuální technologie, špatným příjmem signálu z navigačních družic nebo finálním zpracováním naměřených dat (Gábor et al. 2020b). I v současnosti je například přetrvávajícím problémem to, že u většiny záznamů není k dohledání informace o úrovni neurčitosti spojené s identifikací a georeferencí. Chyby také vznikají u převodů jednotek, ve kterých jsou souřadnice naměřeny a přílišně velký počet desetinných míst u souřadnic naopak falešně implikuje vysokou přesnost záznamu (Soley-Guardia et al. 2024). Často také dochází k záměně polohy pozorovatele a druhu při zaznamenání souřadnic.

Polohová chyba může vést k nesprávnému určení environmentálních podmínek v místech výskytu druhu, což může způsobit nadhodnocení nebo podhodnocení vhodných habitatů. To poté při samotné tvorbě modelu vede k nepřesné predikci distribuce druhu. Vliv polohové chyby je obzvláště významný v oblastech s rychle se měnícími podmínkami, jako jsou horské oblasti. Její dopad se také zvyšuje s jemnějším prostorovým rozlišením environmentálních dat v modelech (Moudrý a Šimová 2012). Proto je důležité při modelování druhové distribuce polohovou chybu zohlednit a, pokud je to možné, minimalizovat ji pomocí vhodných metod, jako je například filtrace záznamů o výskytu druhu (Soley-Guardia et al. 2024). Další způsob jak omezit vliv polohové chyby, přinesla studie Gábora et al. (2022), kteří se o to pokusili zhoršením rozlišení environmentálních dat. Předpoklad byl takový, že modely

používající enviromentální data s vysokým prostorovým rozlišením jsou polohovou chybou v datech o druzích ovlivněny více, protože jednotlivé záznamy se mohou snáze objevit v nevyhovujících podmínkách. Výsledky ale ukázaly, že takový přístup kvalitu modelů nezvýšil.

Sampling bias

Sampling bias je systematická chyba ve sběru dat, která nastává, když metoda sběru přesně neodráží studovanou populaci. Takový vzorek populace není příkladem náhodného výběru, což může vést k zavádějícím či nesprávným závěrům. V kontextu modelování druhové distribuce to může znamenat, že dojde k chybnému určení rozšíření druhu nebo jeho enviromentálních podmínek. K takové odchylce dojde, pokud jsou data o druhu z určitých oblastí sbírána častěji než z jiných, například kvůli jejich snazší přístupnosti nebo ekologické atraktivitě. Dále může být způsobena selektivním sběrem dat o určitých druzích na úkor jiných, což může být ovlivněno například jejich viditelností, rozpoznatelností či stupněm ochrany. Díky tomu dochází k nadhodnocení nebo naopak podhodnocení zastoupení druhu v lokalitě (Monsarrat et al. 2019). Úplné odstranění vlivu sampling bias zůstává výzvou. Jeden z takových přístupů se ho snaží redukovat filtrováním shlukovaných záznamů (Boria et al. 2014).

Velikost výběru

Velikost výběru je v modelování druhové distribuce jednoduše počet jednotlivých pozorování konkrétního studovaného druhu. Ten se může pohybovat od desítek až po tisíce záznamů. Problémem je, že v globálních databázích, které bývají hlavním zdrojem dat o druzích, jsou některé druhy pozorovány velice málo. Zejména to jsou druhy vzácné a ohrožené, tedy ty nejdůležitější z hlediska ochrany přírody. Příliš malý počet záznamů pak způsobuje, že je velice obtížné s takovými daty vytvořit spolehlivý model (Gábor et al. 2020a). Existuje vztah, který říká, že čím více enviromentálních proměnných se při tvorbě modelu používá a čím více jsou vztahy druhů k těmto proměnným komplexní, tím více záznamů v datech o druzích je potřeba k vytvoření spolehlivého modelu (Moudrý a Šímová 2012). Některé studie ale dokázaly, že kvalitní modely se dají vytvořit i s malým počtem záznamů, a to hlavně u druhů, které mají malý areál rozšíření a tolerují jen úzký rozsah enviromentálních podmínek, což je typické pro druhy vzácné (Hernandez et al. 2006).

2.2.2 Data o prostředí

Data o prostředí, častěji spíše enviromentální data, také obsahují určité druhy chyb a neurčitostí. Často se ovšem stává, že jsou tyto chyby při tvorbě modelů druhové distribuce přehlíženy a uživatelé se spíše soustředí na kvalitu dat o druzích. To přitom může vést ke zkresleným odhadům a špatné kvalitě modelů (Stoklosa et al. 2015). Dalším parametrem těchto dat, který je nutné brát v potaz, je doba, k níž se vztahují. Je vhodné, aby mezi dobou pořízení dat o prostředí a dat o druhu nebyl velký rozdíl, zejména pak u enviromentálních proměnných, které se za tuto dobu mohou změnit. Zohlednit je třeba i časový rozdíl mezi samotnými daty o prostředí. Příliš studií se tomuto problému však nevěnuje a podle dosavadních výsledků nemá časový rozdíl na kvalitu modelů příliš velký vliv (Vierling et al. 2014). Problémy při modelování druhové distribuce také může způsobit měřítko dat o prostředí (Moudrý a Šimová 2012).

Neurčitost v datech

Neurčitosti v enviromentálních datech často vznikají při jejich tvorbě. Například klimatická data jsou většinou získávána pomocí meteorologických stanic a jednotlivá měření jsou k dispozici pouze v těchto diskrétních bodech. Aby taková data byla použitelná v modelování druhové distribuce, využívá se interpolace diskrétních bodových měření do spojitě rastrové vrstvy, která obsahuje hodnoty v každém místě mezi bodovými měřeními. Ale jelikož jsou tyto hodnoty pouhými odhady skutečného klimatu v dané lokalitě, mají tato data určitou úroveň neurčitosti, kterou je možné také vyjádřit v prostoru a lze díky tomu odhalit místa náchylná k nadhodnocování a podhodnocování. Pokud jsou tedy informace o neurčitosti v klimatických datech k dispozici, je vhodné je do tvorby modelů zahrnout a zlepšit tak jejich kvalitu (Stoklosa et al. 2015).

Měřítko

Prostorové měřítko enviromentálních dat je dáno jejich prostorovým rozlišením a rozsahem. Bylo zjištěno, že horší rozlišení enviromentálních dat způsobuje špatnou predikční schopnost modelů zkoumajících distribuci vzácných druhů. Ty se mohou například vyskytovat pouze v malém rozsahu podmínek, které nemusejí být v datech s horším rozlišením patrné. Obecně tedy mají lepší kvalitu modely založené na enviromentálních datech s vysokým prostorovým rozlišením

(Moudrý a Šimová 2012). Zkoumané bylo také vyvážení mezi oběma aspekty měřítka, tedy rozlišením a rozsahem. Snižováním prostorového rozlišení se kvalita modelů zhoršovala, ale současné zvětšování geografického rozsahu tento efekt potlačovalo (Connor et al. 2019).

2.3 Virtuální druhy

Jak již bylo zmíněno, modely druhové distribuce jsou velice oblíbeným nástrojem používaným například v makroekologii a biogeografii. Problémy však mohou nastat při jejich testování, ať už je cílem zjistit jejich přesnost nebo odhalit některé problémy vzniklé působením chyb a vlastnostmi vstupních dat (Grimmett et al. 2021). Těmi může být například polohová chyba nebo sampling bias. Odhalit takové problémy se skutečnými daty pořízenými v terénu je často nemožné, a tak vzniká snaha je překonat za pomoci simulací. Díky tomu byl vytvořen koncept virtuálního druhu, u kterého je jeho distribuce simulována uměle definovaným vztahem mezi ním a enviromentálními proměnnými. Virtuální druhy také umožňují ovládat další parametry spojené s jeho distribucí, jako je například druhová prevalence (frekvence míst, na kterých se druh vyskytuje), typ vztahu mezi druhem a prostředím a jeho odezva na něj nebo počet vytvořených nálezů. Těchto parametrů je celá řada a všechny nám dovolují simulovat druh podle konkrétních potřeb (Meynard a Kaplan 2013).

Obecný postup pro vytváření virtuálního druhu shrnují Meynard et al. (2019). Podle nich se dá rozdělit do 5 fází, přičemž poslední fáze zahrnuje evaluaci a testování modelu, která je shodná jako při použití skutečných nálezových dat o druhu. V první fázi skutečné tvorby virtuálního druhu je definován již zmiňovaný vztah mezi druhem a prostředím. Tento vztah lze vyjádřit pomocí tzv. response křivky, která představuje zjednodušenou ekologickou niku druhu. Tato křivka má může mít téměř jakýkoliv tvar a může se měnit podle typu enviromentální proměnné. Takto se vytvoří response funkce, která převádí hodnoty prediktorů na hodnoty vhodnosti prostředí. Tento krok se může zařadit do druhé fáze tvorby virtuálního druhu, kdy se simulovaný vztah aplikuje do prostoru tvořeným enviromentálními proměnnými. Ty mohou být skutečné, nebo je také možné je nasimulovat, což může přinést některé výhody, pokud je cílem například zkoumat vliv vlastností dat o prostředí.

Třetím krokem je převedení získaných hodnot vhodnosti prostředí na distribuci prezencí a absencí. K tomu lze využít 2 přístupů, přičemž každý z nich funguje na odlišném principu. Prvním z nich je převod s použitím prahové hodnoty. Toto byla původní implementace převedení vhodnosti prostředí na prezenci/absenci. Zvolí se prahová hodnota ve funkci vhodnosti a hodnoty nad tímto prahem budou považovány za prezenci a hodnoty pod ním za absenci. Tento přístup byl oblíbený zejména pro svoji jednoduchost a rychlost, ale jeho použití není příliš vhodné (Meynard a Kaplan 2013). Takto se totiž degraduje komplexní funkce vhodnosti prostředí na funkci, kde jsou hodnoty pouze 0 (pod prahem) a 1 (nad prahem). Mnohem vhodnější je tedy druhý přístup, označovaný jako pravděpodobnostní. Zde se pomocí logistické funkce převede vhodnost prostředí na pravděpodobnost výskytu, která nabývá hodnot od 0 do 1. Místa s prezencí a absencí jsou poté vytvořena na základě pravděpodobnosti. Například místo s hodnotou pravděpodobnosti výskytu 0,2 bude ve 2 z 10 případů označeno jako prezence a ve zbylých 8 jako absence (Leroy et al. 2016). Pravděpodobnostní přístup je považovaný za vhodnější jeho doporučení je podpořeno několika studiemi (Grimmett et al. 2021; Meynard et al. 2019; Moudrý 2015). Čtvrtou fází tvorby virtuálního druhu je vytvoření vzorku nálezových dat na získaných místech prezencí a absencí. Virtuální druhy umožňují zvolit konkrétní počet vytvořených záznamů a také nastavit poměr mezi prezencemi a absencemi (Meynard et al. 2019).

V současné době existuje několik možností, jak virtuální druhy vytvářet. Nejčastěji se s takovými nástroji lze setkat ve formě softwarových balíčků. Mezi ně patří například *virtualspecies*, *SDMvspecies*, *RangeShifter* nebo *HexSim* (Leroy et al. 2016). Každý z nich má určité vlastnosti a možnosti. Balíček *virtualspecies* je ve studiích často zmiňován, jelikož dokáže překonávat nedostatky ostatních zmíněných nástrojů. Také využívá pravděpodobnostní přístup k převodu na prezenci/absenci (Moudrý 2015).

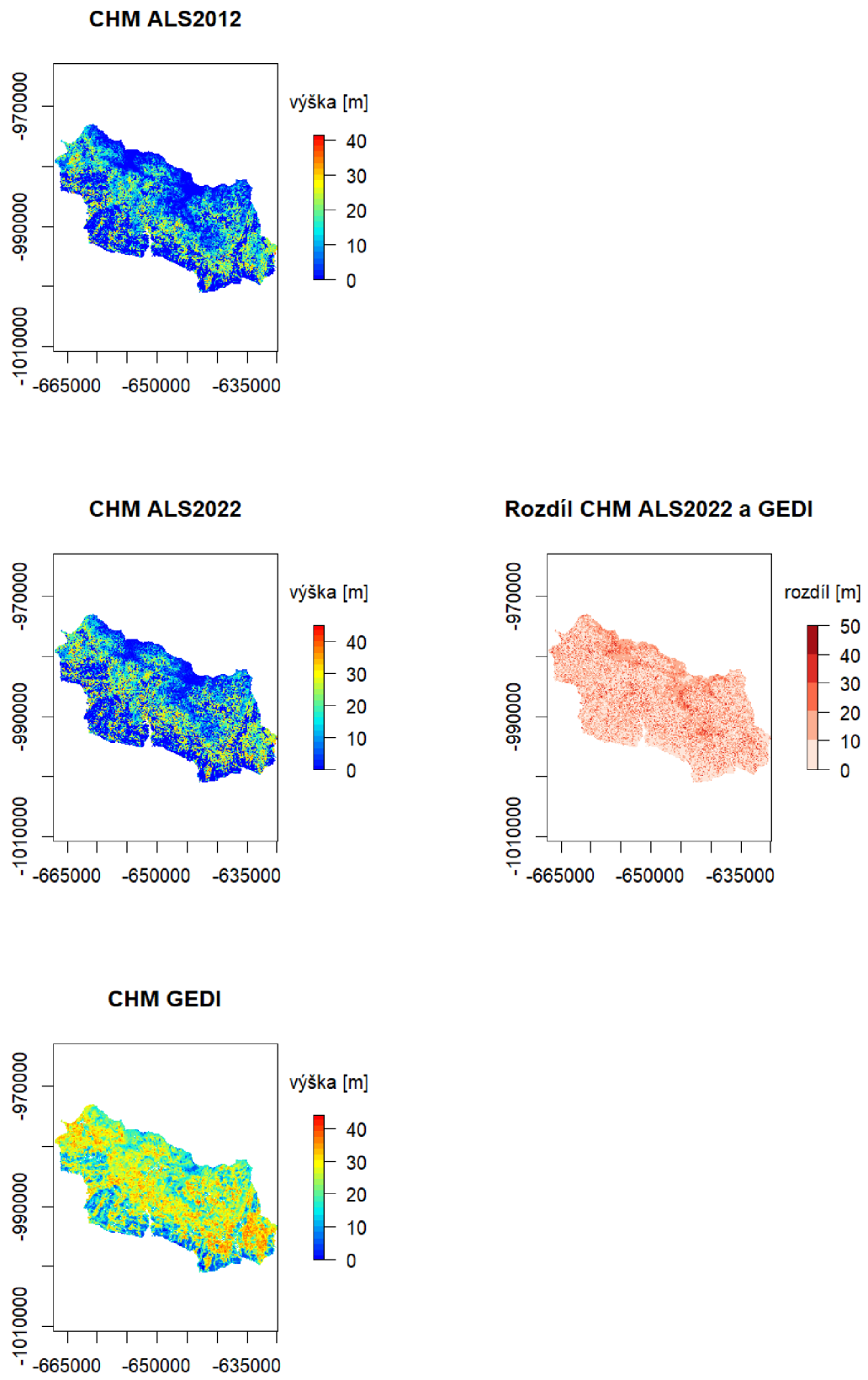
Virtuální druhy mají v modelování druhové distribuce svůj význam. Využívají se hlavně při zkoumání dopadů použití vstupních dat a nastavení parametrů v modelech. S jejich pomocí je možné odhalit určité nedostatky v modelech a zlepšit tak jejich kvalitu (Meynard et al. 2019).

3 Metodika

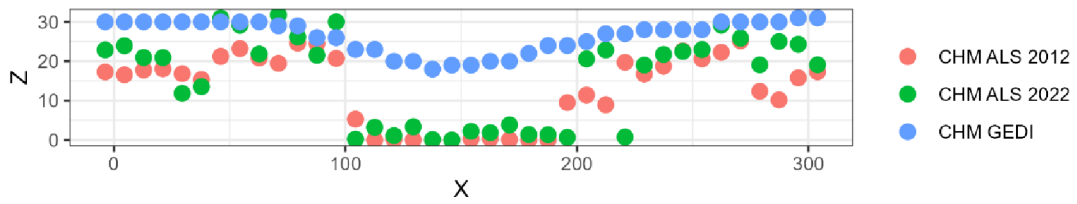
3.1 Použitá data

Testovacím územím byl Krkonošský národní park. V práci byly využity 3 rastrové modely výšky vegetace (Canopy height model – CHM). Cílem bylo zjistit, jak každý z nich ovlivňuje výslednou kvalitu modelu druhové distribuce. Dva CHM byly produkty dat leteckého laserového skenování (Airborne laser scanning – ALS), jeden byl pořízený v roce 2012 měl rozlišení 0,5 m a druhý byl z roku 2022 s rozlišením 1 m. Jako třetí CHM byl zvolen model od Lang et al. (2023). Ti ho vytvořili pomocí deep learning modelu, kde spojili data z družice Sentinel 2 a data laserového senzoru GEDI, umístěného na Mezinárodní vesmírné stanici. Vzniklý model výšky vegetace má globální pokrytí a jeho rozlišení je 10 m. Takto predikovaná data se vztahují k roku 2020. Důvodem pro použití těchto modelů je jejich odlišná kvalita a doba pořízení. Model z dat ALS z roku 2022 byl jakožto přesnější a novější využit ke generování virtuálního druhu. U toho byly dále nasimulovány různé úrovně polohové chyby a tato data vstupovala, společně s každým ze 3 CHM, do modelů druhové distribuce. Následně byla různými metrikami hodnocena jejich vzájemná přesnost. Předpokladem bylo, že modely trénované na datech GEDI budou predikovat pravděpodobnost výskytu s nižší přesností než modely založené na datech leteckého laserového skenování. Dále bylo zjišťováno, jak se bude lišit kvalita modelů s daty ALS z roku 2012 a 2022 a zda starší ale kvalitnější CHM bude mít lepší/horší přesnost než aktuální, ale predikovaný CHM z dat GEDI.

Modely výšky vegetace bylo nutné nejprve upravit, aby měly všechny stejné rozlišení a rozsah. K tomu byl využit program ArcGIS Pro (ESRI 2023), kde bylo ve funkci Aggregate změněno rozlišení rastrů z dat ALS na 10 m, tedy na rozlišení rastru z dat GEDI. V parametru Snap raster bylo nastaveno, aby se buňky všech rastrů přesně překrývaly. Nakonec byly rastry oříznuty na území Krkonošského národního parku. Všechny CHM lze pozorovat na obrázku (*Obrázek 2*). Zde je patrné, jak se mezi sebou jednotlivé modely liší. CHM založený na datech GEDI se od ostatních odchyľuje mnohem více, což dokazuje mapa rozdílů. Obecně se z tohoto dá usoudit, že model GEDI výšku vegetace nadhodnocuje a přidává výšku i v místech, kde se žádná vegetace nevyskytuje (*Obrázek 3*).



Obrázek 2: Modely výšky vegetace z dat ALS z roku 2012, 2022 a z dat GEDI. Vpravo rozdíl mezi modely z odlišných datových zdrojů.



Obrázek 3: Průřez modely výšky vegetace. Zobrazuje nadhodnocování výšek modelem GEDI.

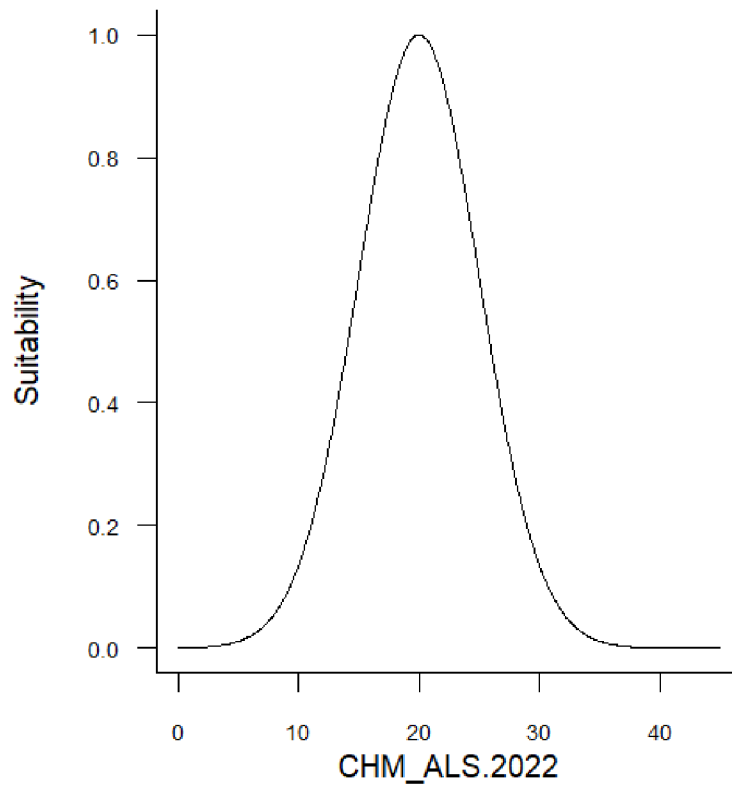
3.2 Generování virtuálního druhu

Virtuální druh byl vytvořen v programu R, v balíčku `virtualspecies`. Proces lze popsat v těchto hlavních krocích: a) definování vztahu mezi druhem a prostředím b) vytvoření mapy vhodnosti prostředí pro virtuální druh, c) převedení vhodnosti prostředí na mapu prezence/absence a d) vytvoření vzorku nálezových dat.

Vztah mezi druhem a prostředím udává, jaké podmínky druh toleruje a může se v nich vyskytovat. Toto chování se zjednodušeně popisuje pomocí tzv. response křivky. U tohoto druhu byla modelována vhodnost prostředí závislá na výšce vegetace, kterou představuje model výšky vegetace vytvořený z dat ALS z roku 2022. Ve funkci `formatFunctions` bylo nastaveno, že křivka má mít tvar normálního rozdělení se střední hodnotou $\mu = 20$ m a směrodatnou odchylkou $\sigma = 5$ m. Hodnota vhodnosti prostředí byla určena podle obecné normalizované Gaussovy funkce (*Rovnice 1*) a nabývá hodnot mezi 0 a 1 (*Obrázek 4*).

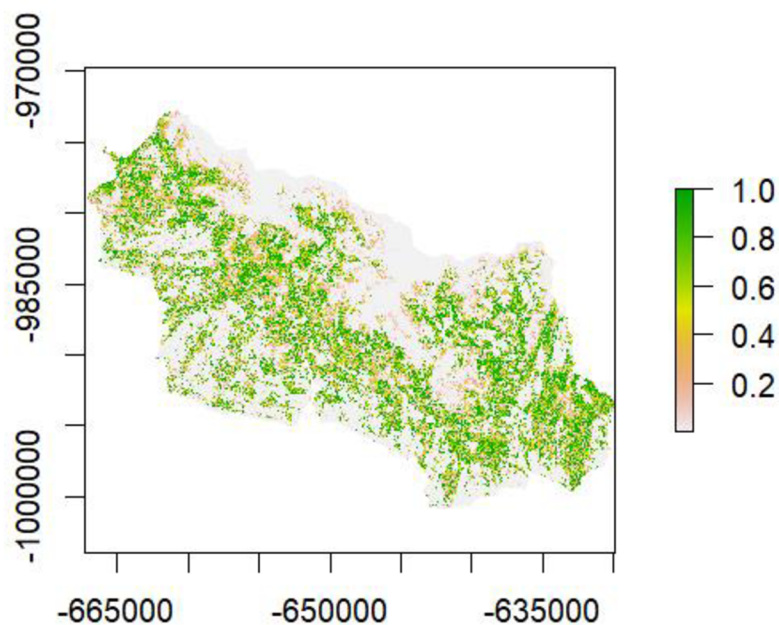
$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Rovnice 1: Obecná normalizovaná Gaussova funkce. Argumenty jsou výška vegetace v daném místě (x), střední hodnota ($\mu = 20$ m) a směrodatná odchylka ($\sigma = 5$ m).



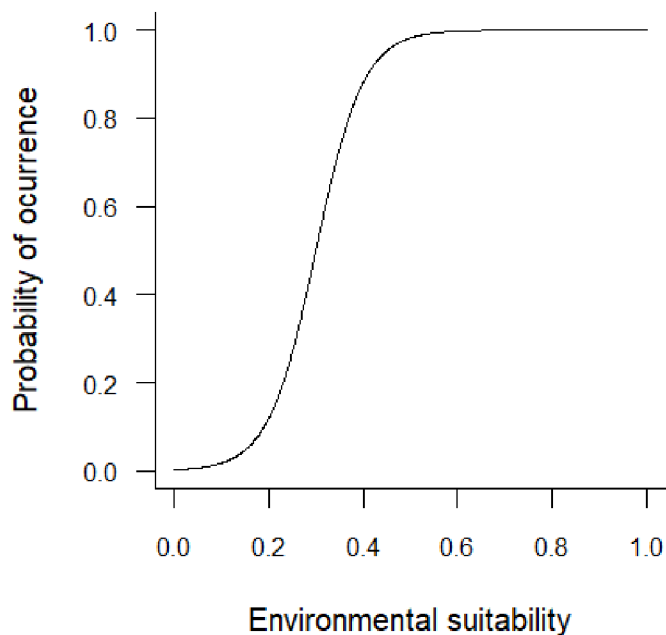
Obrázek 4: Vztah mezi výškou vegetace a vhodností prostředí.

Dále byla vhodnost prostředí pro virtuální druh vypočítána na celém sledovaném území. Výše definovaná funkce byla aplikována na všech buňkách rastrového CHM, k čemuž byla využita funkce generateSpFromFun. Tím byla získána mapa vhodnosti prostředí (*Obrázek 5*).

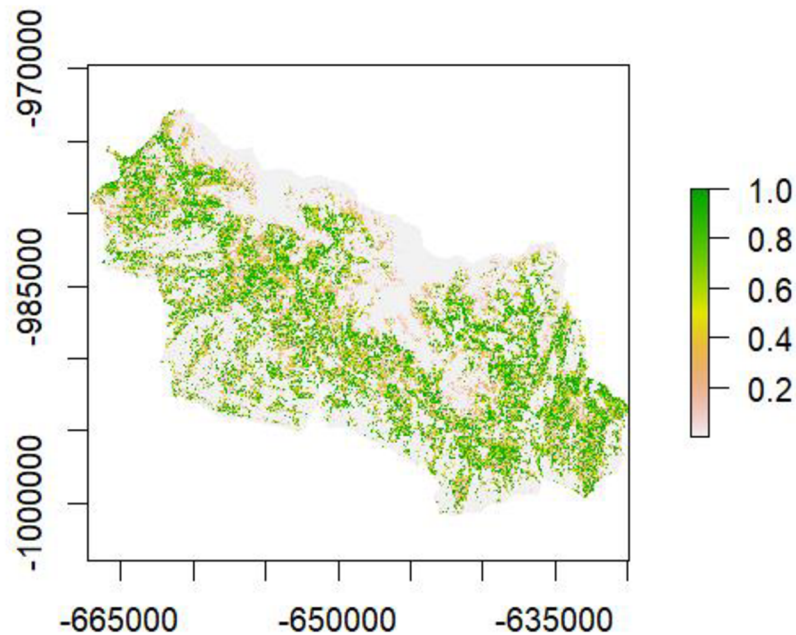


Obrázek 5: Vhodnost prostředí virtuálního druhu

Tato mapa ale stále neukazuje, zda se na konkrétním místě druh vyskytuje. Pouze udává, jak jsou pro něj vhodné podmínky. K získání dat o prezenci/absenci je nutné vhodnost prostředí nejprve převést na pravděpodobnost výskytu. Nejjednodušším způsobem je převod pomocí fixní prahové hodnoty, ale studie ukázaly (Grimmett et al. 2021; Meynard a Kaplan 2013), že takový přístup nereflektuje chování druhu ve skutečném světě a také téměř eliminuje gradient ve vztahu mezi vhodností a faktorem prostředí, který byl definován v předchozím kroku. Z tohoto důvodu byl použit pravděpodobnostní přístup, při kterém byla pravděpodobnost výskytu zjištěna pomocí logistické funkce (Obrázek 6). Tato pravděpodobnost závisí na hodnotě vhodnosti prostředí, byla opět vypočítána pro celou oblast a její hodnoty jsou v rozmezí 0–1 (Obrázek 7). Místa s prezencí/absencí jsou pak vytvořena jako náhodný výběr pixelů vážený pravděpodobností výskytu. K tomuto celému postupu byla použita funkce convertToPA.



Obrázek 6: Logistická funkce zobrazující vztah mezi vhodností prostředí a pravděpodobností výskytu



Obrázek 7: Pravděpodobnost výskytu virtuálního druhu

Posledním krokem generování virtuálního druhu bylo vytvoření vzorku nálezových dat ve formě bodů. Každý bod měl atributy se souřadnicemi a informací o tom, zda se jedná o prezenci či o absenci. K tvorbě této bodové sady sloužila funkce `sampleOccurrences`, jejímž vstupem byl rastr prezence/absence. Počet vytvořených nálezů byl nastaven na 300 a parametr `species prevalence`, který udává poměr mezi prezencí a absencí, měl hodnotu 0,33, tudíž výsledkem byl dvojnásobek absencí oproti prezencím. Na této sadě byla později simulována polohová chyba a také sloužila k trénování a testování modelů. Celý proces tvorby virtuálního druhu byl v R proveden následovně:

```
params_ALS.2022 <- formatFunctions(CHM_ALS.2022 = c(fun = 'dnorm',
mean = 20, sd = 5))
```

```
VS_ALS.2022 <- generateSpFromFun(raster.stack = CHM_ALS.2022,
parameters = params_ALS.2022, plot = FALSE)
```

```
PA.raster_ALS.2022 <- convertToPA(VS_ALS.2022, alpha = -0.05, beta
= 0.3, plot = FALSE)
```

```
PA.sampling_ALS.2022 <- sampleOccurrences(PA.raster_ALS.2022, n =
300, type = "presence-absence", plot = FALSE, sample.prevalence =
0.33)
```

3.3 Simulace polohové chyby

K posouzení vlivu polohové chyby ve vstupních datech na model bylo potřeba tuto chybu nasimulovat, a to posunem souřadnic v libovolném směru v rámci testovaného území. Byly definovány jednotlivé úrovně posunu podle vzdálenosti, o kterou se mohly změnit původní souřadnice. Testovány byly 2 způsoby simulování polohové chyby. V prvním z nich byly posuny provedeny tak, že úrovně na sebe navazovaly, tedy že minimální vzdálenost posunu dané úrovně vždy navazovala na maximální vzdálenost posunu úrovně předchozí. Úrovně byly následující: 5-10 m (E1), 10-50 m (E2), 50-100 m (E3), 100-300 m (E4) a 300-500 m (E5). Ve druhém způsobu simulování polohové chyby byly úrovně posunu vytvořeny tak, že minimální vzdálenost všech úrovní byla stejná a postupně se navyšovala pouze maximální vzdálenost. Úrovně posunu pak vypadaly takto: 5-10 m (E1), 5-50 m (E2), 5-100 m (E3), 5-300 m (E4) a 5-500 m (E5). Cílem bylo zjistit, zda má odlišná reprezentace polohové chyby na modely nějaký vliv. K posunu souřadnic bylo využito těchto dvou funkcí v programovacím jazyce R:

```
inside <- function(x, y, mask){
  ins <- ifelse(x < mask@extent@xmin | x > mask@extent@xmax | y <
mask@extent@ymin | y > mask@extent@ymax, FALSE, TRUE)
  if (ins == TRUE) {
    val <- extract(mask, SpatialPoints(coords = matrix(c(x,y), ncol
= 2)))
    ins <- ifelse(is.na(val), FALSE, TRUE)
  }
  return(ins)
}
```

```
shift <- function(x, y, dist, mask){
  new.x <- x
  new.y <- y
  inside <- FALSE
  while (!inside){
    angle <- runif(1,0,360)*pi/180
```

```

    new.x <- x + dist*sin(angle)
    new.y <- y + dist*cos(angle)
    inside <- inside(new.x, new.y, mask)
  }
  return(c(new.x, new.y))
}

```

Funkce `inside` limitovala posun pouze do oblasti zájmu tak, že při každém posunu souřadnic získala hodnotu rastru v tomto novém místě. Jako rastr zde byl použitý model výšky vegetace, který byl oříznutý na zájmové území. Pokud se posunutý bod bude stále nacházet v oblasti zájmu, bude tomuto místu odpovídat příslušná hodnota buňky rastru v tomto místě. Pokud by ale nový bod byl posunutý za hranice zájmového území, v takovém místě nebude hodnota rastru dostupná (vrátí NA hodnotu) a dojde k jeho vyřazení. To se bude opakovat do té doby, než bude posunutý bod ležet v oblasti zájmu. Funkce `shift` posouvala body v definovaných úrovních polohové chyby (E1 – E5).

Před samotným posunem byla nálezová data prezencí a absencí rozdělena na trénovací a testovací sadu v rovnoměrném poměru. Posuny byly následně provedeny pouze na trénovací sadě a testovací sada obsahovala původní nezměněné souřadnice. Takto bylo vytvořeno 5 úrovní polohové chyby v trénovací sadě, způsobem vzájemně navazujících úrovní:

```

Train_ALS.2022[,c("E1x", "E1y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 5, 10),mask=mask)))

```

```

Train_ALS.2022[,c("E2x", "E2y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 10,
50),mask=mask)))

```

```

Train_ALS.2022[,c("E3x", "E3y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 50,
100),mask=mask)))

```

```
Train_ALS.2022[,c("E4x", "E4y")] <- t(apply(Train_ALS.2022, 1,  
function(.) shift(.["x"], .["y"], dist=runif(1, 100,  
300),mask=mask)))
```

```
Train_ALS.2022[,c("E5x", "E5y")] <- t(apply(Train_ALS.2022, 1,  
function(.) shift(.["x"], .["y"], dist=runif(1, 300,  
500),mask=mask)))
```

Pro druhý způsob simulování polohové chyby, kde každá úroveň posunu začínala na hodnotě 5 m, bylo využito stejných příkazů, jako u prvního způsobu, pouze byly změněny minimální vzdálenosti posunu:

```
Train_ALS.2022[,c("E1x", "E1y")] <- t(apply(Train_ALS.2022, 1,  
function(.) shift(.["x"], .["y"], dist=runif(1, 5, 10),mask=mask)))
```

```
Train_ALS.2022[,c("E2x", "E2y")] <- t(apply(Train_ALS.2022, 1,  
function(.) shift(.["x"], .["y"], dist=runif(1, 5, 50),mask=mask)))
```

```
Train_ALS.2022[,c("E3x", "E3y")] <- t(apply(Train_ALS.2022, 1,  
function(.) shift(.["x"], .["y"], dist=runif(1, 5,  
100),mask=mask)))
```

```
Train_ALS.2022[,c("E4x", "E4y")] <- t(apply(Train_ALS.2022, 1,  
function(.) shift(.["x"], .["y"], dist=runif(1, 5,  
300),mask=mask)))
```

```
Train_ALS.2022[,c("E5x", "E5y")] <- t(apply(Train_ALS.2022, 1,  
function(.) shift(.["x"], .["y"], dist=runif(1, 5,  
500),mask=mask)))
```

Dále byly extrahovány hodnoty všech 3 rastrů výšky vegetace z příslušných poloh všech (nyní i posunutých) bodů trénovací i testovací sady. Obě sady tak byly vytvořeny třikrát a v jedné byly obsaženy výšky vegetace z dat ALS z roku 2022, ve druhé z roku 2012 a ve třetí výšky z predikovaného globálního modelu z dat GEDI.

3.4 Tvorba modelů

Jelikož byla nálezová data ve formě prezencí a absencí, byly použity zobecněné lineární modely, které jsou v modelování druhové distribuce běžně používané (Hellegers et al. 2020; Gábor et al. 2020b; Melo-Merino et al. 2020;). Byly vytvořeny tři skupiny modelů podle toho, jaký model výšky vegetace do nich vstupoval jako prediktor. Ve všech těchto skupinách bylo vytvořeno šest modelů, jeden obsahoval původní neposunutá nálezová data a v ostatních pěti byla nálezová data posunuta podle popsanych úrovní. Tyto modely byly poté použity k predikci pravděpodobnosti výskytu. Nejprve byla predikována na testovacích sadách, které obsahovaly hodnoty z příslušných modelů výšky vegetace. Výsledkem byly predikované pravděpodobnosti výskytu v bodech testovacích sad. Poté byla predikce ještě provedena na celém území, tedy na všech rastrech výšky vegetace. Takovým výsledkem byly opět rastry, tentokrát s hodnotami predikované pravděpodobnosti výskytu.

K vyhodnocení kvality modelů byla použita plocha pod ROC křivkou (area under curve – AUC), která nabývá hodnot mezi 0 a 1, přičemž hodnota 0,5 představuje náhodnou predikci a hodnoty nad 0,9 označují velice kvalitní model (Fielding a Bell 1997). Dále byl vypočten Warrenův index I (Warren et al. 2008) a Schoenerův index D (Schoener 1968). Oba tyto indexy kvantifikují rozdíl mezi skutečnou pravděpodobností výskytu vytvořeného virtuálního druhu a pravděpodobností výskytu predikovanou modely. Porovnávají jejich překryv na celém území a jejich hodnoty se také pohybují v rozsahu 0 až 1, kde 0 znamená žádný překryv a 1 perfektní překryv. Ačkoliv jsou způsoby počítání těchto indexů trochu odlišné, mezi jejich výsledky jsou jen malé rozdíly. Byly vypočteny pomocí následující funkce:

```
index <- function (pred1, pred2, na.rm = TRUE){  
  p1 <- pred1/sum(pred1, na.rm = na.rm)  
  p2 <- pred2/sum(pred2, na.rm = na.rm)  
  SchoenerD <- 1 - 0.5 * sum(abs(p1 - p2), na.rm = na.rm)
```

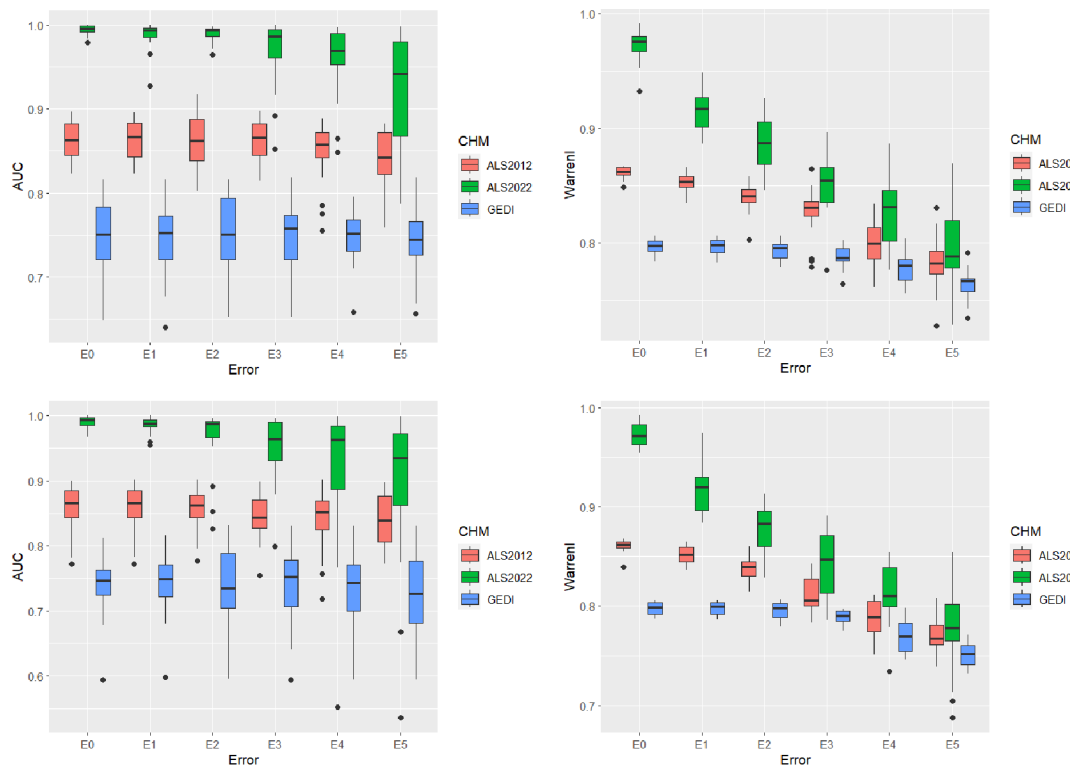


```
HellingerDist <- sqrt(sum((sqrt(p1) - sqrt(p2))^2, na.rm =  
na.rm))  
  
WarrenI <- 1 - ((HellingerDist^2)/2)  
  
data.frame(SchoenerD = SchoenerD, WarrenI = WarrenI,  
HellingerDist = HellingerDist)  
}
```

Celý proces tvorby virtuálního druhu, simulace polohové chyby, tvorby modelů i hodnocení jejich kvality byl 20krát zopakován. Prvek náhody při vytváření mapy prezence/absence a tvorbě vzorku nálezových dat zajišťoval, že výsledky nebyly pokaždé zcela totožné.

4 Výsledky

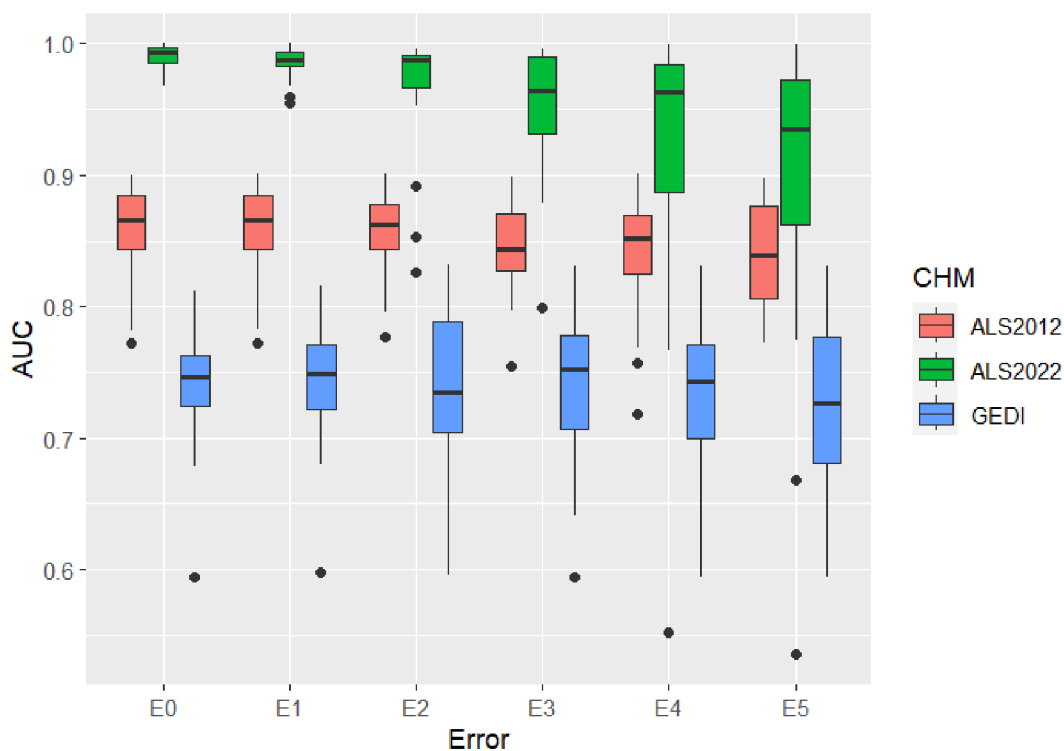
Bylo zjištěno, že mezi modely s odlišným způsobem simulování polohové chyby jsou zcela zanedbatelné rozdíly, což dokazují hodnoty vybraných metrik (*Obrázek 8*). V tomto případě tedy nezáleží, jakým způsobem jsou jednotlivé úrovně definovány a je pravděpodobné, že kvalitu modelů ovlivňují spíše maximální hodnoty polohové chyby v daných úrovních. Dále proto bude věnována pozornost pouze modelům se vzájemně navazujícími úrovněmi polohové chyby.



Obrázek 8: Hodnoty AUC a indexu Warren I pro modely se stejnou minimální vzdáleností posunu (nahore) a pro modely s navzájem navazujícími úrovněmi posunu (dole).

Výsledné hodnoty AUC (*Obrázek 9*) ukazují, jaké kvality dosahují modely s různou úrovní polohové chyby a přesností prediktoru. Nejvyšší kvalitu mají modely založené na aktuálním a přesném CHM z dat leteckého laserového skenování z roku 2022, na kterém byl zároveň simulován virtuální druh. Hodnoty se ve všech úrovních polohové chyby pohybovaly v rozmezí 0,9 – 1, což značí výbornou predikční schopnost modelů. Stále za velice kvalitní lze považovat modely založené na druhém CHM vytvořeném z dat ALS, ale starším z roku 2012. Zde se hodnoty AUC pohybovaly mezi 0,8 – 0,9. Nejméně přesné modely z hlediska vstupního prediktoru byly ty, založené na predikovaném globálním CHM z dat GEDI. Ačkoliv se jeho data vztahují k roku 2020, a má tedy časově blíže k CHM, na kterém byl simulován

virtuální druh, než starý model z roku 2012, je zároveň velmi nekvalitní a podle hodnot AUC vykazuje horší přesnost. Dále je patrné, že AUC není příliš citlivá na polohovou chybu v datech. I přes posun souřadnic až v řádu stovek metrů oproti původním, zůstávají průměrné hodnoty AUC takřka nezměněny (Tabulka 2). U modelů založených na CHM z roku 2022 způsobuje nárůst polohové chyby větší variabilitu hodnot AUC. Ostatní modely na polohovou chybu téměř nereagují.

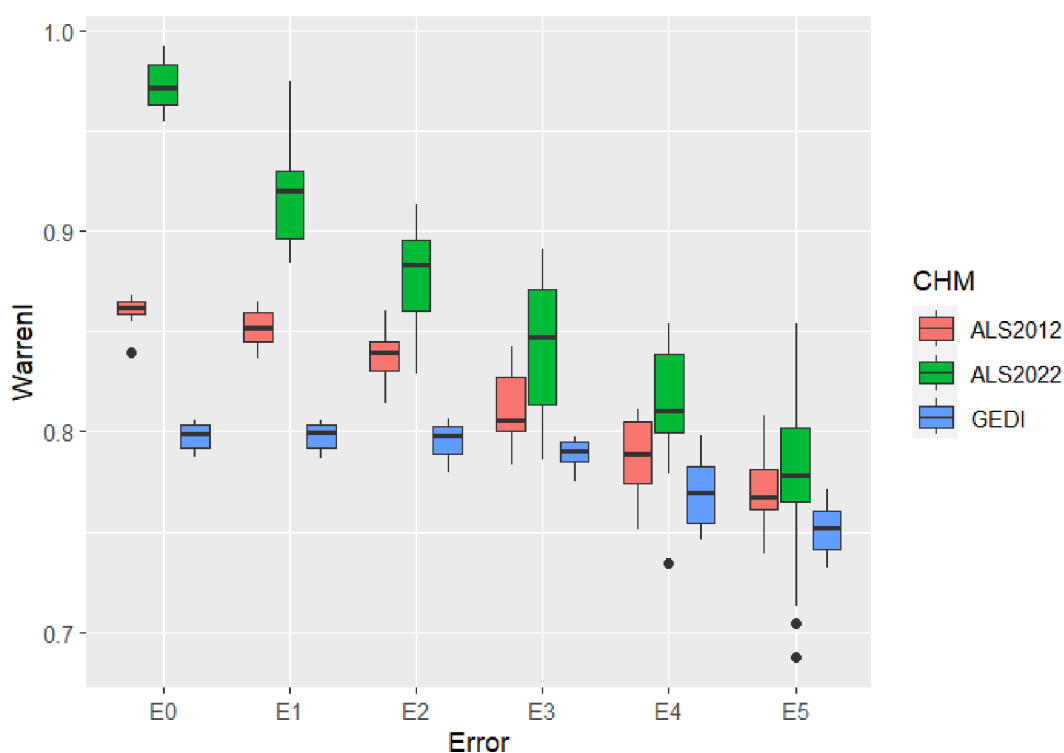


Obrázek 9: Výsledné AUC pro modely s jednotlivými úrovněmi polohové chyby (E0 – bez posunu, E1 - 5-10 m, E2 - 10-50 m, E3 - 50-100 m, E4 - 100-300 m, E5 - 300-500 m) a odlišnými modely výšky vegetace.

		Polohová chyba						
		E0	E1	E2	E3	E4	E5	
AUC	ALS2012	μ	0,86	0,86	0,86	0,84	0,84	0,84
		σ	0,04	0,04	0,03	0,03	0,05	0,04
	ALS2022	μ	0,99	0,99	0,97	0,95	0,91	0,89
		σ	0,01	0,01	0,05	0,05	0,11	0,12
	GEDI	μ	0,74	0,74	0,74	0,74	0,73	0,73
		σ	0,05	0,05	0,06	0,06	0,06	0,06

Tabulka 2: Střední hodnoty μ a směrodatné odchytky σ AUC pro modely s jednotlivými úrovněmi polohové chyby (E0 – bez posunu, E1 - 5-10 m, E2 - 10-50 m, E3 - 50-100 m, E4 - 100-300 m, E5 - 300-500 m) a odlišnými modely výšky vegetace.

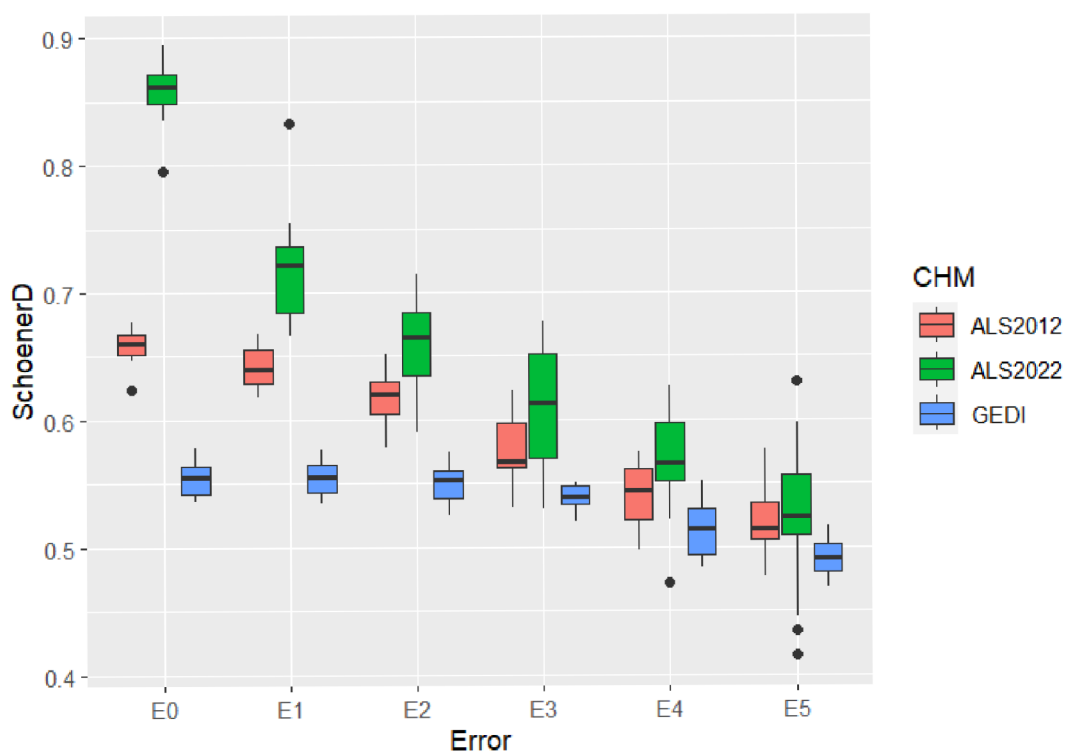
Warrenův index I porovnává překryv původní mapy vhodnosti prostředí s predikovanou pravděpodobností výskytu. Jeho hodnoty opět reflektují schopnost predikovat pravděpodobnost výskytu (*Obrázek 10*). Oproti AUC je zde již patrný vliv posunu souřadnic na kvalitu modelů. Kromě narůstající variability v hodnotách indexu dochází zároveň k poklesu jeho průměrných hodnot (*Tabulka 3*). Podobné výsledky udává také Schoenerův D index (*Obrázek 11*). Ačkoliv se hodnoty (*Tabulka 4*) obou indexů liší, z grafů lze pozorovat, že data sledují stejný trend. Indexy tak na polohovou chybu v datech o druhu reagují téměř totožně. Díky tomu je možné skutečně posoudit vliv polohové chyby v datech o druhu na model.



Obrázek 10: Výsledný Warrenův index I pro modely s jednotlivými úrovněmi polohové chyby (E0 – bez posunu, E1 - 5-10 m, E2 - 10-50 m, E3 - 50-100 m, E4 - 100-300 m, E5 - 300-500 m) a odlišnými modely výšky vegetace.

		Polohová chyba						
		E0	E1	E2	E3	E4	E5	
Warrenův I index	ALS2012	μ	0,86	0,85	0,84	0,81	0,79	0,77
		σ	0,01	0,01	0,01	0,02	0,02	0,02
	ALS2022	μ	0,97	0,92	0,88	0,84	0,81	0,78
		σ	0,01	0,02	0,02	0,03	0,03	0,04
	GEDI	μ	0,8	0,8	0,8	0,79	0,77	0,75
		σ	0,01	0,01	0,01	0,01	0,02	0,01

Tabulka 3: Střední hodnoty μ a směrodatné odchylky σ Warrenova indexu I pro modely s jednotlivými úrovněmi polohové chyby (E0 – bez posunu, E1 - 5-10 m, E2 - 10-50 m, E3 - 50-100 m, E4 - 100-300 m, E5 - 300-500 m) a odlišnými modely výšky vegetace.



Obrázek 11: Výsledný Schoenerův index D pro modely s jednotlivými úrovněmi polohové chyby (E0 – bez posunu, E1 - 5-10 m, E2 - 10-50 m, E3 - 50-100 m, E4 - 100-300 m, E5 - 300-500 m) a odlišnými modely výšky vegetace.

			Polohová chyba					
			E0	E1	E2	E3	E4	E5
Schoenerův D index	ALS2012	μ	0,66	0,64	0,62	0,58	0,54	0,52
		σ	0,01	0,01	0,02	0,03	0,02	0,03
	ALS2022	μ	0,86	0,72	0,66	0,61	0,57	0,53
		σ	0,02	0,04	0,03	0,05	0,04	0,05
	GEDI	μ	0,55	0,55	0,55	0,54	0,51	0,49
		σ	0,01	0,01	0,02	0,01	0,02	0,02

Tabulka 4: Střední hodnoty μ a směrodatné odchylky σ Schoenerova indexu D pro modely s jednotlivými úrovněmi polohové chyby (E0 – bez posunu, E1 - 5-10 m, E2 - 10-50 m, E3 - 50-100 m, E4 - 100-300 m, E5 - 300-500 m) a odlišnými modely výšky vegetace.

Modely se vstupním CHM z přesnějších a aktuálnějších dat ALS jsou posunem souřadnic ovlivněny více než modely se starším CHM a predikovanou výškou vegetace z dat GEDI. U prvních jmenovaných dochází k propadu hodnot indexu již od první úrovně posunu, která se pohybuje od 5 do 10 m (E1) a tímto trendem pokračuje až do nejvyšší úrovně posunu. Naproti tomu u modelů založených na predikované výšce z dat GEDI dochází k prvnímu mírnému poklesu hodnot indexu až u polohové chyby od 50 do 100 m (E3) a k většímu poklesu až u posunu od 100 do 300 m (E4). Ačkoliv jsou tedy tyto modely celkově méně přesné než modely s daty ALS, polohová chyba má na ně menší vliv.

Samotné použití nekvalitních prediktorů má na přesnost modelů také značný vliv. Porovnání dvou skupin modelů založených na CHM z dat ALS ukazuje pokles hodnot indexů při použití starších dat. Ještě výraznější pokles je zaznamenán u modelů založených na CHM z dat GEDI. Pokud bychom hodnotili modely s neposunutými souřadnicemi, tak u Warrenova indexu I hodnoty klesají z průměrných 0,97 (ALS 2022) na 0,86 (ALS 2012) a dokonce až na 0,8 (CHM GEDI). Pro srovnání, takové hodnoty odpovídají přesnosti modelů s daty ALS 2022 s polohovou chybou úrovně E4 – E5, tedy až 500 m.

Z pohledu interakce polohové chyby s nekvalitními prediktory je zřejmé, že výskyt obou chyb současně nezpůsobuje signifikantnější zhoršení přesnosti modelu. Toto chování lze pozorovat u modelů s prediktorem výšky vegetace z predikovaných dat GEDI. Tento prediktor je, jak již bylo zmíněno, velice nepřesný a již sám o sobě snižuje přesnost modelu. Přidání další chyby v podobě posunu souřadnic již ale nevede

k dalšímu poklesu přesnosti. Hodnoty Schoenerova D indexu se pro tyto modely pohybují napříč všemi úrovněmi polohové chyby v rozsahu 0,55 – 0,50. Pokles přesnosti je zde jen minimální. Naproti tomu u modelů založených na přesném prediktoru v podobě výšky vegetace z dat ALS 2022 je klesají hodnoty jen v první úrovni polohové chyby z 0,86 na 0,72.

5 Diskuze

Z výsledků vyplývá, že nedochází k synergickému působení polohové chyby v náleзовých datech o druhu a nepřesných prediktorů na model, nýbrž výskyt jedné chyby zamezuje projevu chyby druhé a není způsobeno další zhoršení modelu. To lze dokázat na modelech založených na starším CHM z dat ALS z roku 2012 a také na predikované výšce z dat GEDI. V prvně jmenovaných je zdrojem nepřesnosti časový rozdíl mezi vytvořením virtuálního druhu a pořízením prediktoru. Prediktory v druhých modelech jsou zatíženy nepřesnostmi už ze své podstaty, že se jedná o predikovanou výšku vegetace, která je sice aktuálnější, ale pochází z vícero datových zdrojů a od skutečného stavu v podobě CHM z dat ALS se výrazně odchyľují. Obě tyto skupiny modelů tedy obsahují nekvalitní prediktory a vliv samotné polohové chyby se téměř neprojevuje. Naproti tomu u modelů založených na kvalitním prediktoru v podobě CHM z dat ALS z roku 2022, na kterém byl generován virtuální druh a jeho míra neurčitosti je tak nízká, se vliv polohové chyby v náleзовých datech o druhu naplno projevuje. Každá další úroveň polohové chyby zde znamená pokles hodnot indexů hodnotících jejich přesnost. Z toho mimo jiné vyplývá, že pokud je očekávána chyba v prediktorech, není potřeba vynakládat snahu o zpřesnění náleзовých dat. Nicméně v případě polohové chyby to platí do hodnot 100–300 m. Od této úrovně i v takových modelech začíná klesat přesnost. Nakonec bylo zjištěno, že odlišný způsob simulování polohové chyby nemá na kvalitu modelů druhové distribuce takřka žádný vliv.

Polohové chybě a jejímu efektu na modely druhové distribuce se věnuje velké množství studií (Gábor et al. 2022; 2023; Hanberry 2013; Naimi et al. 2014; Osborne a Leitão 2009; Smith et al. 2023; Tulowiecki et al. 2015). Například ve studii Gábora et al. (2020b) zkoumali, zda polohová chyba ovlivňuje více modely generalistů, specialistů nebo jejich průměru. Jejich výsledky říkají, že chyba se ve všech třech případech nejvíce projevovala mezi žádným posunem souřadnic a nejmenším simulovaným posunem, který byl v jejich případě 5-10 m, což je stejná úroveň posunu, jako je v této práci. Stejné chování modelů ukázaly i naše výsledky. Jejich hlavním zjištěním bylo, že polohová chyba ovlivňuje více modely specialistů než generalistů. Mimo jiné také potvrdili, že zvýšení počtu záznamů v datech o druzích vede ke zlepšení kvality modelů, což zmiňovali i Moudrý a Šimová (2012). Testování změn v počtu záznamů by mohlo být prospěšné i v této práci. Mohlo by tak být například

zjištěno, zda by zvýšení počtu záznamů vedlo ke zlepšení kvality modelů založených na datech GEDI a dostalo je blíže ke kvalitě modelů založených na datech leteckého laserového skenování. K hodnocení modelů využili kromě AUC a Warrenova indexu I ještě metriku true skill statistics (TSS) a Spearmanův korelační koeficient. Zahrnutí některé z těchto metrik by mohlo rovněž zpřesnit některé výsledky této práce. Mohly by například pomoci vysvětlit, proč zde AUC téměř nereaguje na polohovou chybu v datech. Nutno podotknout, že samotná AUC je často kritizována jako metrika hodnotící přesnost modelů druhové distribuce. Lobo et al. (2008) například ve své studii vysvětlili 5 důvodů, proč může být AUC zavádějící. Mezi hlavní patří skutečnost, že AUC nedokáže vymeziť prostorové rozložení chyb v modelech, a také že zvětšováním geografického rozsahu modelování mimo oblast prezencí se zvyšuje hodnota AUC díky vyššímu poměru správně predikovaných absencí.

Další studii, která zkoumala vliv polohové chyby na modely druhové distribuce, přinesli i Mitchell et al. (2017). Podobně jako Gábor et al. (2020b) testovali modely s různými úrovněmi polohové chyby a také zjišťovali efekt počtu záznamů na kvalitu modelů. I oni potvrdili, že zvýšení počtu záznamů zlepšuje predikční schopnost modelů a ubírá na variabilitě příslušných metrik. Zajímavějším zjištěním v kontextu této práce je však fakt, že v jejich případě rovněž AUC nevykazovalo žádné změny s rostoucí polohovou chybou. Není jasné, z jakého důvodu nejsou hodnoty AUC posunem souřadnic v datech o druhu ovlivněny. Odlišný byl od této práce přístup v počítání. V jejich studii počítali AUC na celém predikovaném rastru vhodnosti prostředí, zatímco v této práci bylo vypočteno na testovací sadě diskrétních bodů, přesto byl výsledek stejný. V závěru studie konstatovali, že na základě tohoto zjištění lze tvořit kvalitní modely druhové distribuce i s určitou úrovní polohové chyby. To je do jisté míry pravda, ale výsledky této práce ukazují, že stejné modely jsou hodnoceny jinak pomocí AUC a jinak pomocí například Warrenova indexu I. Důraz by měl být tedy kladen na zahrnutí metrik fungujících na odlišných principech a nespolehat se pouze na jednu.

Časový rozdíl mezi daty o druhu a o prostředí nemá na modely příliš velký vliv. To potvrzují i ve studii Vierling et al. (2014), kteří rovněž porovnávali modely s časovým rozdílem mezi daty o druhu a o prostředí. Zjistili, že 6 let rozdílu nezpůsobí tak zásadní pokles kvality modelů a je i s takovými daty možné vytvořit použitelnou predikci druhové distribuce, což podporuje i tato práce, kde je mezi dvěma CHM z dat

ALS rozdíl 10 let. Zároveň ale podotýkají, že záleží na tom, jaké prediktory do modelů vstupují a je potřeba být obezřetný, protože některé proměnné se za takovou dobu mohou proměnit více než jiné. Také Hill a Hinsley (2015) dokázali, že časový rozdíl až 12 let mezi pořízením modelů výšky vegetace není velkou překážkou v modelování druhové distribuce. Za zmínku stojí, že jejich oblastí zájmu byl starší les, kde již nedochází k nijak zásadním změnám a výsledek by mohl být jiný na odlišné lokalitě. Na to koneckonců upozorňují i Moudrý et al. (2021), že zatímco jsou u leteckého laserového skenování běžné rozdíly v době pořízení dat stejného území 5 až 10 let, což je pro některé oblasti adekvátní, pro některé rychle se vyvíjející lokality, jako jsou například ranná sukcesní stanoviště, může již takový rozdíl zanášet do modelů chybu. Otázkou je tedy, jak často je vhodné taková data pořizovat, jelikož se jedná o velice nákladnou činnost.

Žádná studie se nepokoušela zjistit, zda odlišná metoda simulování polohové chyby bude mít nějaký vliv na kvalitu modelů. V literatuře se objevují dva způsoby vymezení úrovně polohové chyby, například u Gábora et al. (2023) na sebe úrovně navazují (50-100 m, 100-250 m, 250-500 m...) a v práci Mitchell et al. (2017) všechny začínaly na stejné hodnotě a pouze se navyšovala maximální vzdálenost (0-5 m, 0-25 m, 0-50 m...). V této práci byly otestovány oba způsoby a bylo dokázáno, že jinak definované úrovně posunu souřadnic produkují totožné modely. Jelikož se v úrovních lišila vždy minimální vzdálenost posunu, ale maximální byla v odpovídajících úrovních stejná, je možné, že jsou modely ovlivňovány hlavně maximální možnou vzdáleností posunu.

6 Závěr

Tato práce zkoumala, jak polohová chyba v nálezových datech o druhu ovlivňuje přesnost modelů druhové distribuce založených na prediktorech s různou kvalitou. Za využití virtuálního druhu bylo dokázáno, že nedochází k synergickému působení obou chyb zároveň, ale v případě výskytu jedné chyby zaniká efekt chyby druhé. Tento jev se projevuje u modelů, jehož prediktory nesou určitou míru nepřesnosti a polohová chyba má na ně kvůli tomu menší vliv. U modelů, jejichž prediktory nejsou natolik zatíženy neurčitostmi v datech, dochází ke značnému poklesu přesnosti již od nejnižší úrovně polohové chyby. Na základě těchto zjištění se doporučuje pro modelování druhové distribuce vybírat vždy co nejpřesněji lokalizovaná data o druhu a prediktory vhodné kvality, jelikož ty vedou k nejpřesnějším výsledkům. Pokud však takto kvalitní prediktory nejsou k dispozici, není potřeba vybírat pouze polohově přesná data (čímž se snižuje počet záznamů v datech o druhu), protože by takový krok nepřinesl zvýšení prediktivní schopnosti modelu.

7 Přehled literatury a použitých zdrojů

A. LEE-YAW, Julie, Jenny L. MCCUNE, Samuel PIRONON a Seema N. SHETH, 2022. Species distribution models rarely predict the biology of real populations. *Ecography* [online]. **2022**(6) [vid. 2024-02-28]. ISSN 16000587. Dostupné z: doi:10.1111/ecog.05877

ARENAS-CASTRO, Salvador, Adrián REGOS, Ivone MARTINS, João HONRADO a Joaquim ALONSO, 2022. Effects of input data sources on species distribution model predictions across species with different distributional ranges. *Journal of Biogeography* [online]. **49**(7), 1299–1312 [vid. 2024-02-21]. ISSN 13652699. Dostupné z: doi:10.1111/jbi.14382

BORIA, Robert A, Link E OLSON, Steven M GOODMAN a Robert P ANDERSON, 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* [online]. **275**, 73–77 [vid. 2024-03-08]. ISSN 03043800. Dostupné z: doi:10.1016/j.ecolmodel.2013.12.012

CLAERHOUT, Tim, Diederik STRUBBE, Ramón CASTILLO TORRES, Marie Stéphanie SAMAIN, Lars W CHATROU a Emily VELTJEN, 2023. An integrative study of species distribution modelling and conservation genetics: Magnolia in Hispaniola. *Biodiversity and Conservation* [online]. **32**(4), 1205–1231 [vid. 2024-02-25]. ISSN 15729710. Dostupné z: doi:10.1007/s10531-023-02546-6

CONNOR, Thomas, Andrés VIÑA, Julie A. WINKLER, Vanessa HULL, Ying TANG, Ashton SHORTRIDGE, Hongbo YANG, Zhiqiang ZHAO, Fang WANG, Jindong ZHANG, Zejun ZHANG, Caiquan ZHOU, Wenke BAI a Jianguo LIU, 2019. Interactive spatial scale effects on species distribution modeling: The case of the giant panda. *Scientific Reports* [online]. **9**(1), 1–14. ISSN 20452322. Dostupné z: doi:10.1038/s41598-019-50953-z

ELITH, Jane a John R LEATHWICK, 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* [online]. **40**, 677–697 [vid. 2024-02-25]. ISSN 1543592X. Dostupné z: doi:10.1146/annurev.ecolsys.110308.120159

ESRI, 2023. *ArcGIS Pro*. 2023. Redlands, CA: Environmental Systems

Research Institute.

FERNANDEZ, M., H HAMILTON a L M KUEPPERS, 2013. Characterizing uncertainty in species distribution models derived from interpolated weather station data. *Ecosphere* [online]. **4**(5) [vid. 2024-03-13]. ISSN 21508925. Dostupné z: doi:10.1890/ES13-00049.1

FIELDING, Alan H. a John F. BELL, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* [online]. **24**(1), 38–49. ISSN 03768929. Dostupné z: doi:10.1017/S0376892997000088

FOURCADE, Yoan, Aurélien G. BESNARD a Jean SECONDI, 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography* [online]. **27**(2), 245–256. ISSN 14668238. Dostupné z: doi:10.1111/geb.12684

FRANKLIN, Janet, Frank W. DAVIS, Makihiko IKEGAMI, Alexandra D. SYPHARD, Lorraine E. FLINT, Alan L. FLINT a Lee HANNAH, 2013. Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Global Change Biology* [online]. **19**(2), 473–483. ISSN 13541013. Dostupné z: doi:10.1111/gcb.12051

GÁBOR, Lukáš, Walter JETZ, Muyang LU, Duccio ROCCHINI, Anna CORD, Marco MALAVASI, Alejandra ZARZO-ARIAS, Vojtěch BARTÁK a Vítězslav MOUDRÝ, 2022. Positional errors in species distribution modelling are not overcome by the coarser grains of analysis. *Methods in Ecology and Evolution* [online]. **13**, 2289–2302 [vid. 2024-02-03]. ISSN 2041210X. Dostupné z: doi:10.1111/2041-210X.13956

GÁBOR, Lukáš, Walter JETZ, Alejandra ZARZO-ARIAS, Kevin WINNER, Scott YANCO, Stefan PINKERT, Charles J. MARSH, Matthew S. ROGAN, Jussi MÄKINEN, Duccio ROCCHINI, Vojtěch BARTÁK, Marco MALAVASI, Petr BALEJ a Vítězslav MOUDRÝ, 2023. Species distribution models affected by positional uncertainty in species occurrences can still be ecologically interpretable. *Ecography* [online]. **2023**(6), e06358 [vid. 2024-02-03]. ISSN 16000587. Dostupné z: doi:10.1111/ecog.06358

GÁBOR, Lukáš, Vítězslav MOUDRÝ, Vojtěch BARTÁK a Vincent LECOURE, 2020a. How do species and data characteristics affect species distribution models and when to use environmental filtering? *International Journal of Geographical Information Science* [online]. **34**(8), 1567–1584 [vid. 2024-02-21]. ISSN 13623087. Dostupné z: doi:10.1080/13658816.2019.1615070

GÁBOR, Lukáš, Vítězslav MOUDRÝ, Vincent LECOURE, Marco MALAVASI, Vojtěch BARTÁK, Michal FOGL, Petra ŠÍMOVÁ, Duccio ROCCHINI a Tomáš VÁCLAVÍK, 2020b. The effect of positional error on fine scale species distribution models increases for specialist species. *Ecography* [online]. **43**(2), 256–269. ISSN 16000587. Dostupné z: doi:10.1111/ecog.04687

GRIMMETT, Liam, Rachel WHITSED a Ana HORTA, 2021. Creating virtual species to test species distribution models: the importance of landscape structure, dispersal and population processes. *Ecography* [online]. **44**(5), 753–765 [vid. 2024-02-03]. ISSN 16000587. Dostupné z: doi:10.1111/ecog.05555

HANBERRY, Brice B, 2013. Finer grain size increases effects of error and changes influence of environmental predictors on species distribution models. *Ecological Informatics* [online]. **15**, 8–13 [vid. 2024-03-10]. ISSN 15749541. Dostupné z: doi:10.1016/j.ecoinf.2013.02.003

HELLEGERS, Marjon, Wim A. OZINGA, Arjen HINSBERG, Mark A.J. HUIJBREGTS, Stephan M. HENNEKENS, Joop H.J. SCHAMINÉE, Jürgen DENGLER a Aafke M. SCHIPPER, 2020. Evaluating the ecological realism of plant species distribution models with ecological indicator values. *Ecography* [online]. **43**(1), 161–170. ISSN 16000587. Dostupné z: doi:10.1111/ecog.04291

HERNANDEZ, Pilar A., Catherine H. GRAHAM, Lawrence L. MASTER a Deborah L. ALBERT, 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* [online]. **29**(5), 773–785 [vid. 2024-03-09]. ISSN 1600-0587. Dostupné z: doi:10.1111/J.0906-7590.2006.04700.X

HILL, Ross A. a Shelley A. HINSLEY, 2015. Airborne lidar for woodland habitat quality monitoring: Exploring the significance of lidar data characteristics when modelling organism-habitat relationships. *Remote Sensing* [online]. **7**(4), 3446–

3466. ISSN 20724292. Dostupné z: doi:10.3390/rs70403446

KOZAK, Kenneth H, Catherine H GRAHAM a John J WIENS, 2008. *Integrating GIS-based environmental data into evolutionary biology* [online]. 2008. [vid. 2024-02-25]. ISSN 01695347. Dostupné z: doi:10.1016/j.tree.2008.02.001

KRAMER-SCHADT, Stephanie, Jürgen NIEDBALLA, John D. PILGRIM, Boris SCHRÖDER, Jana LINDENBORN, Vanessa REINFELDER, Milena STILLFRIED, Ilja HECKMANN, Anne K. SCHARF, Dave M. AUGERI, Susan M. CHEYNE, Andrew J. HEARN, Joanna ROSS, David W. MACDONALD, John MATHAI, James EATON, Andrew J. MARSHALL, Gono SEMIADI, Rustam RUSTAM, Henry BERNARD, Raymond ALFRED, Hiromitsu SAMEJIMA, J. W. DUCKWORTH, Christine BREITENMOSER-WUERSTEN, Jerrold L. BELANT, Heribert HOFER a Andreas WILTING, 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions* [online]. **19**(11), 1366–1379 [vid. 2024-03-05]. ISSN 14724642. Dostupné z: doi:10.1111/ddi.12096

LANG, Nico, Walter JETZ, Konrad SCHINDLER a Jan Dirk WEGNER, 2023. A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*.

LEROY, Boris, Christine N. MEYNARD, Céline BELLARD a Franck COURCHAMP, 2016. virtualspecies, an R package to generate virtual species distributions. *Ecography* [online]. **39**(6), 599–607. ISSN 16000587. Dostupné z: doi:10.1111/ecog.01388

LOBO, Jorge M., Alberto JIMÉNEZ-VALVERDE a Raimundo REAL, 2008. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* [online]. **17**(2), 145–151 [vid. 2024-03-22]. ISSN 1466822X. Dostupné z: doi:10.1111/j.1466-8238.2007.00358.x

LOMOLINO, Mark V, 2004. Conservation Biogeography - Introduction. *Frontiers of Biogeography*. (1839), 293–297.

MELO-MERINO, Sara M., Héctor REYES-BONILLA a Andrés LIRA-NORIEGA, 2020. *Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence* [online]. 2020. ISSN 03043800. Dostupné z: doi:10.1016/j.ecolmodel.2019.108837

MEYNARD, Christine N. a David M. KAPLAN, 2013. Using virtual species to study species distributions and model performance. *Journal of Biogeography* [online]. **40**(1), 1–8 [vid. 2024-02-03]. ISSN 1365-2699. Dostupné z: doi:10.1111/JBI.12006

MEYNARD, Christine N., Boris LEROY a David M. KAPLAN, 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? *Ecography* [online]. **42**(12), 2021–2036 [vid. 2024-02-03]. ISSN 1600-0587. Dostupné z: doi:10.1111/ECOG.04385

MILLER, Jennifer, 2010. *Species distribution modeling* [online]. 2010. ISSN 17498198. Dostupné z: doi:10.1111/j.1749-8198.2010.00351.x

MITCHELL, Peter J, Jacquomo MONK a Laurie LAURENSEN, 2017. Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution* [online]. **8**(1), 12–21 [vid. 2024-02-21]. ISSN 2041210X. Dostupné z: doi:10.1111/2041-210X.12645

MONSARRAT, Sophie, Andre F. BOSHOFF a Graham I.H. KERLEY, 2019. Accessibility maps as a tool to predict sampling bias in historical biodiversity occurrence records. *Ecography* [online]. **42**(1), 125–136. ISSN 16000587. Dostupné z: doi:10.1111/ecog.03944

MOUDRÝ, Vítězslav, 2015. *Modelling species distributions with simulated virtual species* [online]. 1. srpen 2015. B.m.: John Wiley & Sons, Ltd. [vid. 2024-02-21]. ISSN 13652699. Dostupné z: doi:10.1111/jbi.12552

MOUDRÝ, Vítězslav, Vincent LECOURS, Kateřina GDULOVÁ, Lukáš GÁBOR, Lucie MOUDRÁ, Jan KROPÁČEK a Jan WILD, 2018. On the use of global DEMs in ecological modelling and the accuracy of new bare-earth DEMs. *Ecological Modelling* [online]. **383**, 3–9 [vid. 2024-03-26]. ISSN 03043800. Dostupné z: doi:10.1016/j.ecolmodel.2018.05.006

MOUDRÝ, Vítězslav, Lucie MOUDRÁ, Vojtěch BARTÁK, Vladimír BEJČEK, Kateřina GDULOVÁ, Markéta HENDRYCHOVÁ, David MORAVEC, Petr MUSIL, Duccio ROCCHINI, Karel ŠŤASTNÝ, Ondřej VOLF a Miroslav ŠÁLEK, 2021. The role of the vegetation structure, primary productivity and

senescence derived from airborne LiDAR and hyperspectral data for birds diversity and rarity on a restored site. *Landscape and Urban Planning* [online]. **210**, 104064 [vid. 2024-03-22]. ISSN 01692046. Dostupné z: doi:10.1016/j.landurbplan.2021.104064

MOUDRÝ, Vítězslav a Petra ŠÍMOVÁ, 2012. Influence of positional accuracy, sample size and scale on modelling species distributions: A review. *International Journal of Geographical Information Science* [online]. **26**(11), 2083–2095. ISSN 13658816. Dostupné z: doi:10.1080/13658816.2012.721553

NAIMI, Babak, Nicholas A.S. HAMM, Thomas A GROEN, Andrew K SKIDMORE a Albertus G TOXOPEUS, 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecography* [online]. **37**(2), 191–203 [vid. 2024-02-03]. ISSN 09067590. Dostupné z: doi:10.1111/j.1600-0587.2013.00205.x

OSBORNE, Patrick E. a Pedro J. LEITÃO, 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity and Distributions* [online]. **15**(4), 671–681 [vid. 2024-03-17]. ISSN 13669516. Dostupné z: doi:10.1111/j.1472-4642.2009.00572.x

PECCHI, Matteo, Maurizio MARCHI, Vanessa BURTON, Francesca GIANNETTI, Marco MORIONDO, Iacopo BERNETTI, Marco BINDI a Gherardo CHIRICI, 2019. Species distribution modelling to support forest management. A literature review [online]. [vid. 2024-02-26]. Dostupné z: doi:10.1016/j.ecolmodel.2019.108817

SCHOENER, Thomas W., 1968. The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna. *Ecology* [online]. **49**(4), 704–726 [vid. 2024-03-14]. ISSN 0012-9658. Dostupné z: doi:10.2307/1935534

SMITH, Adam B., Stephen J. MURPHY, David HENDERSON a Kelley D. ERICKSON, 2023. Including imprecisely georeferenced specimens improves accuracy of species distribution models and estimates of niche breadth. *Global Ecology and Biogeography* [online]. **32**(3), 342–355. ISSN 14668238. Dostupné z: doi:10.1111/geb.13628

SOLEY-GUARDIA, Mariano, Diego F. ALVARADO-SERRANO a Robert P.

ANDERSON, 2024. *Top ten hazards to avoid when modeling species distributions: a didactic guide of assumptions, problems, and recommendations* [online]. 2024. B.m.: John Wiley & Sons, Ltd. [vid. 2024-02-21]. ISSN 16000587. Dostupné z: doi:10.1111/ecog.06852

STOKLOSA, Jakub, Christopher DALY, Scott D. FOSTER, Michael B. ASHCROFT a David I. WARTON, 2015. A climate of uncertainty: Accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution* [online]. **6**(4), 412–423 [vid. 2024-03-05]. ISSN 2041210X. Dostupné z: doi:10.1111/2041-210X.12217

TULOWIECKI, Stephen J., Chris P.S. LARSEN a Yi Chen WANG, 2015. Effects of positional error on modeling species distributions: a perspective using presettlement land survey records. *Plant Ecology* [online]. **216**(1), 67–85. ISSN 15735052. Dostupné z: doi:10.1007/s11258-014-0417-9

VIERLING, Kerri T., Charles E. SWIFT, Andrew T. HUDAK, Jody C. VOGELER a Lee A. VIERLING, 2014. How much does the time lag between wildlife field-data collection and LiDAR-data acquisition matter for studies of animal distributions A case study using bird communities. *Remote Sensing Letters* [online]. **5**(2), 185–193. ISSN 21507058. Dostupné z: doi:10.1080/2150704X.2014.891773

VOLIS, Sergei a Komiljon TOJIBAEV, 2021. Defining critical habitat for plant species with poor occurrence knowledge and identification of critical habitat networks. *Biodiversity and Conservation* [online]. **30**(12), 3603–3611. ISSN 15729710. Dostupné z: doi:10.1007/s10531-021-02265-w

WARREN, Dan L., Richard E. GLOR a Michael TURELLI, 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* [online]. **62**(11), 2868–2883 [vid. 2024-02-24]. ISSN 00143820. Dostupné z: doi:10.1111/j.1558-5646.2008.00482.x

WILLIAMS, Kristen J, Lee BELBIN, Michael P AUSTIN, Janet L STEIN a Simon FERRIER, 2012. Which environmental variables should I use in my biodiversity model? *International Journal of Geographical Information Science* [online]. **26**(11), 2009–2047 [vid. 2024-03-27]. ISSN 13658816. Dostupné z: doi:10.1080/13658816.2012.698015

ZURELL, Damaris, Janet FRANKLIN, Christian KÖNIG, Phil J BOUCHET, Carsten F DORMANN, Jane ELITH, Guillermo FANDOS, Xiao FENG, Gurutzeta GUILLERA-ARROITA, Antoine GUISAN, José J. LAHOZ-MONFORT, Pedro J LEITÃO, Daniel S PARK, A Townsend PETERSON, Giovanni RAPACCIUOLO, Dirk R SCHMATZ, Boris SCHRÖDER, Josep M. SERRA-DIAZ, Wilfried THUILLER, Katherine L YATES, Niklaus E ZIMMERMANN a Cory MEROW, 2020. A standard protocol for reporting species distribution models. *Ecography* [online]. **43**(9), 1261–1277. ISSN 16000587. Dostupné z: doi:10.1111/ecog.04960

8 Přílohy

Příloha A – Skript v R zahrnující podstatné metodické kroky

Část 1 – Načtení balíčků, dat a jejich příprava

```
# nacteni balicku
library(PresenceAbsence)
library(virtualspecies)
library(raster)
library(tidyverse)
library(sf)
library(dismo)

# nastaveni adresare
setwd("*cesta/do/adresare*")

# vstupni data
krnap <- read_sf(dsn = "KRMAP/KRMAP_hranice.shp")
CHM_ALS.2012 <- raster("CHM_ALS2012/chm_2012.tif")
CHM_ALS.2022 <- raster("CHM_ALS2022/chm_2022.tif")
CHM_GEDI <- raster("CHM_GEDI/chm_GEDI.tif")

# nastaveni NA hodnot
CHM_ALS.2022[is.na(CHM_ALS.2022)] <- 0
CHM_ALS.2022[CHM_ALS.2022 < 0 | CHM_ALS.2022 > 45 ] <- NA
CHM_ALS.2022 <- mask(CHM_ALS.2022, krnap)
CHM_GEDI[CHM_GEDI > 60] <- NA
names(CHM_ALS.2012) <- c("CHM_ALS.2012")
names(CHM_ALS.2022) <- c("CHM_ALS.2022")
names(CHM_GEDI) <- c("CHM_GEDI")

# priprava dat
stack_ALS.2012 <- stack(CHM_ALS.2012)
stack_ALS.2022 <- stack(CHM_ALS.2022)
stack_gedi <- stack(CHM_GEDI)
names(stack_ALS.2012) <- c("CHM_ALS.2012")
names(stack_ALS.2022) <- c("CHM_ALS.2022")
```

```
names(stack_gedi) <- c("CHM_GEDI")
mask <- CHM_ALS.2022
```

Část 2 – Generování virtuálního druhu

```
# generovani virtual species
```

```
print("generovani suitability rastru a virtual species")
```

```
params_ALS.2022 <- formatFunctions(CHM_ALS.2022 = c(fun = 'dnorm',
mean = 20, sd = 5))
```

```
VS_ALS.2022 <- generateSpFromFun(raster.stack = CHM_ALS.2022,
parameters = params_ALS.2022, plot = FALSE)
```

```
pred1 <- as.data.frame(VS_ALS.2022$suitab.raster@data@values)
```

Část 3 – Definování funkcí

```
# funkce limitující posun pouze do oblasti zajmu
```

```
inside <- function(x, y, mask){
```

```
  ins <- ifelse(x < mask@extent@xmin | x > mask@extent@xmax | y <
mask@extent@ymin | y > mask@extent@ymax, FALSE, TRUE)
```

```
  if (ins == TRUE) {
```

```
    val <- extract(mask, SpatialPoints(coords = matrix(c(x,y), ncol
= 2)))
```

```
    ins <- ifelse(is.na(val), FALSE, TRUE)
```

```
  }
```

```
  return(ins)
```

```
}
```

```
# funkce posouva zaznamy podle zadane vzdalenosti
```

```
shift <- function(x, y, dist, mask){
```

```
  new.x <- x
```

```

new.y <- y
inside <- FALSE
while (!inside){
  angle <- runif(1,0,360)*pi/180
  new.x <- x + dist*sin(angle)
  new.y <- y + dist*cos(angle)
  inside <- inside(new.x, new.y, mask)
}
return(c(new.x, new.y))
}

# funkce pro rozdeleni na trenovaci a testovaci sadu
splitdf <- function(dataframe, seed=NULL) {
  if (!is.null(seed)) set.seed(seed)
  index <- 1:nrow(dataframe)
  trainindex <- sample(index, trunc(length(index)*5/10))
  trainset <- dataframe[trainindex, ]
  testset <- dataframe[-trainindex, ]
  list(trainset=trainset,testset=testset)
}

# funkce pro niche overlap
index <- function (pred1, pred2, na.rm = TRUE)
{
  p1 <- pred1/sum(pred1, na.rm = na.rm)
  p2 <- pred2/sum(pred2, na.rm = na.rm)
  SchoenerD <- 1 - 0.5 * sum(abs(p1 - p2), na.rm = na.rm)
}

```

```

    HellingerDist <- sqrt(sum((sqrt(p1) - sqrt(p2))^2, na.rm =
na.rm))

    WarrenI <- 1 - ((HellingerDist^2)/2)

    data.frame(SchoenerD = SchoenerD, WarrenI = WarrenI,
HellingerDist = HellingerDist)
}

```

Část 4 – Cyklus pro vytvoření nálezoých dat, simulování polohové chyby, tvorbu modelů a výpočet metrik

FOR CYKLUS

```
for (i in 1:20) {
```

prevedeni na P/A

```
  print("prevedeni na P/A")
```

```
  PA.raster_ALS.2022 <- convertToPA(VS_ALS.2022, alpha = -0.05,
beta = 0.3, plot = FALSE)
```

sampling occurences

```
  print("sampling occurences")
```

```
  PA.sampling_ALS.2022 <- sampleOccurrences(PA.raster_ALS.2022, n =
301, type = "presence-absence", plot = FALSE, sample.prevalence =
0.33)
```

priprava dat na posun zaznamu

```
  print("priprava dat na posun zaznamu")
```

```
  DATA.Prep_ALS.2022 <-
as.data.frame(PA.sampling_ALS.2022$sample.points)
```

```
  s_ALS.2022 <- nrow(DATA.Prep_ALS.2022)
```

```
  DATA.Prep_ALS.2022$ID <- 1:s_ALS.2022
```

```
  split_ALS.2022 <- splitdf(DATA.Prep_ALS.2022)
```

```
  Train_ALS.2022 <- split_ALS.2022$trainset
```

```

Test_ALS.2022 <- split_ALS.2022$testset

# posun E1 - E5 -> E1 = 5-10m; E2 = 10-50m; E3 = 50-100m; E4 = 100-
300m; E5 = 300-500m

print("posun ALS2022")

Train_ALS.2022[,c("E1x", "E1y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 5, 10),mask=mask)))

Train_ALS.2022[,c("E2x", "E2y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 5, 50),mask=mask)))

Train_ALS.2022[,c("E3x", "E3y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 5,
100),mask=mask)))

Train_ALS.2022[,c("E4x", "E4y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 5,
300),mask=mask)))

Train_ALS.2022[,c("E5x", "E5y")] <- t(apply(Train_ALS.2022, 1,
function(.) shift(.["x"], .["y"], dist=runif(1, 5,
500),mask=mask)))

# pojmenovani sloupcu a ulozeni hodnot

# ALS2012

print("pojmenovani sloupcu a ulozeni hodnot")

ID <- Train_ALS.2022$ID

XY_E0 <- Train_ALS.2022[,1:2]

XY_E1 <- Train_ALS.2022[,6:7]

XY_E2 <- Train_ALS.2022[,8:9]

XY_E3 <- Train_ALS.2022[,10:11]

XY_E4 <- Train_ALS.2022[,12:13]

XY_E5 <- Train_ALS.2022[,14:15]

Observed_Train <- Train_ALS.2022[4]

```



```

Value_E0.A.2012 <- data.frame(extract(CHM_ALS.2012,
SpatialPoints(coords = XY_E0)))

Value_E1.A.2012 <- data.frame(extract(CHM_ALS.2012,
SpatialPoints(coords = XY_E1)))

Value_E2.A.2012 <- data.frame(extract(CHM_ALS.2012,
SpatialPoints(coords = XY_E2)))

Value_E3.A.2012 <- data.frame(extract(CHM_ALS.2012,
SpatialPoints(coords = XY_E3)))

Value_E4.A.2012 <- data.frame(extract(CHM_ALS.2012,
SpatialPoints(coords = XY_E4)))

Value_E5.A.2012 <- data.frame(extract(CHM_ALS.2012,
SpatialPoints(coords = XY_E5)))

ALS.2012_Train.E0 <- data.frame(ID, XY_E0, Observed_Train,
Value_E0.A.2012)

names(ALS.2012_Train.E0) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2012")

ALS.2012_Train.E1 <- data.frame(ID, XY_E1, Observed_Train,
Value_E1.A.2012)

names(ALS.2012_Train.E1) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2012")

ALS.2012_Train.E2 <- data.frame(ID, XY_E2, Observed_Train,
Value_E2.A.2012)

names(ALS.2012_Train.E2) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2012")

ALS.2012_Train.E3 <- data.frame(ID, XY_E3, Observed_Train,
Value_E3.A.2012)

names(ALS.2012_Train.E3) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2012")

ALS.2012_Train.E4 <- data.frame(ID, XY_E4, Observed_Train,
Value_E4.A.2012)

```

```

names(ALS.2012_Train.E4) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2012")

ALS.2012_Train.E5 <- data.frame(ID, XY_E5, Observed_Train,
Value_E5.A.2012)

names(ALS.2012_Train.E5) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2012")

ID <- Test_ALS.2022$ID

XY <- Test_ALS.2022[,1:2]

Observed <- Test_ALS.2022[4]

Value_T.A.2012 <- data.frame(extract(CHM_ALS.2012,
SpatialPoints(coords = XY)))

ALS.2012_Test <- data.frame(ID, XY, Observed, Value_T.A.2012)

names(ALS.2012_Test) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2012")

# ALS2022

ID <- Train_ALS.2022$ID

XY_E0 <- Train_ALS.2022[,1:2]

XY_E1 <- Train_ALS.2022[,6:7]

XY_E2 <- Train_ALS.2022[,8:9]

XY_E3 <- Train_ALS.2022[,10:11]

XY_E4 <- Train_ALS.2022[,12:13]

XY_E5 <- Train_ALS.2022[,14:15]

Observed_Train <- Train_ALS.2022[4]

Value_E0.A.2022 <- data.frame(extract(CHM_ALS.2022,
SpatialPoints(coords = XY_E0)))

Value_E1.A.2022 <- data.frame(extract(CHM_ALS.2022,
SpatialPoints(coords = XY_E1)))

```

```

Value_E2.A.2022 <- data.frame(extract(CHM_ALS.2022,
SpatialPoints(coords = XY_E2)))

Value_E3.A.2022 <- data.frame(extract(CHM_ALS.2022,
SpatialPoints(coords = XY_E3)))

Value_E4.A.2022 <- data.frame(extract(CHM_ALS.2022,
SpatialPoints(coords = XY_E4)))

Value_E5.A.2022 <- data.frame(extract(CHM_ALS.2022,
SpatialPoints(coords = XY_E5)))

ALS.2022_Train.E0 <- data.frame(ID, XY_E0, Observed_Train,
Value_E0.A.2022)

names(ALS.2022_Train.E0) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2022")

ALS.2022_Train.E1 <- data.frame(ID, XY_E1, Observed_Train,
Value_E1.A.2022)

names(ALS.2022_Train.E1) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2022")

ALS.2022_Train.E2 <- data.frame(ID, XY_E2, Observed_Train,
Value_E2.A.2022)

names(ALS.2022_Train.E2) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2022")

ALS.2022_Train.E3 <- data.frame(ID, XY_E3, Observed_Train,
Value_E3.A.2022)

names(ALS.2022_Train.E3) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2022")

ALS.2022_Train.E4 <- data.frame(ID, XY_E4, Observed_Train,
Value_E4.A.2022)

names(ALS.2022_Train.E4) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2022")

ALS.2022_Train.E5 <- data.frame(ID, XY_E5, Observed_Train,
Value_E5.A.2022)

```

```

names(ALS.2022_Train.E5) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2022")

ID <- Test_ALS.2022$ID

XY <- Test_ALS.2022[,1:2]

Observed <- Test_ALS.2022[4]

Value_T.A.2022 <- data.frame(extract(CHM_ALS.2022,
SpatialPoints(coords = XY)))

ALS.2022_Test <- data.frame(ID, XY, Observed, Value_T.A.2022)

names(ALS.2022_Test) <- c("ID", "X", "Y", "Observed",
"CHM_ALS.2022")

```

GEDI

```

ID <- Train_ALS.2022$ID

XY_E0 <- Train_ALS.2022[,1:2]

XY_E1 <- Train_ALS.2022[,6:7]

XY_E2 <- Train_ALS.2022[,8:9]

XY_E3 <- Train_ALS.2022[,10:11]

XY_E4 <- Train_ALS.2022[,12:13]

XY_E5 <- Train_ALS.2022[,14:15]

Observed_Train <- Train_ALS.2022[4]

Value_E0.G <- data.frame(extract(CHM_GEDI, SpatialPoints(coords =
XY_E0)))

Value_E1.G <- data.frame(extract(CHM_GEDI, SpatialPoints(coords =
XY_E1)))

Value_E2.G <- data.frame(extract(CHM_GEDI, SpatialPoints(coords =
XY_E2)))

Value_E3.G <- data.frame(extract(CHM_GEDI, SpatialPoints(coords =
XY_E3)))

```

```

Value_E4.G <- data.frame(extract(CHM_GEDI, SpatialPoints(coords =
XY_E4)))

Value_E5.G <- data.frame(extract(CHM_GEDI, SpatialPoints(coords =
XY_E5)))

GEDI_Train.E0 <- data.frame(ID, XY_E0, Observed_Train,
Value_E0.G)

names(GEDI_Train.E0) <- c("ID", "X", "Y", "Observed", "CHM_GEDI")

GEDI_Train.E1 <- data.frame(ID, XY_E1, Observed_Train,
Value_E1.G)

names(GEDI_Train.E1) <- c("ID", "X", "Y", "Observed", "CHM_GEDI")

GEDI_Train.E2 <- data.frame(ID, XY_E2, Observed_Train,
Value_E2.G)

names(GEDI_Train.E2) <- c("ID", "X", "Y", "Observed", "CHM_GEDI")

GEDI_Train.E3 <- data.frame(ID, XY_E3, Observed_Train,
Value_E3.G)

names(GEDI_Train.E3) <- c("ID", "X", "Y", "Observed", "CHM_GEDI")

GEDI_Train.E4 <- data.frame(ID, XY_E4, Observed_Train,
Value_E4.G)

names(GEDI_Train.E4) <- c("ID", "X", "Y", "Observed", "CHM_GEDI")

GEDI_Train.E5 <- data.frame(ID, XY_E5, Observed_Train,
Value_E5.G)

names(GEDI_Train.E5) <- c("ID", "X", "Y", "Observed", "CHM_GEDI")

ID <- Test_ALS.2022$ID
XY <- Test_ALS.2022[,1:2]
Observed <- Test_ALS.2022[4]

Value_T.G <- data.frame(extract(CHM_GEDI, SpatialPoints(coords =
XY)))

```

```

GEDI_Test <- data.frame(ID, XY, Observed, Value_T.G)

names(GEDI_Test) <- c("ID", "X", "Y", "Observed", "CHM_GEDI")

# GLM model a predikce novych hodnot

# ALS2012

print("GLM model a predikce novych hodnot ALS2012")

Model_E0.A.2012 <- glm(formula =
Observed~CHM_ALS.2012+I(CHM_ALS.2012^2), family = binomial, data =
ALS.2012_Train.E0)

Model_E1.A.2012 <- glm(formula =
Observed~CHM_ALS.2012+I(CHM_ALS.2012^2), family = binomial, data =
ALS.2012_Train.E1)

Model_E2.A.2012 <- glm(formula =
Observed~CHM_ALS.2012+I(CHM_ALS.2012^2), family = binomial, data =
ALS.2012_Train.E2)

Model_E3.A.2012 <- glm(formula =
Observed~CHM_ALS.2012+I(CHM_ALS.2012^2), family = binomial, data =
ALS.2012_Train.E3)

Model_E4.A.2012 <- glm(formula =
Observed~CHM_ALS.2012+I(CHM_ALS.2012^2), family = binomial, data =
ALS.2012_Train.E4)

Model_E5.A.2012 <- glm(formula =
Observed~CHM_ALS.2012+I(CHM_ALS.2012^2), family = binomial, data =
ALS.2012_Train.E5)

predict_E0.A.2012 <- predict(Model_E0.A.2012, newdata =
ALS.2012_Test, type = "response")

predict_E1.A.2012 <- predict(Model_E1.A.2012, newdata =
ALS.2012_Test, type = "response")

predict_E2.A.2012 <- predict(Model_E2.A.2012, newdata =
ALS.2012_Test, type = "response")

```

```

predict_E3.A.2012 <- predict(Model_E3.A.2012, newdata =
ALS.2012_Test, type = "response")

predict_E4.A.2012 <- predict(Model_E4.A.2012, newdata =
ALS.2012_Test, type = "response")

predict_E5.A.2012 <- predict(Model_E5.A.2012, newdata =
ALS.2012_Test, type = "response")

pred_index_E0.A.2012 <- predict(stack_ALS.2012, Model_E0.A.2012,
type = "response")

pred_index_E1.A.2012 <- predict(stack_ALS.2012, Model_E1.A.2012,
type = "response")

pred_index_E2.A.2012 <- predict(stack_ALS.2012, Model_E2.A.2012,
type = "response")

pred_index_E3.A.2012 <- predict(stack_ALS.2012, Model_E3.A.2012,
type = "response")

pred_index_E4.A.2012 <- predict(stack_ALS.2012, Model_E4.A.2012,
type = "response")

pred_index_E5.A.2012 <- predict(stack_ALS.2012, Model_E5.A.2012,
type = "response")

predicted_E0.A.2012 <- as.vector(predict_E0.A.2012)
predicted_E1.A.2012 <- as.vector(predict_E1.A.2012)
predicted_E2.A.2012 <- as.vector(predict_E2.A.2012)
predicted_E3.A.2012 <- as.vector(predict_E3.A.2012)
predicted_E4.A.2012 <- as.vector(predict_E4.A.2012)
predicted_E5.A.2012 <- as.vector(predict_E5.A.2012)

# ALS2022

print("GLM model a predikce novych hodnot ALS2022")

Model_E0.A.2022 <- glm(formula =
Observed~CHM_ALS.2022+I(CHM_ALS.2022^2), family = binomial, data =
ALS.2022_Train.E0)

```

```
Model_E1.A.2022 <- glm(formula =  
Observed~CHM_ALS.2022+I(CHM_ALS.2022^2), family = binomial, data =  
ALS.2022_Train.E1)
```

```
Model_E2.A.2022 <- glm(formula =  
Observed~CHM_ALS.2022+I(CHM_ALS.2022^2), family = binomial, data =  
ALS.2022_Train.E2)
```

```
Model_E3.A.2022 <- glm(formula =  
Observed~CHM_ALS.2022+I(CHM_ALS.2022^2), family = binomial, data =  
ALS.2022_Train.E3)
```

```
Model_E4.A.2022 <- glm(formula =  
Observed~CHM_ALS.2022+I(CHM_ALS.2022^2), family = binomial, data =  
ALS.2022_Train.E4)
```

```
Model_E5.A.2022 <- glm(formula =  
Observed~CHM_ALS.2022+I(CHM_ALS.2022^2), family = binomial, data =  
ALS.2022_Train.E5)
```

```
predict_E0.A.2022 <- predict(Model_E0.A.2022, newdata =  
ALS.2022_Test, type = "response")
```

```
predict_E1.A.2022 <- predict(Model_E1.A.2022, newdata =  
ALS.2022_Test, type = "response")
```

```
predict_E2.A.2022 <- predict(Model_E2.A.2022, newdata =  
ALS.2022_Test, type = "response")
```

```
predict_E3.A.2022 <- predict(Model_E3.A.2022, newdata =  
ALS.2022_Test, type = "response")
```

```
predict_E4.A.2022 <- predict(Model_E4.A.2022, newdata =  
ALS.2022_Test, type = "response")
```

```
predict_E5.A.2022 <- predict(Model_E5.A.2022, newdata =  
ALS.2022_Test, type = "response")
```

```
pred_index_E0.A.2022 <- predict(stack_ALS.2022, Model_E0.A.2022,  
type = "response")
```

```
pred_index_E1.A.2022 <- predict(stack_ALS.2022, Model_E1.A.2022,  
type = "response")
```



```

    pred_index_E2.A.2022 <- predict(stack_ALS.2022, Model_E2.A.2022,
type = "response")

    pred_index_E3.A.2022 <- predict(stack_ALS.2022, Model_E3.A.2022,
type = "response")

    pred_index_E4.A.2022 <- predict(stack_ALS.2022, Model_E4.A.2022,
type = "response")

    pred_index_E5.A.2022 <- predict(stack_ALS.2022, Model_E5.A.2022,
type = "response")

    predicted_E0.A.2022 <- as.vector(predict_E0.A.2022)

    predicted_E1.A.2022 <- as.vector(predict_E1.A.2022)

    predicted_E2.A.2022 <- as.vector(predict_E2.A.2022)

    predicted_E3.A.2022 <- as.vector(predict_E3.A.2022)

    predicted_E4.A.2022 <- as.vector(predict_E4.A.2022)

    predicted_E5.A.2022 <- as.vector(predict_E5.A.2022)

# GEDI

    print("GLM model a predikce novych hodnot GEDI")

    Model_E0.G <- glm(formula = Observed~CHM_GEDI+I(CHM_GEDI^2),
family = binomial, data = GEDI_Train.E0)

    Model_E1.G <- glm(formula = Observed~CHM_GEDI+I(CHM_GEDI^2),
family = binomial, data = GEDI_Train.E1)

    Model_E2.G <- glm(formula = Observed~CHM_GEDI+I(CHM_GEDI^2),
family = binomial, data = GEDI_Train.E2)

    Model_E3.G <- glm(formula = Observed~CHM_GEDI+I(CHM_GEDI^2),
family = binomial, data = GEDI_Train.E3)

    Model_E4.G <- glm(formula = Observed~CHM_GEDI+I(CHM_GEDI^2),
family = binomial, data = GEDI_Train.E4)

    Model_E5.G <- glm(formula = Observed~CHM_GEDI+I(CHM_GEDI^2),
family = binomial, data = GEDI_Train.E5)

```

```

predict_E0.G <- predict(Model_E0.G, newdata = GEDI_Test, type =
"response")

predict_E1.G <- predict(Model_E1.G, newdata = GEDI_Test, type =
"response")

predict_E2.G <- predict(Model_E2.G, newdata = GEDI_Test, type =
"response")

predict_E3.G <- predict(Model_E3.G, newdata = GEDI_Test, type =
"response")

predict_E4.G <- predict(Model_E4.G, newdata = GEDI_Test, type =
"response")

predict_E5.G <- predict(Model_E5.G, newdata = GEDI_Test, type =
"response")

pred_index_E0.G <- predict(stack_gedi, Model_E0.G, type =
"response")

pred_index_E1.G <- predict(stack_gedi, Model_E1.G, type =
"response")

pred_index_E2.G <- predict(stack_gedi, Model_E2.G, type =
"response")

pred_index_E3.G <- predict(stack_gedi, Model_E3.G, type =
"response")

pred_index_E4.G <- predict(stack_gedi, Model_E4.G, type =
"response")

pred_index_E5.G <- predict(stack_gedi, Model_E5.G, type =
"response")

predicted_E0.G <- as.vector(predict_E0.G)

predicted_E1.G <- as.vector(predict_E1.G)

predicted_E2.G <- as.vector(predict_E2.G)

predicted_E3.G <- as.vector(predict_E3.G)

predicted_E4.G <- as.vector(predict_E4.G)

```

```

predicted_E5.G <- as.vector(predict_E5.G)

# vypocet AUC

# ALS2012

print("vypocet AUC ALS2012")

DAT_E0.A.2012 <- data.frame(ALS.2012_Test$ID,
ALS.2012_Test$Observed, predicted_E0.A.2012)

DAT_E0.A.2012 <- na.omit(DAT_E0.A.2012)

DAT_E1.A.2012 <- data.frame(ALS.2012_Test$ID,
ALS.2012_Test$Observed, predicted_E1.A.2012)

DAT_E1.A.2012 <- na.omit(DAT_E1.A.2012)

DAT_E2.A.2012 <- data.frame(ALS.2012_Test$ID,
ALS.2012_Test$Observed, predicted_E2.A.2012)

DAT_E2.A.2012 <- na.omit(DAT_E2.A.2012)

DAT_E3.A.2012 <- data.frame(ALS.2012_Test$ID,
ALS.2012_Test$Observed, predicted_E3.A.2012)

DAT_E3.A.2012 <- na.omit(DAT_E3.A.2012)

DAT_E4.A.2012 <- data.frame(ALS.2012_Test$ID,
ALS.2012_Test$Observed, predicted_E4.A.2012)

DAT_E4.A.2012 <- na.omit(DAT_E4.A.2012)

DAT_E5.A.2012 <- data.frame(ALS.2012_Test$ID,
ALS.2012_Test$Observed, predicted_E5.A.2012)

DAT_E5.A.2012 <- na.omit(DAT_E5.A.2012)

AUC_E0.A.2012 <- auc(DAT_E0.A.2012, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E0.A.2012 <- rbind(AUC_ALL_E0.A.2012, AUC_E0.A.2012)

AUC_E1.A.2012 <- auc(DAT_E1.A.2012, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E1.A.2012 <- rbind(AUC_ALL_E1.A.2012, AUC_E1.A.2012)

```

```

AUC_E2.A.2012 <- auc(DAT_E2.A.2012, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E2.A.2012 <- rbind(AUC_ALL_E2.A.2012, AUC_E2.A.2012)

AUC_E3.A.2012 <- auc(DAT_E3.A.2012, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E3.A.2012 <- rbind(AUC_ALL_E3.A.2012, AUC_E3.A.2012)

AUC_E4.A.2012 <- auc(DAT_E4.A.2012, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E4.A.2012 <- rbind(AUC_ALL_E4.A.2012, AUC_E4.A.2012)

AUC_E5.A.2012 <- auc(DAT_E5.A.2012, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E5.A.2012 <- rbind(AUC_ALL_E5.A.2012, AUC_E5.A.2012)

# ALS2022

print("vypocet AUC ALS2022")

DAT_E0.A.2022 <- data.frame(ALS.2022_Test$ID,
ALS.2022_Test$Observed, predicted_E0.A.2022)

DAT_E0.A.2022 <- na.omit(DAT_E0.A.2022)

DAT_E1.A.2022 <- data.frame(ALS.2022_Test$ID,
ALS.2022_Test$Observed, predicted_E1.A.2022)

DAT_E1.A.2022 <- na.omit(DAT_E1.A.2022)

DAT_E2.A.2022 <- data.frame(ALS.2022_Test$ID,
ALS.2022_Test$Observed, predicted_E2.A.2022)

DAT_E2.A.2022 <- na.omit(DAT_E2.A.2022)

DAT_E3.A.2022 <- data.frame(ALS.2022_Test$ID,
ALS.2022_Test$Observed, predicted_E3.A.2022)

DAT_E3.A.2022 <- na.omit(DAT_E3.A.2022)

```

```

DAT_E4.A.2022 <- data.frame(ALS.2022_Test$ID,
ALS.2022_Test$Observed, predicted_E4.A.2022)

DAT_E4.A.2022 <- na.omit(DAT_E4.A.2022)

DAT_E5.A.2022 <- data.frame(ALS.2022_Test$ID,
ALS.2022_Test$Observed, predicted_E5.A.2022)

DAT_E5.A.2022 <- na.omit(DAT_E5.A.2022)

AUC_E0.A.2022 <- auc(DAT_E0.A.2022, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E0.A.2022 <- rbind(AUC_ALL_E0.A.2022, AUC_E0.A.2022)

AUC_E1.A.2022 <- auc(DAT_E1.A.2022, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E1.A.2022 <- rbind(AUC_ALL_E1.A.2022, AUC_E1.A.2022)

AUC_E2.A.2022 <- auc(DAT_E2.A.2022, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E2.A.2022 <- rbind(AUC_ALL_E2.A.2022, AUC_E2.A.2022)

AUC_E3.A.2022 <- auc(DAT_E3.A.2022, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E3.A.2022 <- rbind(AUC_ALL_E3.A.2022, AUC_E3.A.2022)

AUC_E4.A.2022 <- auc(DAT_E4.A.2022, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E4.A.2022 <- rbind(AUC_ALL_E4.A.2022, AUC_E4.A.2022)

AUC_E5.A.2022 <- auc(DAT_E5.A.2022, st.dev = TRUE, which.model =
1, na.rm = FALSE)

AUC_ALL_E5.A.2022 <- rbind(AUC_ALL_E5.A.2022, AUC_E5.A.2022)

# GEDI

print("vypocet AUC GEDI")

DAT_E0.G <- data.frame(GEDI_Test$ID, GEDI_Test$Observed,
predicted_E0.G)

```

```

DAT_E0.G <- na.omit(DAT_E0.G)

DAT_E1.G <- data.frame(GEDI_Test$ID, GEDI_Test$Observed,
predicted_E1.G)

DAT_E1.G <- na.omit(DAT_E1.G)

DAT_E2.G <- data.frame(GEDI_Test$ID, GEDI_Test$Observed,
predicted_E2.G)

DAT_E2.G <- na.omit(DAT_E2.G)

DAT_E3.G <- data.frame(GEDI_Test$ID, GEDI_Test$Observed,
predicted_E3.G)

DAT_E3.G <- na.omit(DAT_E3.G)

DAT_E4.G <- data.frame(GEDI_Test$ID, GEDI_Test$Observed,
predicted_E4.G)

DAT_E4.G <- na.omit(DAT_E4.G)

DAT_E5.G <- data.frame(GEDI_Test$ID, GEDI_Test$Observed,
predicted_E5.G)

DAT_E5.G <- na.omit(DAT_E5.G)

AUC_E0.G <- auc(DAT_E0.G, st.dev = TRUE, which.model = 1, na.rm =
FALSE)

AUC_ALL_E0.G <- rbind(AUC_ALL_E0.G, AUC_E0.G)

AUC_E1.G <- auc(DAT_E1.G, st.dev = TRUE, which.model = 1, na.rm =
FALSE)

AUC_ALL_E1.G <- rbind(AUC_ALL_E1.G, AUC_E1.G)

AUC_E2.G <- auc(DAT_E2.G, st.dev = TRUE, which.model = 1, na.rm =
FALSE)

AUC_ALL_E2.G <- rbind(AUC_ALL_E2.G, AUC_E2.G)

AUC_E3.G <- auc(DAT_E3.G, st.dev = TRUE, which.model = 1, na.rm =
FALSE)

AUC_ALL_E3.G <- rbind(AUC_ALL_E3.G, AUC_E3.G)

```

```

AUC_E4.G <- auc(DAT_E4.G, st.dev = TRUE, which.model = 1, na.rm =
FALSE)

AUC_ALL_E4.G <- rbind(AUC_ALL_E4.G, AUC_E4.G)

AUC_E5.G <- auc(DAT_E5.G, st.dev = TRUE, which.model = 1, na.rm =
FALSE)

AUC_ALL_E5.G <- rbind(AUC_ALL_E5.G, AUC_E5.G)

# vypocet niche overlap

# ALS2012

print("vypocet overlap ALS2012")

pred2_E0.A.2012 <- as.data.frame(pred_index_E0.A.2012)
pred2_E1.A.2012 <- as.data.frame(pred_index_E1.A.2012)
pred2_E2.A.2012 <- as.data.frame(pred_index_E2.A.2012)
pred2_E3.A.2012 <- as.data.frame(pred_index_E3.A.2012)
pred2_E4.A.2012 <- as.data.frame(pred_index_E4.A.2012)
pred2_E5.A.2012 <- as.data.frame(pred_index_E5.A.2012)

NO_E0.A.2012 <- index(pred1, pred2_E0.A.2012, na.rm = TRUE)
NO_ALL_E0.A.2012 <- rbind(NO_ALL_E0.A.2012,NO_E0.A.2012)
NO_E1.A.2012 <- index(pred1, pred2_E1.A.2012, na.rm = TRUE)
NO_ALL_E1.A.2012 <- rbind(NO_ALL_E1.A.2012,NO_E1.A.2012)
NO_E2.A.2012 <- index(pred1, pred2_E2.A.2012, na.rm = TRUE)
NO_ALL_E2.A.2012 <- rbind(NO_ALL_E2.A.2012,NO_E2.A.2012)
NO_E3.A.2012 <- index(pred1, pred2_E3.A.2012, na.rm = TRUE)
NO_ALL_E3.A.2012 <- rbind(NO_ALL_E3.A.2012,NO_E3.A.2012)
NO_E4.A.2012 <- index(pred1, pred2_E4.A.2012, na.rm = TRUE)
NO_ALL_E4.A.2012 <- rbind(NO_ALL_E4.A.2012,NO_E4.A.2012)

```

```

NO_E5.A.2012 <- index(pred1, pred2_E5.A.2012, na.rm = TRUE)
NO_ALL_E5.A.2012 <- rbind(NO_ALL_E5.A.2012,NO_E5.A.2012)

# ALS2022

print("vypocet overlap ALS2022")

pred2_E0.A.2022 <- as.data.frame(pred_index_E0.A.2022)
pred2_E1.A.2022 <- as.data.frame(pred_index_E1.A.2022)
pred2_E2.A.2022 <- as.data.frame(pred_index_E2.A.2022)
pred2_E3.A.2022 <- as.data.frame(pred_index_E3.A.2022)
pred2_E4.A.2022 <- as.data.frame(pred_index_E4.A.2022)
pred2_E5.A.2022 <- as.data.frame(pred_index_E5.A.2022)

NO_E0.A.2022 <- index(pred1, pred2_E0.A.2022, na.rm = TRUE)
NO_ALL_E0.A.2022 <- rbind(NO_ALL_E0.A.2022,NO_E0.A.2022)
NO_E1.A.2022 <- index(pred1, pred2_E1.A.2022, na.rm = TRUE)
NO_ALL_E1.A.2022 <- rbind(NO_ALL_E1.A.2022,NO_E1.A.2022)
NO_E2.A.2022 <- index(pred1, pred2_E2.A.2022, na.rm = TRUE)
NO_ALL_E2.A.2022 <- rbind(NO_ALL_E2.A.2022,NO_E2.A.2022)
NO_E3.A.2022 <- index(pred1, pred2_E3.A.2022, na.rm = TRUE)
NO_ALL_E3.A.2022 <- rbind(NO_ALL_E3.A.2022,NO_E3.A.2022)
NO_E4.A.2022 <- index(pred1, pred2_E4.A.2022, na.rm = TRUE)
NO_ALL_E4.A.2022 <- rbind(NO_ALL_E4.A.2022,NO_E4.A.2022)
NO_E5.A.2022 <- index(pred1, pred2_E5.A.2022, na.rm = TRUE)
NO_ALL_E5.A.2022 <- rbind(NO_ALL_E5.A.2022,NO_E5.A.2022)

# GEDI

print("vypocet overlap GEDI")

pred2_E0.G <- as.data.frame(pred_index_E0.G)
pred2_E1.G <- as.data.frame(pred_index_E1.G)

```



```

pred2_E2.G <- as.data.frame(pred_index_E2.G)
pred2_E3.G <- as.data.frame(pred_index_E3.G)
pred2_E4.G <- as.data.frame(pred_index_E4.G)
pred2_E5.G <- as.data.frame(pred_index_E5.G)
NO_E0.G <- index(pred1, pred2_E0.G, na.rm = TRUE)
NO_ALL_E0.G <- rbind(NO_ALL_E0.G,NO_E0.G)
NO_E1.G <- index(pred1, pred2_E1.G, na.rm = TRUE)
NO_ALL_E1.G <- rbind(NO_ALL_E1.G,NO_E1.G)
NO_E2.G <- index(pred1, pred2_E2.G, na.rm = TRUE)
NO_ALL_E2.G <- rbind(NO_ALL_E2.G,NO_E2.G)
NO_E3.G <- index(pred1, pred2_E3.G, na.rm = TRUE)
NO_ALL_E3.G <- rbind(NO_ALL_E3.G,NO_E3.G)
NO_E4.G <- index(pred1, pred2_E4.G, na.rm = TRUE)
NO_ALL_E4.G <- rbind(NO_ALL_E4.G,NO_E4.G)
NO_E5.G <- index(pred1, pred2_E5.G, na.rm = TRUE)
NO_ALL_E5.G <- rbind(NO_ALL_E5.G,NO_E5.G)
}

```

Část 5 – Spojení výsledků a tvorba grafů

```
# spojeni vysledku
```

```
# ALS2012
```

```
AUC_ALL_E0.A.2012$Error <- "E0"
```

```
AUC_ALL_E1.A.2012$Error <- "E1"
```

```
AUC_ALL_E2.A.2012$Error <- "E2"
```

```
AUC_ALL_E3.A.2012$Error <- "E3"
```

```
AUC_ALL_E4.A.2012$Error <- "E4"
```

```

AUC_ALL_E5.A.2012$Error <- "E5"

AUC_ALL_A.2012 <- rbind(AUC_ALL_E0.A.2012, AUC_ALL_E1.A.2012,
AUC_ALL_E2.A.2012, AUC_ALL_E3.A.2012, AUC_ALL_E4.A.2012,
AUC_ALL_E5.A.2012)

NO_ALL_E0.A.2012$Error <- "E0"

NO_ALL_E1.A.2012$Error <- "E1"

NO_ALL_E2.A.2012$Error <- "E2"

NO_ALL_E3.A.2012$Error <- "E3"

NO_ALL_E4.A.2012$Error <- "E4"

NO_ALL_E5.A.2012$Error <- "E5"

Niche_ALL_A.2012 <- rbind(NO_ALL_E0.A.2012, NO_ALL_E1.A.2012,
NO_ALL_E2.A.2012, NO_ALL_E3.A.2012, NO_ALL_E4.A.2012,
NO_ALL_E5.A.2012)

# ALS2022

AUC_ALL_E0.A.2022$Error <- "E0"

AUC_ALL_E1.A.2022$Error <- "E1"

AUC_ALL_E2.A.2022$Error <- "E2"

AUC_ALL_E3.A.2022$Error <- "E3"

AUC_ALL_E4.A.2022$Error <- "E4"

AUC_ALL_E5.A.2022$Error <- "E5"

AUC_ALL_A.2022 <- rbind(AUC_ALL_E0.A.2022, AUC_ALL_E1.A.2022,
AUC_ALL_E2.A.2022, AUC_ALL_E3.A.2022, AUC_ALL_E4.A.2022,
AUC_ALL_E5.A.2022)

NO_ALL_E0.A.2022$Error <- "E0"

NO_ALL_E1.A.2022$Error <- "E1"

NO_ALL_E2.A.2022$Error <- "E2"

NO_ALL_E3.A.2022$Error <- "E3"

```

```

NO_ALL_E4.A.2022$Error <- "E4"

NO_ALL_E5.A.2022$Error <- "E5"

Niche_ALL_A.2022 <- rbind(NO_ALL_E0.A.2022, NO_ALL_E1.A.2022,
NO_ALL_E2.A.2022, NO_ALL_E3.A.2022, NO_ALL_E4.A.2022,
NO_ALL_E5.A.2022)

# GEDI

AUC_ALL_E0.G$Error <- "E0"

AUC_ALL_E1.G$Error <- "E1"

AUC_ALL_E2.G$Error <- "E2"

AUC_ALL_E3.G$Error <- "E3"

AUC_ALL_E4.G$Error <- "E4"

AUC_ALL_E5.G$Error <- "E5"

AUC_ALL_G <- rbind(AUC_ALL_E0.G, AUC_ALL_E1.G, AUC_ALL_E2.G,
AUC_ALL_E3.G, AUC_ALL_E4.G, AUC_ALL_E5.G)

NO_ALL_E0.G$Error <- "E0"

NO_ALL_E1.G$Error <- "E1"

NO_ALL_E2.G$Error <- "E2"

NO_ALL_E3.G$Error <- "E3"

NO_ALL_E4.G$Error <- "E4"

NO_ALL_E5.G$Error <- "E5"

Niche_ALL_G <- rbind(NO_ALL_E0.G, NO_ALL_E1.G, NO_ALL_E2.G,
NO_ALL_E3.G, NO_ALL_E4.G, NO_ALL_E5.G)

# pojmenovani datasetu

AUC_ALL_A.2012$CHM <- "ALS2012"

AUC_ALL_A.2022$CHM <- "ALS2022"

AUC_ALL_G$CHM <- "GEDI"

```

```

AUC_ALL <- rbind(AUC_ALL_A.2012,AUC_ALL_A.2022, AUC_ALL_G)

Niche_ALL_A.2012$CHM <- "ALS2012"

Niche_ALL_A.2022$CHM <- "ALS2022"

Niche_ALL_G$CHM <- "GEDI"

Niche_ALL <- rbind(Niche_ALL_A.2012, Niche_ALL_A.2022, Niche_ALL_G)

# boxploty

auc <- ggplot(data = AUC_ALL, aes(x=Error, y=AUC)) +
  geom_boxplot(aes(fill=CHM));auc

schoenerD <- ggplot(data = Niche_ALL, aes(x=Error, y=SchoenerD)) +
  geom_boxplot(aes(fill=CHM));schoenerD

warrenI <- ggplot(data = Niche_ALL, aes(x=Error, y=WarrenI)) +
  geom_boxplot(aes(fill=CHM));warrenI

hellingerDist <- ggplot(data = Niche_ALL, aes(x=Error,
y=HellingerDist)) +
  geom_boxplot(aes(fill=CHM));hellingerDist

```