

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ
KATEDRA APLIKOVANÉ GEOINFORMATIKY A ÚZEMNÍHO PLÁNOVÁNÍ



Fakulta životního
prostředí

Vliv rozlišení environmentálních proměnných na modely druhové distribuce

DIPLOMOVÁ PRÁCE

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

Diplomant: Matěj Hlavička

2017



Česká zemědělská univerzita v Praze
Fakulta životního prostředí

ZADÁNÍ DIPLOMOVÉ PRÁCE

Autor práce:	Matěj Hlavička
Studijní program:	Environmentální vědy
Obor:	Aplikovaná ekologie
Vedoucí práce:	Ing. Vítězslav Moudrý, Ph.D.
Garantující pracoviště:	Katedra aplikované geoinformatiky a územního plánování
Jazyk práce:	Čeština
Název práce:	Vliv rozlišení environmentálních proměnných na modely druhové distribuce
Název anglicky:	Influence of grain size on species distribution models
Cíle práce:	Cílem práce je posoudit vliv různého rozlišení environmentálních proměnných na kvalitu modelů druhové distribuce. <ul style="list-style-type: none">- Generování virtuálního druhu- Vytvoření modelů druhové distribuce s daty různého rozlišení- 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 generování virtuálních druhů a vytvořit virtuální druh, na jehož příkladu otestuje vliv různého rozlišení environmentálních proměnných na kvalitu modelu.
Doporučený rozsah práce:	40 - 50 stran
Klíčová slova:	Formulace klíčových slov je úkolem autora.

Doporučené zdroje informací:

1. Gottschalk, T. K., Aue, B., Hotes, S., & Ekschmitt, K. (2011). Influence of grain size on species–habitat models. *Ecological Modelling*, 222(18), 3403-3412.
2. LEROY, Boris, Christine N. MEYNARD, Céline BELLARD a Franck COURCHAMP. Virtualspecies, an R package to generate virtual species distributions. *Ecography*. 2015, , 001-009. DOI: 10.1111/ecog.01388. ISSN 09067590.
3. MEYNARD, Christine N. a David M. KAPLAN. The effect of a gradual response to the environment on species distribution modeling performance. *Ecography*. 2012, 35(6), 499-509. DOI: 10.1111/j.1600-0587.2011.07157.x. ISSN 09067590
4. MILLER, Jennifer. Species Distribution Modeling. *Geography Compass*. 2010, 4(6), 490-509. DOI: 10.1111/j.1749-8198.2010.00351.x. ISSN 17498198
5. MOUDRÝ, Vítězslav. Modelling species distributions with simulated virtual species. *Journal of Biogeography*. 2015,42(8). DOI: 10.1111/jbi.12552. ISBN 10.1111/jbi.12552.

Předběžný termín 2017/18 ZS - FŽP
obhajoby:

Elektronicky schváleno: 8. 11. 2017
doc. Ing. Petra Šimová, Ph.D.
Vedoucí katedry

Elektronicky schváleno: 8. 11. 2017
prof. RNDr. Vladimír Bejček, CSc.
Děkan

PROHLÁŠENÍ

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně, pod vedením Ing. Vítězslava Moudrého, Ph. D. a uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal. Prohlašuji, že tištěná verze se shoduje s verzí odevzdanou přes Univerzitní informační systém.

V Praze dne 15. 12. 2017

.....

PODĚKOVÁNÍ

Děkuji svému školiteli Vítězslavu Moudrému za řadu podnětných konzultací a cenné odborné rady při zpracovávání celé práce. Taktéž mu děkuji za poskytnutí LiDARové topografické vrstvy Krkonošského národního parku.

Abstrakt

Druhové distribuční modely (SDM) se staly mocným nástrojem v mnoha oborech díky své schopnosti nalézt vazbu mezi organismem a jeho prostředím. Tyto modely jsou schopny predikovat rozšíření druhů na dosud neprozkoumaných místech či v kontextu klimatických změn, čímž nalezly uplatnění například v ochraně přírody, populační ekologii, invazní biologii nebo bioklimatologii. Mnoho studií se již zabývalo vlivem kvality vstupních dat na výkon těchto modelů, avšak vliv rozlišení environmentálních prediktorů stále není podrobně prozkoumán. Vysoce kvalitní environmentální vrstvy zobrazují více detailů a tím lépe reprezentují heterogenní krajinu, což by mělo pomoci SDM lépe definovat vazbu mezi druhem a prostředím. Několik studií měnilo velikost zrna (a tím i rozlišení prediktorů), ale zatím žádná z nich nezkoumala efekt postupného snižování rozlišení prediktorů s ohledem na rozlišení dat o výskytu druhů. V této práci jsem vytvořil 3 různé modely druhové distribuce (250 m, 500 m, 1 km) založené na několika prediktorech (svažitost, orientace svahu, průměrná teplota v červnu) v různém rozlišení (5 m, 10 m, 20 m, 50 m, 100 m) a na prezenčně-absenčních datech virtuálního druhu, abych předešel možnému šumu vzniklého z nespolehlivých druhových dat. Zájmovou oblastí byl Krkonošský národní park, jenž je heterogenní z hlediska svažitosti i orientace. Navzdory očekávání, samotné rozlišení prediktorů nehrálo téměř žádnou roli ve výkonu SDM a pouze 250m model profitoval ze zlepšování kvality prediktorů, avšak jen v minimální míře. Na druhou stranu výkon SDM ovlivňovalo rozlišení dat o výskytu druhů, a to při všech rozlišeních prediktorů. Nicméně i tak byl rozdíl mezi hodnotami AUC mezi 250m a 1km modelem malý (průměrně 0,07). Tyto výsledky korespondují s většinou dosud provedených studií, které také pozorovaly pouze malý pokles v kvalitě modelů. Vliv kvality prediktorů na výkon SDM by si však rozhodně zasloužil další výzkum na širší paletě rozlišení dat o výskytech druhů. Nalezení hranice, kde začne rozlišení prediktorů znatelně ovlivňovat výkon SDM, by nám pomohlo v rozhodování, zda se vyplatí investovat čas i peníze na získání nejkvalitnějších vrstev prediktorů a při jaké kvalitě dat o výskytech druhů by jejich rozlišení mělo vůbec začít hrát výraznou roli.

Klíčová slova: druhové distribuční modely, virtuální druh, rozlišení, GLM, DMT

Abstract

Species distribution models (SDM) have become a powerful tool in many disciplines for its ability to find a link between species and their environment. As being able to predict species distributions on yet unsampled sites or with respect to climate change, SDMs have found its place in conservation ecology, population ecology, invasion biology, bioclimatology and many other disciplines. A great effort have been put into finding out how the quality of input data affects SDM performance, but the effects of the resolution of environmental are still not well researched. High quality environmental layers can carry more details and are better at representing heterogenous landscape, which should help the SDMs to establish strong species-environment bond. Other works have been manipulating with different grain sizes (and thus changing the quality of environmental predictors), but no one yet discovered the effect of stepwise coarsening the resolution of predictors with respect to different grain sizes of species data. In this paper I created 3 models in different grain sizes (250 m, 500 m, 1 km) based on a set of environmental predictors (aspect, slope, mean temperature in June) in multiple scales (5 m, 10 m, 20 m, 50 m, 100 m) and presence-absence species data from virtual distribution to avoid the possible confounding effects of unreliable data sources. The study area was Krkonoše National Park, which is greatly heterogenous concerning both aspect and slope. In contrary to my expectations, the resolution of predictors *per se* played almost no role in SDM performance and only 250m model benefited from better quality predictors, but in minor magnitude. On the other hand, the grain size of species data did play a role in SDM performance and it was observed on all predictor scales, although the magnitude of difference in AUC values between 1km and 250m models wasn't so high (about 0,07). This corresponds with most of related papers that experienced similar decrease of performance. The effects of high quality predictors would definitely deserve further research on wider range of model grain sizes. Finding a threshold where the scale of predictors begins to significantly affect SDM performance would help to define if is worth to invest financial and temporal resources for obtaining best quality environmental layers and which grain sizes would benefit the best from these layers.

Keywords: species distribution models, virtual species, resolution, GLM, DEM

Obsah

1	Úvod.....	9
2	Cíle práce	10
3	Literární rešerše	11
3.1	Druhovité distribuční modely a jejich význam	11
3.2	Proces tvorby SDM.....	12
3.3	Vstupní data pro SDM	13
3.3.1	Data o výskytech druhů	13
3.3.2	Data o prostředí.....	15
3.4	Evaluace modelů	16
3.5	Problematika rozlišení vstupních dat	18
3.6	Úskalí spojená s SDM	19
3.6.1	Překážky pramenící z kvality dat o výskytech	20
3.7	Virtuální druhy a jejich přínos pro SDM.....	21
4	Metodika.....	23
4.1	Metodologie postupu	24
4.2	Vymezení a charakteristika zájmového území	24
4.3	Zdroje vstupních dat	25
4.3.1	WorldClim.....	25
4.3.2	DEM LiDAR	26
4.4	Úprava vstupních dat.....	26
4.4.1	Průměrná teplota	26
4.4.2	Svažítost.....	26
4.4.3	Orientace svahu	27
4.5	Virtuální druh	29
4.6	Příprava dat pro modelování druhové distribuce.....	31
4.6.1	Metoda průměrných hodnot podmínek prostředí.....	32
4.6.2	Metoda relativního zastoupení vyhovujících podmínek	32
4.7	Modelování druhové distribuce a evaluace	33
5	Výsledky	35
5.1	Metoda průměrných hodnot podmínek prostředí.....	35
5.2	Metoda relativního zastoupení vyhovujících podmínek	36
6	Diskuze.....	38
7	Závěr	41
8	Přehled literatury a použitých zdrojů	42

1 Úvod

Druhové distribuční modely mají v ekologii již dlouhou tradici a v posledních letech stále nabývají na popularitě v důsledku potřeby reagovat na probíhající klimatické změny, pokles biodiverzity a biologické invaze (Yackulic et al., 2013). S těmito změnami roste i tlak na co nejlepší pochopení procesů ovlivňující rozšíření organismů na Zemi a modely druhové distribuce jsou v tomto směru vítaným pomocníkem. Jejich rozmach byl ještě více umocněn příchodem technologií GIS a dálkového průzkumu Země, které s sebou přinesly lepší dostupnost digitálních dat a širokou škálu prostorových analýz (Miller, 2010). Druhové distribuční modely tak mají k dispozici velký výběr environmentálních vrstev ve stále se zvyšující kvalitě. Vyšší rozlišení přináší výhody v detailnější reprezentaci prostředí, kde lze zachytit i nenápadné změny (například v typu povrchu), které by byly při nižším rozlišení skryty (Gottschalk et al., 2011). A protože od druhových distribučních modelů očekáváme co nejlepší popis vazby mezi organismem a jeho prostředím, zdá se být používání co nejkvalitnějších environmentálních vrstev jasnou volbou. Nicméně studie zabývající se tímto tématem nepřinášejí tak jednoznačné výsledky. Navíc do celého procesu modelování zasahuje mnoho proměnných a zdroje potenciálních chyb pramenících z použitých vstupních dat mohou srovnávání mezi jednotlivými studiemi komplikovat.

Možným řešením je použití virtuálního druhu, u kterého jsou zvoleny pouze relevantní environmentální proměnné a kde je i nulové riziko, že budou vstupní data zatížena chybou při určování druhu nebo v metodice terénního sběru. Díky nim se lze ve srovnávacích studiích zaměřit čistě na technickou stránku věci. Ve své práci využiji právě koncept virtuálního druhu a podívám se, jakým způsobem může rozlišení environmentálních proměnných a dat o výskytu druhu (rozlišení modelu) ovlivnit predikční schopnost druhových distribučních modelů.

2 Cíle práce

Cílem této práce je zjistit, jakým způsobem může rozlišení environmentálních proměnných vstupujících do modelů druhové distribuce v různé kvalitě ovlivnit jejich predikční schopnost. Hypotéza je taková, že postupným snižováním kvality vstupních dat bude docházet k horšímu výkonu u všech modelů.

V této práci jsem stanovil dílčí cíle:

- 1) Vygenerování virtuálního druhu nad vybranou sadou environmentálních proměnných.
- 2) Vytvoření modelů druhové distribuce na základě dat o výskytu druhů v různé kvalitě pomocí environmentálních vrstev v několika rozlišeních.
- 3) Vyhodnocení predikční schopnosti modelů z hlediska predikované prezence či absence.

3 Literární rešerše

3.1 Druhové distribuční modely a jejich význam

Druhové distribuční modely (též známé pod zkratkou SDM - Species Distribution Models), lze obecně definovat jako modely kombinující informace o prostorovém rozšíření druhů spolu s podmínkami prostředí, v nichž se druh nalézá (například klimatické či půdní parametry). Model na základě těchto vstupních dat stanovuje vhodné podmínky k udržení životaschopné populace (Pearson et al., 2010; Araújo et Peterson, 2012). SDM jsou využívány zejména z těchto důvodů: i) snaha o pochopení vztahu mezi organismem a prostředím, ii) zjištění, které oblasti jsou více nebo méně vhodné pro život zájmového druhu, iii) interpolace druhové distribuce v oblastech, kde nebyl proveden sběr dat o výskytech, iv) extrapolace druhové distribuce do oblastí mimo jejich přirozený areál (Elith et Graham, 2009). Ke všem zmíněným bodům mohou SDM přinést cenné informace, a proto se ve vědeckých kruzích těší velké oblibě napříč obory. Zejména v ochraně přírody, pro kterou je predikce druhové distribuce klíčový element (Hirzel et al., 2001), nám SDM mohou pomoci odhadnout velikost populace ohroženého druhu (Gottschalk et al., 2011) nebo naleznout i doposud neznámé populace a získat tak podklady pro management nutný k ochraně biotopu či pro sestavení reintrodukčního programu (Graham et al., 2004).

Kromě ochrany přírody zažívají velký rozmach v oblastech, jako jsou biogeografie, makroekologie, ekologie invazních druhů, evoluční biologie a zkoumání vlivu klimatických změn na šíření organismů (včetně patogenů a jejich vektorů) (Guisan et Thuiller, 2005; Pearson, 2010; Meynard et Kaplan, 2013; Miller, 2014). Například v invazní biologii lze s využitím záznamů o rozšíření invazního druhu v domovském areálu modelovat jeho budoucí rozšíření na místech, kam byl introdukován (Graham et al., 2004; Guillera-Aroita et al., 2015). Bioklimatologové zase mohou na základě aktuálního rozšíření druhů simulovat pomocí SDM mapy budoucích výskytů v kontextu klimatických změn (Duputié et al., 2014), nicméně stejně jako u všech extrapolujících modelů je jejich hodnocení ošemetné, protože následky změn klimatu, jež jsou modelovány, buď ještě nenastaly vůbec, nebo jsou probíhající změny zatím nedostatečně zdokumentované. Stejně tak při snaze o predikci distribuce invazních druhů záběr modelu přesahuje region, odkud pochází vzorek vstupních dat, což celou situaci komplikuje (Araújo et al., 2005). Přestože nám SDM nemusí poskytnout vždy přesné a konkrétní odpovědi na položené otázky, mohou

pomocí alespoň určit směr dalšího výzkumu, odhalit potenciální rizika lidského počínání nebo zefektivnit vynakládání finančních prostředků v daném oboru.

3.2 Proces tvorby SDM

Do druhových distribučních modelů vstupují data o výskytu zkoumaného organismu (např. ve formě bodů či čtverců síťového mapování) a relevantní environmentální prediktory. Pomocí statistických algoritmů mohou tyto korelativní modely odhalit ekologické nároky druhu a posléze predikovat jeho rozšíření (Elith et al., 2006; Pearson, 2010; Guillera-Aroita et al., 2015). Modelovacích algoritmů existuje celá řada a dělí se obecně do dvou skupin podle typu vstupních dat o výskytech. První skupina modelů pracuje s ryze prezenčními daty (např. BIOCLIM, DOMAIN, LIVES), zatímco zbylé modely vyžadují, kromě prezenčních dat, i informace o absencích (resp. pseudo-absencích) (např. GLM, GAM, MARS), popřípadě data o výskytech ostatních druhů (např. GDM) (Elith et al., 2006). Výběr vhodného algoritmu by měl reflektovat cíle studie a povahu vstupních dat (velikost vzorku, rozmístění prezenčních dat v prostoru). Nelze tedy jednoznačně určit, který z algoritmů je nejlepší (Aguirre-Gutiérrez et al., 2013). Kromě volby algoritmu je důležitý i výběr měřítka, ve kterém budeme celý proces modelování provádět. Zde musíme přihlídnout nejen k cílům studie a kvalitě dostupných vstupních dat (Elith et al., 2009), ale i k vlastnostem zkoumaného druhu, u kterých nás zajímá hlavně velikost areálu a způsob využívání přírodních zdrojů (odlišné pro sesilní a mobilní organismy) (Guisan et Thuiller, 2005). Podle Guisan et Thuiller (2005) lze celý proces tvorby SDM shrnout do šesti dílčích kroků:

- 1) Konceptualizace, formulace hypotéz.
- 2) Příprava vstupních dat. Tento krok zahrnuje získání dat o výskytech (ať už vlastním sběrem či z jiného zdroje) a vrstev s relevantními environmentálními prediktory. Dále v tomto kroku podle zvoleného měřítka upravíme všechna vstupní data (typicky pomocí GIS softwaru).
- 3) Fitování (kalibrace) modelu. Zvolený model na základě vstupních dat vypočítá afinitu druhu ke zvoleným prediktorům. Výstupem jsou zde hodnoty vhodnosti prostředí, které každý kvadrát v modelovaném území definují v intervalu od 0 do 1, kde 0 značí prostředí naprosto nevhodné pro život a 1 ideální prostředí.
- 4) Evaluace modelu. Pomocí evaluačních metrik je hodnocena predikční schopnost modelu.

5) Predikce druhového rozšíření. Hodnoty vhodnosti prostředí vystupující z modelu můžeme vizualizovat pomocí GIS softwaru na mapě a interpretovat výsledky. Mnoho studií však u těchto kontinuálních hodnot nezůstává a dále je převádí na binární rastry vyjadřující predikovanou prezenci či absenci. K tomuto je nezbytné reklasifikovat kontinuální rastr podle uměle zvolené meze (např. všechny hodnoty vhodnosti prostředí nad 0,6 budou považovány za prezenci, zbytek jako absence). Podle Guillera-Aroita et al. (2015) se však bohužel tímto krokem ztrácí mnoho cenných informací, které model poskytuje, nicméně reklasifikace je velmi situační a pro některé studie může být užitečná více (vymezení zón ochrany), pro jiné méně (predikce rozšíření invazních druhů).

6) Aplikace získaných poznatků.

3.3 Vstupní data pro SDM

3.3.1 Data o výskytech druhů

Při modelování druhové distribuce jsou tato data asi nejcitlivějším článkem celého řetězu, jelikož na jejich kvalitě stojí a padá kalibrace SDM. A získat kvalitní druhová data není triviální. Podle typu mohou do SDM vstupovat: a) pouze prezenční data (sada obsahuje jen záznamy o výskytech), b) prezenčně-absenční data (model zde porovnává podmínky prostředí na místech, kde se druh vyskytuje, s místy, kde se nevyskytuje), c) tzv. *presence-background* data. U této varianty model používá namísto absenčních dat tzv. pseudo-absenční data (v zájmové lokalitě je vybrána skupina bodů (např. v určité minimální vzdálenosti od bodů výskytu) sloužící jako absenční záznamy) (Pearson, 2010; Barbet-Massin et al., 2012; Guillera-Aroita et al., 2015). Ve srovnávacích studiích týkajících se SDM jsou však využívány téměř výhradně poslední dva zmíněné typy.

Druhová data lze získat buď vlastním terénním sběrem anebo od třetích stran, kterými jsou nejčastěji muzejní zdroje, univerzitní databáze, herbářové sbírky nebo amatérská činnost (Varela et al., 2014). Právě tyto velké nálezové databáze jsou typickým příkladem ryze prezenčních dat (Miller, 2010; Yackulic et al., 2013). Jejich výhodou je dostupnost a široký záběr napříč taxony i regiony, avšak dokud si k nim neopatříme i absenční či pseudo-absenční záznamy, nelze je použít u naprosté většiny modelovacích algoritmů (Miller, 2010). Některé nálezové databáze sice mohou svá data označovat jako prezenčně-absenční, avšak pokud si nejsme naprosto jisti jejich spolehlivostí, je lepší je považovat za čistě prezenční (Graham et al., 2004).

Vliv jednotlivých typů druhových dat na vypovídací schopnost SDM zkoumalo již několik studií a vyšlo najevo, že každá ze skupin má svá specifika. Ryze prezenční data například vůbec neberou v potaz podmínky panující ve zbytku zájmové lokality (Guillera-Arroita et al., 2015), což lze do jisté míry vyřešit dodáním pseudo-absenčních záznamů. Zásadnějším nedostatkem je však ignorance druhové prevalence (frekvence výskytu) (Elith et al., 2006), s čímž souvisí fakt, že nedovedeme odhalit případy, kdy je zkoumaný druh vzácný (a dobře zdokumentovaný) nebo běžně se vyskytující (a zdokumentovaný špatně). Tyto dva jevy umí odlišit pouze prezenčně-absenční data (Ferrier et al., 2002, Guillera-Arroita et al., 2015). Ryze prezenční data by také měla být podle Yackulic et al. (2013) sesbírána ideálně pomocí náhodného či systematického samplingu, abychom předešli špatné kalibraci modelu způsobené nedostatečným pokrytím zájmové lokality.

Elith et al. (2006) ve své srovnávací studii zjistili, že modely pracující s prezenčně-absenčními daty měly lepší predikční schopnost než modely postavené jen na prezenčních datech. I Elith et Graham (2009) spojují použití prezenčně-absenčních dat s lepší kalibrací modelu. Bohužel se však s nimi pojí několik nevýhod. Pokud se absenční záznamy nachází na místech, které jsou pro druh jinak vhodné, ale neobývá je v důsledku mezidruhových interakcí, migračních bariér či disturbancí, správná kalibrace modelu bude problematická (Miller, 2010). Ještě černější scénář nastává v případě tzv. falešných absencí (druh na místě nebyl zaznamenán, ale ve skutečnosti se tam nachází), které přímo poškozují kalibraci modelu (Pearson, 2010). Modely postavené na těchto chybných absenčních záznamech pak mají podobnou (ne-li horší) vypovídací schopnost jako při použití náhodně vygenerovaných pseudo-absenčních dat. Absenční data je také velice těžké získat, a to i u snadno rozpoznatelných sesilných druhů (například stromy) (Duputié et al., 2014).

Budeme-li se raději spoléhat na vlastní sběr druhových dat, správná metodika by měla pokrýt celý areál výskytu zkoumaného druhu i všechny environmentální gradienty, kde se nachází (Jarnevich et al., 2015), čehož můžeme dosáhnout náhodným, stratifikovaným (sběr dat napříč gradientem klíčových prediktorů) či systematickým sběrem (Guisan et Thuiller, 2005). Další vliv na výkon modelu má i velikost vzorku. Čím větší máme vzorek druhových dat, tím lepší predikční schopnost modelu můžeme očekávat (Wisz et al., 2008; Mitchell et al., 2017). Odlišně se však tento vliv projevuje u modelování generalistů (kde je potřeba větší

vzorek dat) a specialistů (méně bodů stačí na pochopení vztahu druh-prostředí) (van Proosdij et al., 2015).

Je zřejmé, že každý typ druhových dat má svá pro a proti. V drtivé většině případů jsou k dispozici pouze prezenční data (Hernandez et al., 2006; Pearson, 2010; Guillera-Arroita et al., 2015) a odkážeme-li se na nálezové databáze, je třeba mít na paměti, že se mezi sebou budou lišit designem sběru (Varela et al., 2014). Výběr vhodného zdroje tedy bude záviset především na cíli studie a množství prostředků, které můžeme obětovat (jak finanční, tak časové).

3.3.2 Data o prostředí

Neméně důležitým článkem při modelování druhové distribuce jsou data o prostředí (prediktory). Nové metody a technologie dálkového průzkumu Země s sebou přináší možnost použití robustních a detailních digitálních modelů terénu (Elith et Leathwick, 2009) a k tomu navíc umožňují snímat povrch i v jiném než viditelném spektru, což rozšiřuje škálu použitelných environmentálních proměnných (např. index NDVI). Díky prostorovým analýzám v GIS softwaru můžeme z primárních dat získat i další odvozené proměnné, jako například vzdálenost od měst či silnic (Miller, 2010). Nejčastěji se ale používají vrstvy související s klimatem (teplota, množství srážek), topografií (nadmořská výška, svažitost), půdním typem a *landcover* (Miller, 2010; Pearson, 2010). U klimatických dat je však potřeba zachovat jistou opatrnost, jelikož vznikají interpolací výstupů z meteorologických stanic, které nejsou rovnoměrně rozmístěné a v některých oblastech je jejich počet příliš nízký na to, abychom mohli získat spolehlivé klimatické vrstvy ve vysokém rozlišení (Parra et al., 2004). Zahrnout tedy můžeme jak kontinuální, tak kategorické proměnné (ty jsou často před použitím v SDM dále upravovány, např. výpočtem relativního zastoupení dané proměnné v mapovacím kvadrátu). Podmínkou úspěšného modelování je ale z nepřeberného množství možností vybrat prediktory, které jsou ekologicky relevantní ke zkoumanému druhu (Parra et al., 2004). Při jejich výběru je nutné dávat bedlivý pozor na multikolinearitu (některé prediktory jsou mezi sebou korelované) (Guisan et Thuiller, 2005), což je zásadní zejména u snah o extrapolaci druhové distribuce mimo přirozený areál (např. modelování invazních druhů) (Jarnevich et al., 2015).

Doposud byla řeč jen o abiotických podmínkách. Biotické interakce však mají rovnocenný vliv na rozšíření druhů, stejně jako teplota, orientace svahu, apod.. Někteří autoři se snažili tento fakt reflektovat pomocí SDM tím, že mezi abiotické proměnné přidali i data o rozšíření ostatních druhů (nebo jen kompetitorů) a chtěli

se tak přiblížit modelování realizované niky. Nicméně Araújo et Guisan (2006) připomínají, že mezidruhové interakce často probíhají v menším měřítku, než je rozlišení samotného modelu. Vyhne-li se druh svému kompetitorovi, mohou spolu koexistovat i v malé oblasti, tudíž pokud v mapovacím kvadrátu podmíníme výskyt zájmového druhu absencí jeho kompetitora, dopustíme se chyby. Jiný případ však nastává, pokud je zkoumaný druh úzce vázán na jiný (lovec-kořist, mutualismus, symbióza). Zde nám výskyt takových druhů opravdu může pomoci s predikcí, a to i v případě, že jsou ostatní abiotické prediktory homogenně zastoupené (čili méně vypovídající) (Gutiérrez et al., 2005). V určitých případech tedy lze s biotickými interakcemi počítat, ale jejich použití je bohužel i tak ošemetné kvůli jejich dynamičnosti (intenzita se mění nejen v čase, ale i mezi populacemi jednoho druhu) (Miller et Holloway, 2015).

3.4 Evaluace modelů

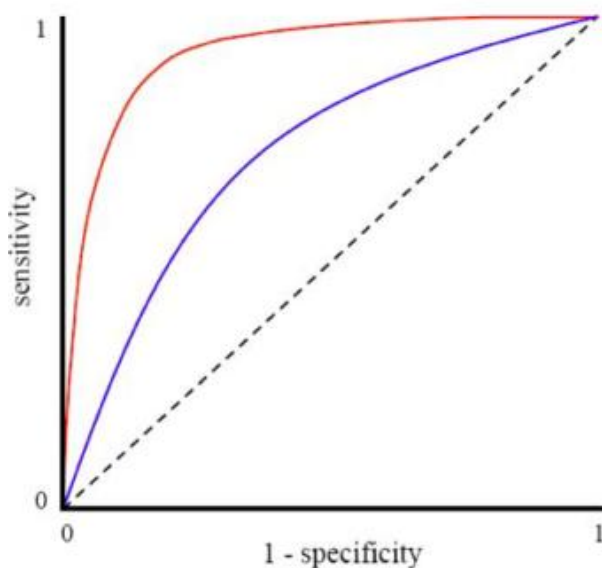
Abychom mohli zjistit predikční schopnost modelu, je nutné jeho výstupy evaluovat. Možnosti jsou dvě. V první řadě můžeme konfrontovat predikovanou distribuci s daty pocházejícími z nezávislého sběru. Druhou možností je použít vstupní druhová data. K tomu, aby se u druhé možnosti výkon modelu netestoval na identické sadě dat, na kterých byl kalibrován, se podle Araújo et al. (2005) využívá několik technik. 1) náhodné rozdělení dat na kalibrační vzorek (např. 70 %) a testovací vzorek (30 %), 2) křížová validace (data se nejprve rozdělí na několik skupin a poté je v každé iteraci postupně každá z nich použita jako testovací a zbylé jako kalibrační), 3) bootstrapping (náhodný výběr ze souboru s opakováním), 4) jackknifing (podsoubory jsou vytvářeny postupným vypouštěním jednotlivých pozorování). Všechny tyto metody mají společné to, že pracují s podmnožinou stávajících dat. Iterační metody (u technik 2, 3 a 4) následně sestaví několik modelů, jejichž výstupy se zprůměrují (Pearson, 2010).

Samotné evaluační metriky lze rozdělit na tzv. *threshold dependent* a *threshold independent*. *Threshold dependent* metody nejdříve převádí sadu kontinuálních hodnot (vyjadřující vhodnost prostředí od 0 do 1; jsou výstupem každého modelu) na binární hodnoty presence/absence (Miller, 2010). Ke stanovení meze, nad kterou se druh vyskytuje a pod kterou nikoliv, nabízí evaluační algoritmus řadu technik (např. snaha o maximalizaci senzitivity či specificity, maximalizace PCC, atd.). Po stanovení meze lze sestavit tzv. chybovou matici (kontingenční tabulka), která porovnává vztah mezi vstupními daty a výstupy modelu a udává nám počet nálezů:

a) skutečně pozitivních, b) falešně pozitivních, c) falešně negativních, d) skutečně negativních. Podmínkou jejího sestavení je však použití absenčních či pseudo-absenčních dat v modelu (Pearson, 2010). Z chybové matice můžeme následně získat další ukazatele. Senzitivita/specificita vyjadřují relativní počet správně předpovězených prezencí/absencí, PCC (Percentage Correctly Classified) udává procento správně predikovaných lokalit.

Threshold independent metriky jsou odvozeny z tzv. ROC křivky (Receiver Operating Characteristic Curve), jež měří predikční schopnost modelu napříč všemi možnými mezemi (výše zmíněné techniky stanovily pouze jedinou mez). Na osách ROC křivky je vynesena senzitivita a 1-specificita (obrázek č. 1) a vyjadřuje vztah mezi podílem správně předpovězených prezencí (senzitivita) a podílem špatně předpovězených absencí (1-specificita). Z této křivky je poté odvozena hodnota AUC, která reprezentuje plochu pod ROC křivkou (Area Under the ROC Curve). AUC měří schopnost modelu rozlišovat mezi pozorovanými prezencemi a absencemi, je-li každý záznam vybrán náhodně (Elith et Graham, 2009; Pearson, 2010) a nabývá hodnot od 0 do 1, kde 1 značí perfektní model (na 100 % spolehlivé odlišení míst, kde je druh přítomen, od míst, kde přítomen není) a 0,5 znamená spolehlivost podobnou náhodě (interpretace hodnot $< 0,5$ již nemá smysl) (Elith et al., 2006). Největší výhodou AUC je to, že hodnotí model z hlediska prezencí a absencí současně, zatímco senzitivita/specificita pouze ukazují, jak si model vedl v jedné kategorii (prezence/absence) a nezohledňuje, kolik chyb udělal ve druhé (McPherson et al., 2004).

Obrázek č. 1 - Ilustrace ROC křivek. Červená křivka vyjadřuje perfektní predikční schopnost modelu, fialová reprezentuje průměrný model. Přerušovaná čára značí predikční schopnost rovnou náhodě (Pearson, 2010).



3.5 Problematika rozlišení vstupních dat

Volba rozlišení (velikost zrna) je další z důležitých faktorů, který může ovlivnit predikční schopnost modelu (Guisan et al., 2007). V případě environmentálních proměnných se se zhoršováním rozlišení postupně smazávají rozdíly mezi jedinečnými rysy povrchu (čím je krajina více heterogenní, tím více informací ztrácíme). U druhů, které mají specifické nároky na habitat, se při modelování v nízkém rozlišení vytrácí zřetelná vazba druh-prostředí nutná pro správnou kalibraci modelu (Gottschalk et al., 2011), což má ještě výraznější dopad při modelování druhové distribuce organismů obývajících mikrohabitaty (Pradervand et al., 2013). S daty o výskytech, pocházejících z nálezových databází, se zase pojí problém, že rozlišení zvolených environmentálních proměnných je často mnohem vyšší (Guisan et al., 2007) (např. výskyty jsou mapovány v rozlišení 1 km, zatímco vrstvy prediktorů mají rozlišení 5 m). Environmentální vrstvy je pak nutné zhoršit do rozlišení dat o výskytech, aby s nimi mohl model pracovat. Existují však i případy, kdy se naopak data o výskytech v horším rozlišení převádí do rozlišení vyššího, aby korespondovala s kvalitními environmentálními daty. Tento proces se nazývá *downscaling* (např. Bombi et D'Amen, 2012).

Podle Pradervand et al. (2013) není vliv rozlišení environmentálních proměnných dopodrobna prozkoumán kvůli omezené dostupnosti a ceně velmi kvalitních dat, což se postupně v posledních letech mění k lepšímu. Náskok (co se týče rozlišení) nad druhovými daty se tak bude ještě více prohlubovat. Honba za co nejvyšším rozlišením však také není samospásná. I u velmi kvalitních druhových dat se může objevit prostorová nepřesnost (chybné určení GPS souřadnic) a modelování ve vysokém rozlišení může v důsledku těchto chyb způsobit ještě více škody než při nižším rozlišení. I malý odklon prezenčního záznamu od reálné pozice znamená změnu hodnot environmentálních proměnných a následný výpočet chybného vztahu mezi druhem a prostředím (Guisan et al., 2007; van Gils et al., 2012), což bylo prokázáno v případech, kdy se modeluje s malým počtem záznamů (Mitchell et al., 2017). Další problém je spojen s měřítkem, které jsme si pro modelování zvolili. Organismy se bohužel odmítají řídit jednoduchými zákonitostmi a na podmínky prostředí reagují jinak na velké a malé škále (van Gils et al., 2012) (při modelování na úrovni kontinentů mohou nejvíce reagovat na změny průměrné teploty, zatímco na malém měřítku může být hlavním hybatelem jejich přítomnosti např. zastoupení listnatých stromů).

Opět zde platí tvrzení, že rozlišení, ve kterém se chystáme modelovat, by kromě cílů studie mělo zohledňovat i atributy zkoumaného druhu, jeho způsob využívání

přírodních zdrojů a hlavně povahu a druh nepřesností, které s sebou nesou vstupní data jak ve vysokém, tak i v nízkém rozlišení.

3.6 Úskalí spojená s SDM

Druhové distribuční modely mají nepochybně potenciál stát se mocným pomocníkem ve všech oborech, které je už nyní využívají. Bohužel při snahách o co nejlepší pochopení vztahů mezi organismy a jejich prostředím narážíme na velké množství překážek, z nichž jsem některé zmínil v kapitolách o vstupních datech. Obecné problémy přináší už jen chování zkoumaných organismů. Zatímco u sesilných druhů je relativně snadné popsat prostředí, ve kterém žijí, mobilní druhy mohou využívat zdroje v celém svém areálu, což snahu o kvantifikaci vazby mezi druhem a prostředím značně komplikuje (Elith et Leathwick, 2009). Naší reakcí může být zvětšení mapovacích kvadrátů tak, abychom pokryli akční radius daného živočicha, avšak s tím je zase spojena patřičná úprava vrstev prediktorů do nižšího rozlišení a riziko ztráty informací o prostředí. Dalším problémem jsou odlišné ekologické strategie organismů. Specialisté, kteří mají úzkou ekologickou niku, se modelují lépe než generalisté, jelikož snáze podchytíme jejich vazbu k prostředí (Hernandez et al., 2006). U generalistů je tato vazba méně zřetelná a navíc může být jejich široké rozšíření dáno lokálními adaptacemi či fenotypovou plasticitou a vztah k prostředí tak bude pro SDM neuchopitelný. Toto hrozí zejména u invazních druhů. Vedle všeho je nutné dodat, že SDM nepočítají s populační dynamikou a mezidruhovými interakcemi (až na výjimky). Jejich výstupem jsou tedy data, která vyjadřují spíše fundamentální niku a která mohou být co do počtu predikovaných prezencí optimističtější než realita (Parra et al., 2004). K tomuto je nutné přihlídnout zejména při provádění reintrodukčních programů.

Přesunu-li se od druhových aspektů k environmentálním datům, špatná predikční schopnost modelu může být způsobena také volbou nevhodných prediktorů. Pokud nejsou naše znalosti o zkoumaném druhu dostatečné a nezohledníme proměnnou, která je pro něj klíčová, výstupy modelu nebudou uspokojivé (Hernandez et al., 2006). Můžeme sice do vstupních dat zahrnout desítky vrstev v prvotřídní kvalitě, avšak reaguje-li druh nejvíce na úplně jinou proměnnou, jeho vazbu k prostředí přesto neobjevíme.

3.6.1 Překážky pramenící z kvality dat o výskytech

Jakožto základní kámen úspěšného modelování si i zde data o výskytech zaslouží vlastní podkapitolu. Nejprve bych chtěl zmínit chybovost způsobenou zkresleným sběrem dat, tzv. *sampling bias*. Jedná se o jev, kdy jsou při terénním sběru dat preferovány dobře přístupné (poblíž lidských sídel a silnic) nebo zajímavé lokality (zvláště chráněná území). Nashromážděná data tak nejsou náhodná, systematická ani stratifikovaná a při modelování hrozí nedostatečné pochopení vztahu mezi druhem a prostředím. Podle Guillera-Aroita et al. (2015) však není jádro problému v neukázněnosti při sběru dat, nýbrž hlavně v rozložení environmentálních proměnných. Budeme-li například provádět terénní sběr převážně v okolí vesnic či měst, kde je menší zastoupení lesů, model může dojít k chybnému závěru, že druh preferuje nezalesněná stanoviště. Modely založené pouze na prezenčních datech by tímto utrpěly téměř smrtelnou ránu. Naštěstí při použití pseudo-absenčních či absenčních dat budou následky tohoto počínání méně fatální (absenční záznamy na nezalesněných místech mimo lukrativní lokality pomohou tuto nesprávně vypočítanou afinitu částečně vyrovnat). V reakci na zkreslená či nekompletní druhová data Araújo et Guisan (2006) navrhuje provádět tzv. *sub-sampling* na již prozkoumaných lokalitách. To znamená získat na inkriminovaných místech nová data pomocí náhodného či stratifikovaného sběru. Další možnou variantou je podle Varela et al. (2014) použití environmentálního filtrování, při kterém se odebírají repetitivní záznamy nacházející se v podobných podmínkách (s hlediska zvoleného prediktoru).

Jiným kamenem úrazu je validita druhových dat, zejména pokud čerpáme z nálezových databází, u nichž je systém a metodika sběru nejasná. Zmapované výskyty mohou být již neaktuální anebo důsledkem tzv. *source-sink* dynamiky. Druh tak mohl být pozorován na pro něj nevhodném stanovišti a jen díky nedaleké zdrojové populaci jsou jeho počty udržovány neustálým přísunem migrujících jedinců. Absence také mohou ve skutečnosti znamenat nesprávné určení druhu (Pearson, 2010), nedostatečný terénní průzkum nebo prostý fakt, že druh unikl pozornosti výzkumníků (Graham et al., 2004; Elith et al., 2006; Jarnevich et al., 2015). Taktéž prezenční záznamy jsou podmíněny tím, zda druh na lokalitě skutečně žije a pravděpodobností, že je lokalita zahrnuta do terénního sběru (Yakulic et al., 2013). Toto bychom měli mít na paměti hlavně při používání pseudo-absenčních dat (což je případ drtivé většiny studií zabývajících se SDM), aby nenastala situace, při které vygenerujeme pseudo-absence na místa, kde vůbec nebyl proveden terénní sběr a druh tam mezitím může klidně prosperovat.

Nepřesná druhová data také přináší překážku pro evaluaci. Tím, že skutečné rozšíření druhu není přesně známé, se použitím jakékoliv evaluační techniky pracující s podmnožinou vstupních dat testuje výkon modelu opět na těch samých nepřesných datech. Řešením by bylo použití dat z nezávislého sběru (Elith et al., 2006), což je však časově i finančně náročné a také zřídka využívané (Araújo et al., 2005).

V neposlední řadě je třeba zmínit i rozlišení druhových dat. Abychom mohli využívat SDM pro sestavení plánů péče konkrétního ohroženého druhu v určité oblasti, je třeba mít k dispozici data o výskytech ve vysokém rozlišení. Bohužel běžné rozlišení v nálezových databázích je zhruba 1 km, což pro podchycení vztahu mezi druhem a prostředím není dostatečné (van Gils et al., 2012; Miller, 2014). Zatímco díky rozmachu nových technologií dálkového průzkumu získáváme postupně přístup k velmi kvalitním environmentálním vrstvám, data o výskytech, jakožto hlavní pilíř úspěšného modelování, stále zaostávají u velké většiny taxonů (Jetz et al., 2011).

3.7 Virtuální druhy a jejich přínos pro SDM

Z předchozí kapitoly, ve které jsem shrnul některá úskalí provázející SDM, vyplývá, že úspěšné modelování stojí a padá na dvou faktorech: kvalitních druhových datech a relevantních environmentálních prediktorech. Studie věnující se SDM a jejich porovnávání tyto faktory sice berou na vědomí (a jsou hojně diskutované), avšak míru, jakou jednotlivé chyby (či kombinace více chyb) ve vstupních datech ovlivňují výkon modelů, lze jen těžko kvantifikovat. Testovat tedy různé techniky modelování z hlediska přesnosti je s reálnými daty takřka nemožné (Meynard et Kaplan, 2013). Koncept virtuálního druhu vznikl přesně z těchto důvodů. Princip spočívá ve vygenerování rastru vhodnosti prostředí nad vybranou sadou environmentálních vrstev za pomoci odezvových funkcí (např. lineární, kvadratické, logistické, Gaussova, atd.), které definují reakci virtuálního druhu na dané podmínky. Z tohoto rastru se následně pomocí konverze (o obou používaných přístupech se zmíním v metodice) vytvoří sada prezenčně-absenčních dat, opět reprezentovaná rastrem (tentokrát binárním) - ten zobrazuje virtuální realitu, tedy výskyt námi stvořeného druhu. Nad těmito daty můžeme následně provést virtuální sampling (analogický sběru dat v terénu) a s nasbíraným vzorkem zacházet stejně jako při modelování distribuce reálných druhů.

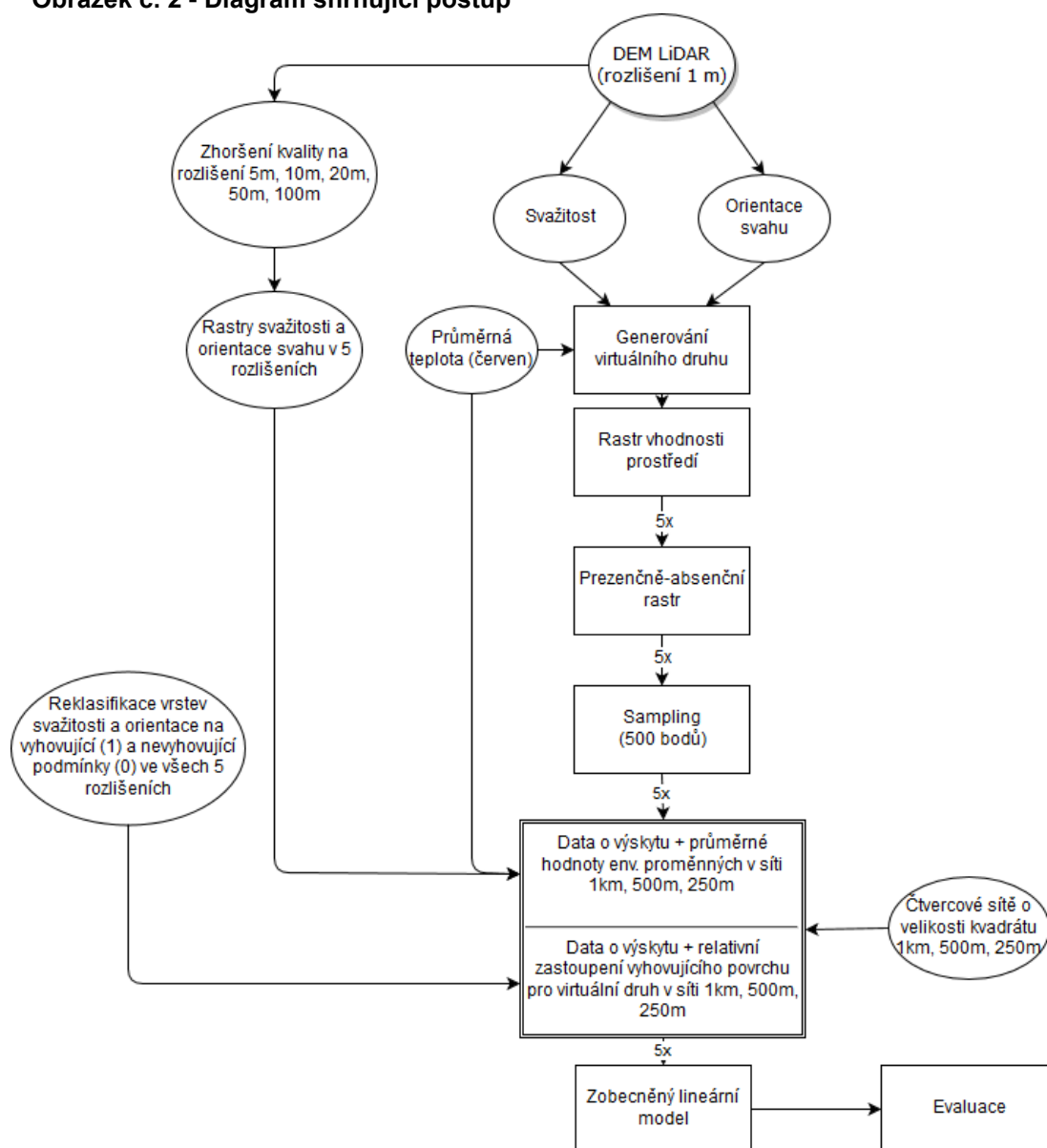
Obrovskou výhodou tohoto přístupu je, že můžeme porovnávat výstupy SDM s předem známou druhovou distribucí (virtuální) a zjistit, jak volba modelu a jeho

parametrů ovlivňuje jeho schopnost tuto virtuální distribuci replikovat (Miller, 2014). Simulovat můžeme i různé aspekty SDM (např. *sampling bias*, různá míra prevalence) bez rizika vzniku šumu pocházejícího z nespolehlivých vstupních dat (Moudrý, 2014). Všechna vstupní data včetně environmentálních proměnných zde tedy máme plně pod kontrolou (Hirzel et al., 2001) a můžeme se zaměřit na vliv jednotlivých segmentů na výkon modelů. Možnosti jsou široké. Nad virtuálními daty o výskytech lze provést sběr prezenčních i prezenčně-absenčních dat a testovat jejich vliv na výkon různých modelů. Lze také zkoumat vliv náhodného, systematického, stratifikovaného nebo i zkresleného sběru dat nebo vliv velikosti vzorku. Nemůže se zde stát, že model bude špatně kalibrován kvůli výběru nevhodných prediktorů. Je sice pravda, že definování virtuálního druhu na omezené sadě environmentálních vrstev značně zjednodušuje realitu, avšak díky tomu je možné se zaměřit na čistě technickou stránku věci. Pokud budeme znát přednosti a největší slabiny toho kterého modelu, můžeme získané poznatky uplatnit i v praxi a před reálným sběrem dat navrhnout metodiku ušitou na míru vlastnostem zkoumaného druhu a s ohledem na cíl studie či zájmovou oblast. Toto je obrovský přínos studií zabývajících se virtuálními druhy.

4 Metodika

Úpravu vstupních dat (vrstvy environmentálních proměnných) a následně prostorové analýzy jsem provedl v softwaru ArcMap 10.3. Pro vygenerování virtuálního druhu, sběr dat, model druhové distribuce a evaluaci modelu jsem použil software R 3.3.1 s nadstavbou R Studio 1.0.44, balíčky *Data.table*, *Virtualspecies*, *Raster* a *PresenceAbsence*.

Obrázek č. 2 - Diagram shrnující postup



4.1 Metodologie postupu

Celá procedura je představena na obrázku č. 2. Zájmovou lokalitou je část Krkonošského národního parku. Vstupní data reprezentují 3 environmentální proměnné, které jsem pro tuto práci zvolil: průměrná teplota v měsíci červnu, svažitost a orientace svahu. Topografické vrstvy jsem vytvořil na základě poskytnutého digitálního modelu reliéfu terénu (LiDAR) zahrnující zkoumanou oblast, zatímco klimatická data pochází z volně dostupné databáze WorldClim. Na těchto třech vrstvách jsem vygeneroval virtuální druh se specifickou reakcí na dané parametry prostředí. Vzniknuvší rastr vhodnosti prostředí jsem převedl na prezenčně-absenční rastr a na něm provedl sampling, jehož výstupem je seznam bodů s informacemi o prezenci či absenci virtuálního druhu. Modelování druhové distribuce jsem provedl s pomocí zobecněného lineárního modelu (GLM), který jsem sestavil pro 3 různá rozlišení (1 km, 500 m a 250 m). Do těchto 3 variant modelu následně vstupují prezenčně-absenční data o výskytech spolu s hodnotami environmentálních proměnných pocházejících ze vstupních vrstev o 5 rozlišeních (5 m, 10 m, 20 m, 50 m a 100 m) a je testována jejich přesnost s ohledem na kvalitu vstupních dat.

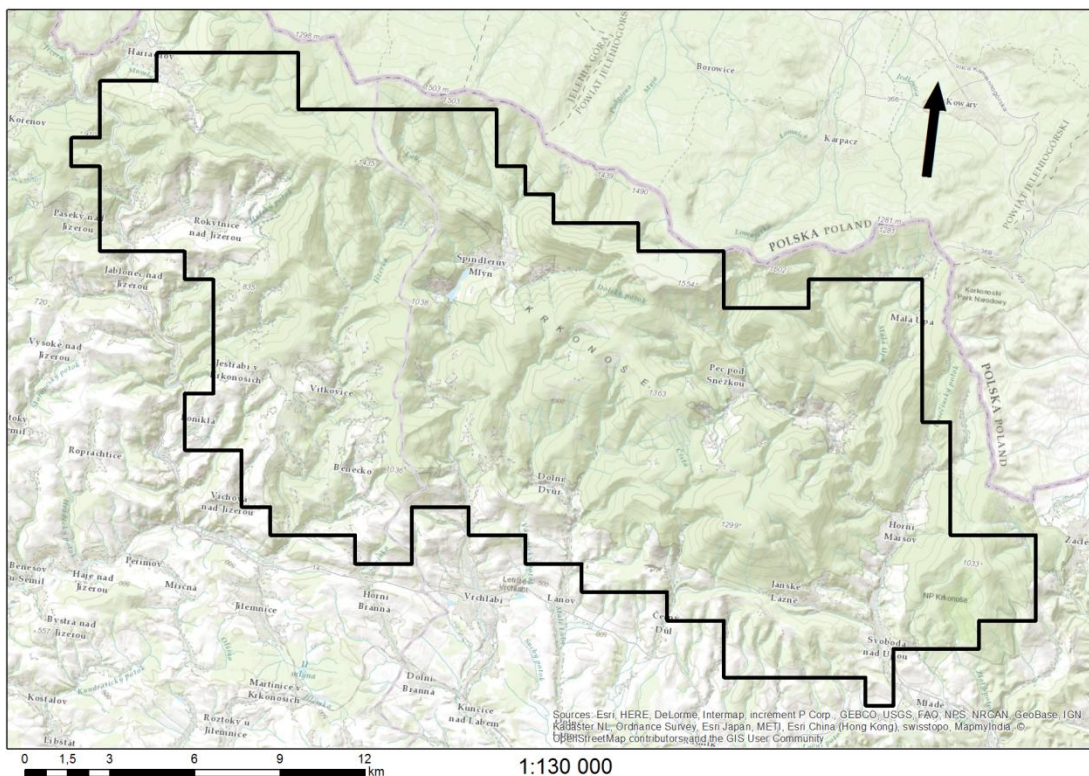
4.2 Vymezení a charakteristika zájmového území

Zkoumaná lokalita se nachází na území Krkonošského národního parku a pokrývá oblast o rozloze 397 km² (obr. č. 3). Krkonošský národní park byl vyhlášen v roce 1963 nařízením vlády a od roku 1992 je spolu s národním parkem na polské straně zařazen mezi biosférické rezervace UNESCO. Celé území také spadá pod evropsky významné lokality (soustava NATURA 2000) a jeho část je vymezena jako ptačí oblast. Předmětem ochrany zde jsou nejen rostliny a živočichové, ale i tradičně obhospodařované louky (KRNAP, 2017).

Z hydrologického hlediska území spadá do úmoří Severního a Baltského moře (rozvodí tvoří Slezský hřbet). Dominující řeky jsou zde Labe a Odra, avšak oblast je protkána více než 140 vodními toky, pro které je charakteristický prudký spád a značná heterogenita podélného sklonu koryta. Typické jsou četné dešťové i sněhové srážky způsobené převládajícím západním prouděním, které přináší vzduchové masy od Atlantiku. Proudění přes horský val pak kromě srážek přináší i vlhčí, chladnější klima oceánického charakteru. Národní park se pyšní také vysokou biodiverzitou. Žije zde přes 15000 druhů bezobratlých živočichů, přes 400 druhů obratlovců, 500 druhů mechorostů, přes 250 druhů lišejníků a cévnaté rostliny čítají

přes 1200 taxonů. Z ochránářského pohledu stojí za zmínku též vysoký počet endemitů (přes 40 % endemických druhů české květeny se nachází právě zde) (KRNAP, 2017).

Obrázek č. 3 - Vymezení zájmové lokality



4.3 Zdroje vstupních dat

4.3.1 WorldClim

WorldClim zahrnuje sady klimatických dat z celého světa o prostorovém rozlišení přibližně 1 km v souřadnicovém systému WGS 1984 a nabízí své využití v mapování, v prostorových analýzách GIS a v modelování druhové distribuce. Zdrojem těchto vrstev je množina dat získaných z měření meteorologických stanic, následně interpolovaných na celý povrch. Mezi dostupné klimatické proměnné patří například minimální, maximální a průměrná teplota, množství srážek, rychlost větru, solární radiace. Z časového hlediska data navíc nabízí i průměrné roční hodnoty, sezonalitu (např. teplotní rozptyl v průběhu roku) a extrémní jevy počasí (např. teplota v nejméně chladném/nejteplejším měsíci) (Hijmans et. al, 2005).

V této práci jsem databázi WorldClim využil k získání klimatické proměnné - průměrné červnové teploty v letech 1970 - 2000.

4.3.2 DEM LiDAR

Zkratka LiDAR pochází z anglického *Light Detection and Ranging* a zastřešuje soubor optických technologií sloužících k zaznamenání povrchu těles a měření vzdáleností. Princip spočívá ve vysílání laserového paprsku (emitem) a jeho následné detekci (detektorem) po odrazu od povrchu země nebo tělesa - podmínkou je, že emitor i detektor jsou umístěny ve stejné rovině. Aby bylo možné kromě tvaru objektu změřit i vzdálenost od snímače, jsou do přístrojového vybavení navíc zařazeny velmi přesné hodiny, které dokážou zaznamenat velmi krátký časový úsek mezi vysláním a detekcí odraženého paprsku. Pro mapování terénu se používají letadla vybavená patřičnou technologií spolu s GPS přístrojem pro určení polohy. Všechny tyto výhody skýtají široké využití například v tvorbě velmi přesných 3D modelů budov či celých měst, mapování vegetačního pokryvu a inženýrských sítí (nadzemní vedení). Jedno z nejvýznamnějších uplatnění technologie LiDAR je také v tvorbě digitálních modelů terénu (DMT či DEM), kde může přesnost mapování přesáhnout i rozlišení 1 m (Dolanský, 2004).

Digitální výškový model Krkonošského národního parku s velikostí buňky 1 m jsem zde použil pro tvorbu topografických vrstev svažitosti terénu a orientace svahu.

4.4 Úprava vstupních dat

4.4.1 Průměrná teplota

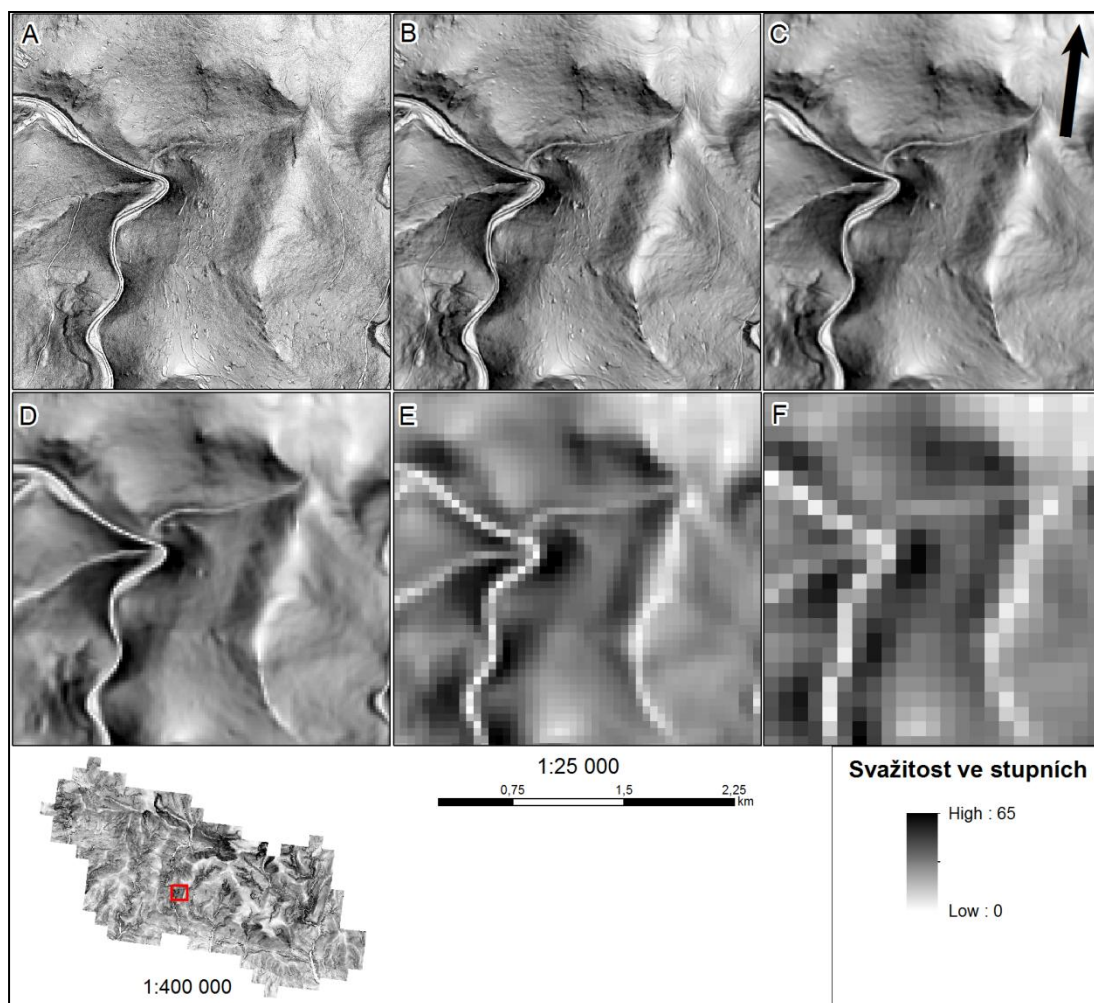
Zdrojová vrstva průměrné červnové teploty byla v souřadnicovém systému WGS1984, proto jsem ji nejprve pomocí nástroje *Project* převedl do souř. systému S-JTSK. Všechny vstupní vrstvy pro vygenerování virtuálního druhu musí mít totožnou velikost buňky, což průměrná teplota s rozlišením 1 km jako jediná nesplňovala, a tak jsem využil nástroj *Resample* k nastavení rozlišení 1 m. Původní hrubá prostorová informace zůstala zachována, pouze je nyní vrstva kompatibilní s daty vytvořenými z DEM LiDAR.

4.4.2 Svažitost

Výchozí vrstvou je digitální výškový model s rozlišením 1 m. K analýze svažitosti jsem v ArcMapu použil funkci *Slope*, která pro každou buňku rastru vypočítá gradient změny v nadmořské výšce vůči okolním 9 buňkám a ve výsledku je každé buňce přiřazena hodnota svažitosti ve stupních. Zatímco pro vygenerování

virtuálního druhu jsem použil nejpřesnější rozlišení 1 m², GLM bude jako vstupní data využívat 5 variant svažitosti o horším rozlišení (5 m, 10 m, 20 m, 50 m a 100 m). Nejdříve jsem tedy výchozí vrstvu nadmořské výšky pomocí nástroje *Aggregate* (metoda *Mean*) zhoršil do kýžených rozlišení a poté z nich pomocí nástroje *Slope* vypočítal svažitost (obr. č. 4).

Obrázek č. 4 - Porovnání kvality rastrů svažitosti pro velikost buňky a) 1 m, b) 5 m, c) 10 m, d) 20 m, e) 50 m, f) 100 m

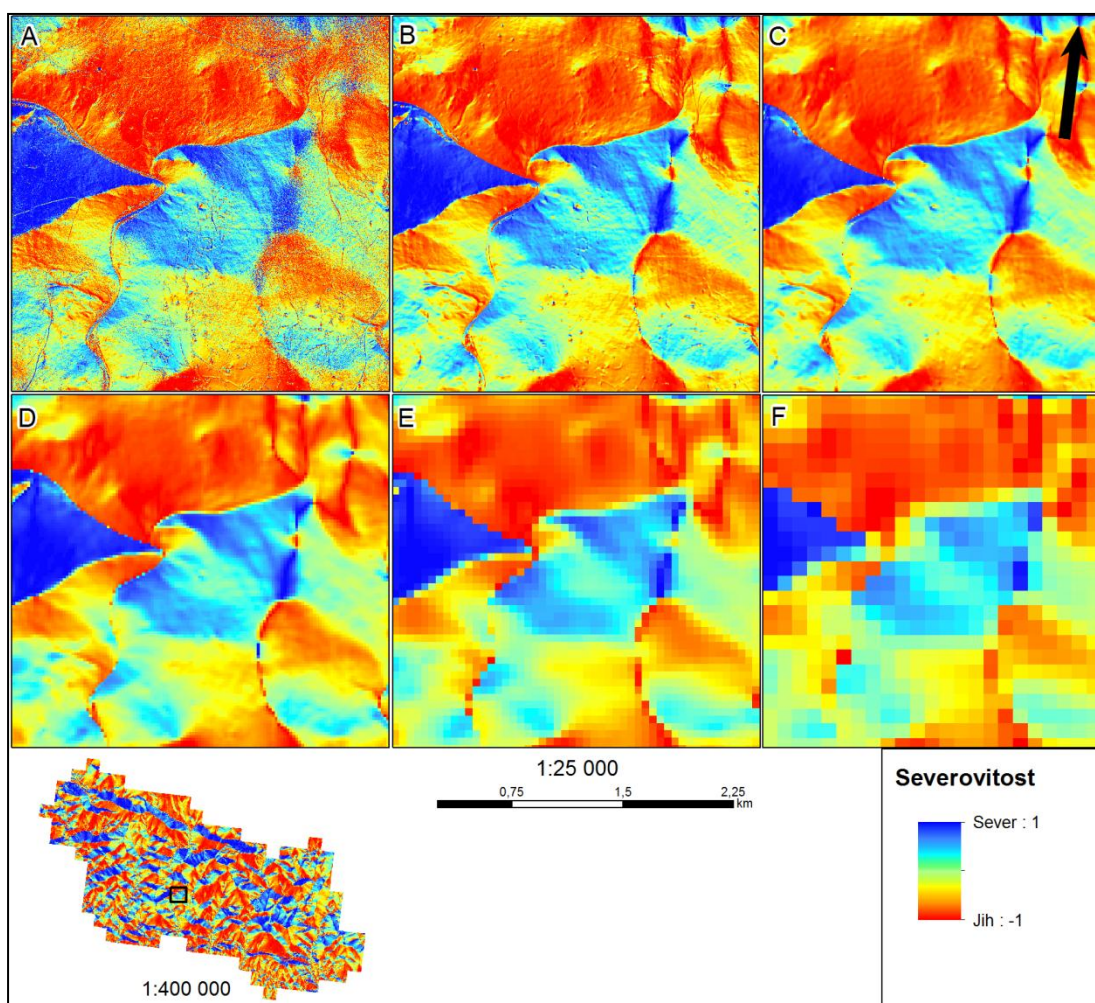


4.4.3 Orientace svahu

Aby mohla být vstupní data použita pro generování virtuálního druhu a modelování druhové distribuce, je nezbytné, aby byl gradient podmínek prostředí kontinuální (teplota i svažitost toto splňují). U orientace svahu je však gradient přetržen při přechodu z 359° na 0°, což znemožňuje definování afinity druhu ke světové straně pomocí Gaussovy funkce. Z tohoto důvodu jsem se inspiroval jedním z nástrojů programu GRASS GIS - *northernness-easterness*, který pomocí funkcí sinus a cosinus převádí orientaci svahu ve stupních na kontinuální gradient (jednotková

kružnice). Pro účely této práce bylo potřeba definovat virtuální druh, jenž nesnáší severně orientované svahy, zvolil jsem tedy k převodu funkci cosinus, která nabývá hodnot od 1 (sever, 0°), přes 0 (západ/východ) do -1 (jih, 180°). Takto je možné stanovit životní optimum druhu na hodnotu -1 (jih) s postupným snižováním vhodnosti prostředí z obou stran směrem k severu.

Obrázek č. 5 - Porovnání kvality rastrů „severovitosti“ pro velikost buňky a) 1 m, b) 5 m, c) 10 m, d) 20 m, e) 50 m, f) 100 m



Postup je analogický s předchozí kapitolou. Z rastrů nadmořské výšky o 5 rozlišeních jsem v ArcMapu nejprve nástrojem *Aspect* vytvořil vrstvy nesoucí informaci o orientaci svahu ve stupních a ty jsem následně pomocí funkce *Cos*, převedl na rastry vyjadřující hodnoty „severovitosti“ (obr. č. 5) (*northernness* postrádá vhodný český ekvivalent, pozn. autora).

4.5 Virtuální druh

K vytvoření virtuálního druhu jsem použil volně dostupný balíček *virtualspecies* pro software R. V prvním kroku jsem sdružil všechny 3 rastry environmentálních proměnných (exportované z ArcMapu do formátu GTiff) do tzv. RasterStacku a nadefinoval odezvové funkce ke každé proměnné. Tyto funkce vyjadřují vztah virtuálního druhu k podmínkám prostředí a výstupem je jediný rastr vhodnosti prostředí (*environmental suitability*), jehož buňky nabývají hodnot od 0 (naprosto nevhodné pro život) do 1 (ideální podmínky pro výskyt) (obr. č. 6).

```
library(virtualspecies)
library(raster)

#Nahrání vstupních dat
bio1 <- raster("C:\\DP\\teplota.tif")
bio2 <- raster("C:\\DP\\slope.tif")
bio3 <- raster("C:\\DP\\north.tif")

bio<-stack(bio1,bio2,bio3)
names(bio)<-c("bio1","bio2","bio3")

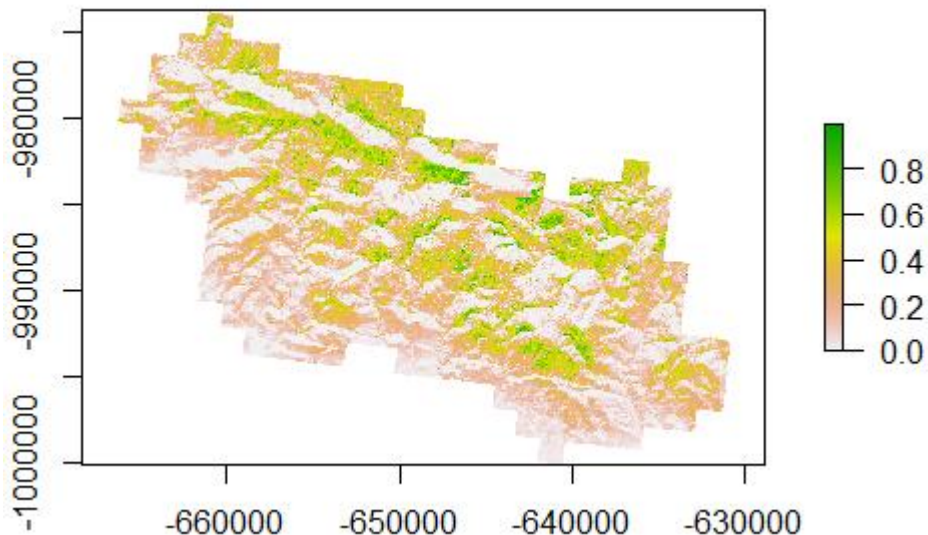
# Definice odezvoových funkcí
parametry <- formatFunctions(bio1 = c(fun = 'dnorm', mean = 120, sd = 30),
                             bio2 = c(fun = 'betaFun', p1 = 0, p2 = 60, alpha = 0.6, gamma = 0.08),
                             bio3 = c(fun = 'custnorm', mean = -1, diff = 1.6, prob = 0.99))

# vygenerování rastru vhodnosti prostředí
virtual.sp <- generateSpFromFun(raster.stack = bio,
                               parameters = parametry,
                               plot = TRUE)
```

Pro získání skutečné informace o výskytech je třeba rastr vhodnosti prostředí nejprve převést na prezenčně-absenční rastr. Až donedávna se pro tento účel široce využívala tzv. prahová metoda, která stanovila pevnou mez, nad níž se druh vyskytuje vždy a pod níž se nevyskytuje nikdy (např. všechny buňky s hodnotou vhodnosti prostředí nad 0,5 budou počítány jako prezenze, zbylé jako absence). Od této metody se postupně upouští, jelikož produkuje méně realistická data o výskytu. V první řadě se takto odbourává smysl informace plynoucí z rastru vhodnosti prostředí (hodnoty od 0 do 1 reprezentují již jen holá čísla sloužící k pevnému dichotomickému dělení). Za druhé hůře reflektuje reálné chování druhů. Při terénním průzkumu můžeme snadno narazit na případ, kdy je pozorovaný druh na vhodné lokalitě nepřítomen (kompetice, predace, migrační bariéry, náhodný jev) nebo naopak přítomen v nevyhovujících podmínkách (mutualismus, facilitace, důsledky migrace) (Leroy et al., 2015). Za třetí tímto způsobem vzniká vždy naprosto stejná sada dat nehledě na počet opakování (na rozdíl od pravděpodobnostního přístupu) (Meynard et Kaplan, 2013, Miller, 2014). Na prezenčně-absenčních rastroch tato metoda vytváří shluky pevně oddělených

výskytů, jejichž hranici může tvořit vrstevnice, izoterma či jiná izočára kontinuálního rastru vstupních dat.

Obrázek č. 6: Rastr vhodnosti prostředí virtuálního druhu

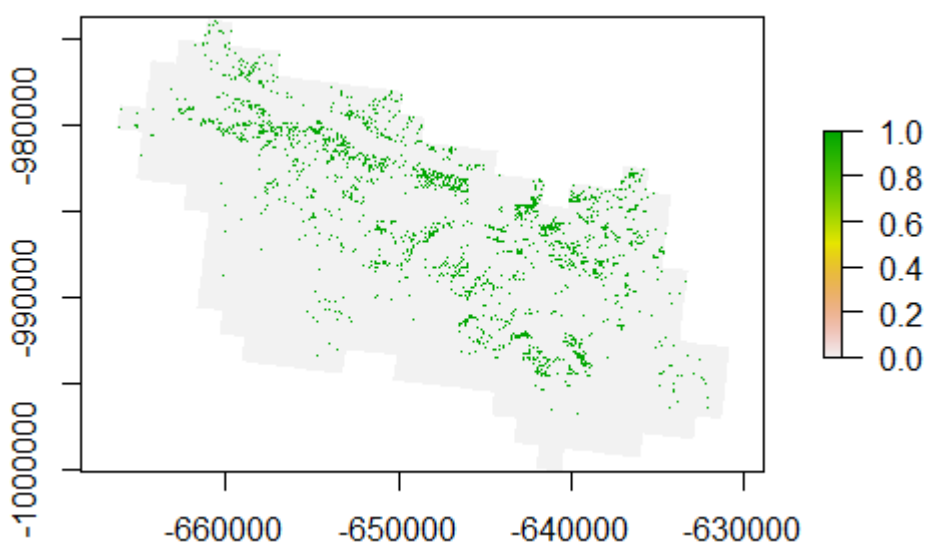


K překonání této nevýhody se používá pravděpodobnostní přístup namísto meze. Hodnoty vhodnosti prostředí jsou zde pomocí logistické či lineární funkce převedeny na pravděpodobnost výskytu, přičemž tvar funkce je v plné režii uživatele. Podobně jako v reálném světě se tak virtuální druh může objevit v celém environmentálním gradientu (se šancí na výskyt určenou podle povahy podmínek prostředí) (Moudrý, 2014). Ve skriptu se upravují pouze hodnoty alfa (sklon křivky) a beta (poloha inflexního bodu). Ve své práci jsem zvolil alfa = -0,07 a beta = 0,65. Konverzi do prezenčně-absenčního rastru jsem provedl 5x, čímž dostávám 5 odlišných verzí výskytů virtuálního druhu (obr. č. 7).

```
# Generování prezenčně-absenčního rastru
p_a <- convertToPA(virtual.sp,
                  beta = 0.65, alpha = -0.07,
                  plot = TRUE)
```

Pro využití dat k modelování druhové distribuce se na prezenčně-absenčních rastrech provádí sampling (analogie sběru dat v terénu). Před jeho započítím jsem v ArcMapu získané rastry ořízl nástrojem *Clip* pouze na zájmovou oblast (dle obr. č. 3).

Obrázek č. 7: Ukázka jedné varianty prezenčně-absenčního rastru



Následně jsem stanovil počet „odběrů“ na 500 bodů, provedl náhodný sběr na každém z 5 rastrů a získaná data exportoval do tabulky, která pro každý bod obsahuje souřadnice x, y a informace o výskytu (0/1).

```
# sampling 500 bodů
presence.points <- sampleOccurrences(p_a,
                                   n = 500,
                                   type = "presence-absence")

body <- presence.points$sample.points ## extrakce bodů výskytu
write.table (sampling$sample.points,"C:/DP/p_a.txt", sep=";",row.names=FALSE) ## uložení tabulky s výskyty
```

Všech 5 tabulek jsem následně importoval do ArcMapu (File > Add Data > Add XY Data) a vytvořil z nich bodové vrstvy.

4.6 Příprava dat pro modelování druhové distribuce

Zde se postup rozděluje do dvou větví. V prvním přístupu budou do GLM o rozlišení 250 m, 500 m a 1 km vstupovat data o výskytech a rastry environmentálních proměnných s velikostí buňky 5 m, 10 m, 20 m, 50 m a 100 m. Ve druhém přístupu rastry teploty, svažitosti a orientace svahu převedu na binární rastry vyjadřující pouze prostředí vhodné (1) a nevhodné (0). V GLM se pak místo s průměrnými hodnotami bude počítat s relativním zastoupením vyhovujícího prostředí pro každé rozlišení modelu.

4.6.1 Metoda průměrných hodnot podmínek prostředí

V rámci studované oblasti jsem nejprve vytvořil 3 čtvercové sítě polygonů (nástroj *Create Fishnet*) pro každé rozlišení modelu: 1 km, 500 m, 250 m. Na všechny vstupní vrstvy o rozlišení 5 m, 10 m, 20 m, 50 m a 100 m jsem použil nástroj *Aggregate*, který vytvořil nové rastry o velikosti buněk 1 km, 500 m a 250 m s průměrnými hodnotami původních dat. K bodovým vrstvám výskytů druhu jsem posléze pomocí nástroje *Extract Multi-values to Points* tyto průměrné hodnoty přiřadil. Na všechny 3 sítě jsem nakonec použil funkci *Spatial Join*, která ke každému čtverci připojila hodnoty průměrných podmínek prostředí a informace o prezenci/absenci druhu. Příliš velké kvadráty sítě přinášejí úskalí, že v místech, kde se na 1 polygonu nachází 2 a více pozorovaných prezencí/absencí, je ne/přítomnost druhu počítána pouze jako jeden záznam. Se zvyšující se velikostí zrna tedy postupně přicházíme o cenná data a snižujeme velikost vzorku (model s rozlišením 250 m bude mít vyšší počet záznamů než model s rozlišením 1 km). Řešením může být zachování počtu pozorovaných prezencí/absencí v každém kvadrátu a tím i zachování velikosti vzorku (Guisan et al., 2007), nicméně z důvodu hrozící pseudoreplikace jsem pro další analýzu počet záznamů v kvadrátu nezohlednil (čili pokud se v polygonu sítě vyskytla alespoň jedna prezence, výsledná hodnota je 1, nehledě na počet absencí či dalších prezencí). Výsledkem byly 3 polygonové vrstvy nesoucí informaci o výskytu druhu, počtu pozorování a průměrné hodnoty všech environmentálních proměnných v 5 rozlišeních. Atributové tabulky (celkem 15) jsem exportoval do formátu .csv pro potřeby modelování.

4.6.2 Metoda relativního zastoupení vyhovujících podmínek

Nevýhodou metody průměrování je, že ji v heterogenní krajině mohou při výpočtu silně ovlivňovat extrémní hodnoty. Kvadráty sítě reprezentující rozlišení modelu například zahrnují místa, kde povětšinou rovinný terén střídá prudké zvýšení sklonu na jednom okraji anebo se skokově mění hodnoty orientace svahu (překrývá-li kvadrát vrchol kopce či horský hřeben). V tomto metodickém přístupu jsem tedy namísto průměrných hodnot environmentálních proměnných použil hodnoty relativního zastoupení vyhovujících podmínek (práh rozlišující prostředí na vhodné/nehodné jsem definoval na základě znalosti odezвовých funkcí při generování virtuálního druhu), jelikož pro modelování může být důležitá i informace, že buňku tvoří například ze 75 % vyhovující podmínky (kromě průměrů abiotických proměnných ovlivnitelných extrémními hodnotami nevyhovujících míst).

Zdrojová data jsem vytvořil pomocí nástroje *Reclassify* ze vstupních vrstev podle následujícího klíče:

	Nové hodnoty rastru	
	1	0
Rastr svažitosti	$x = \langle 25; \text{MAX} \rangle$	$x = \langle 0; 25 \rangle$
Rastr orientace svahu	$x = \langle -1; 0,6 \rangle$	$x = \langle 0,6; 1 \rangle$

V dalším kroku jsem nástrojem *Tabulate Area* vypočítal počet pixelů s hodnotou 1 v každém kvadrátu tří sítí. Relativní zastoupení vyhovujících podmínek je pak: (počet pixelů s hodnotou 1) / (celkový počet pixelů). Vypočítané hodnoty jsem následně připojil nástrojem *Join* k polygonovým vrstvám již nesoucím informace o výskytech druhu a tabulky exportoval do formátu .csv.

4.7 Modelování druhové distribuce a evaluace

Pro potřeby modelování druhové distribuce jsem použil techniku tzv. křížové validace, při které je soubor dat náhodně rozdělen na několik skupin (v této práci na 5), přičemž jedna skupina slouží pro evaluaci modelu (otestování predikční schopnosti modelu) a zbylé jsou kalibrační. Počet iterací je pak roven počtu skupin a evaluační vzorek dat postupně tvoří každá z nich. Výstupy z jednotlivých iterací modelu jsou následně průměrovány.

Samotné modelování jsem provedl pomocí zobecněného lineárního modelu (GLM) a jako prediktory použil hodnoty environmentálních proměnných v 5 rozlišeních (vrstva teploty jen v rozlišení 1 km). V druhém metodickém přístupu však do GLM jako prediktory vstupují pouze svažitost a orientace svahu (vrstva teploty v rozlišení 1 km reklasifikovaná do binární podoby by ohrozila nafitování modelu).

Jako hodnotící ukazatele modelu jsem zvolil AUC, senzitivitu a specifitu. Aby bylo možné vypočítat poslední dvě zmíněné hodnoty, je třeba nejdříve stanovit mez výskytu (*threshold*). Zde jsem vybral metodu *MaxSens+Spec*, jež má za cíl maximalizovat senzitivitu a specifitu. Výsledky všech hodnotících ukazatelů každého z 5 vzorků jsem zprůměroval za jednotlivá rozlišení modelu i rozlišení vstupních dat.

V poslední řadě jsem výstupy GLM (hodnoty vhodnosti prostředí pro každý kvadrát použitých tří rozlišení modelu) importoval do ArcMapu, připojil k nim pomocí nástroje

Join komplementární hodnoty pocházející z rastru vhodnosti prostředí vytvořeném při generování virtuálního druhu a vypočítal Pearsonův korelační koeficient pro jednotlivá rozlišení modelu i vstupních dat. Použitý skript je na obrázku č. 8.

Obrázek č. 8 - Skript použitý pro modelování a evaluaci

```

library(PresenceAbsence)
DATA <- read.table("D:\\Zaloha M\\csv\\250m_pa02.csv", header=T, sep=";") # postupne nactani vstupnich dat
attach(DATA)

## Krizova validace

row_number <- nrow(DATA);row_number      # Nahodne rozdeleni vzorku na 5 skupin
group <- rep(1:5, length = row_number)
group <- sample(group)
D <- numeric(0)
D_all <- numeric(0)
PA_all <- numeric(0)

-----

## GLM iterovano pro 5 skupin

for(i in 1:5) { group.glm <- glm(pa~slope_5m+temp_1km+north_5m,family= binomial, data=DATA[group != i,])
  predict_s1 <-predict(group.glm, newdata=DATA[group == i,], type="response")

  D <- as.vector(predict_s1)
  D_all <- c(D_all,D)
  PA <- DATA[group == i,c("IDENT","pa")]
  PA_all <- rbind(PA_all,PA)
}
PA_all <- PA_all[,c("IDENT","pa")]
DAT <- data.frame (PA_all, D_all)

-----

## Vypocet hodnoticich ukazateleu

optresh <- optimal.thresholds(DAT, opt.methods="MaxSens+Spec", which.model = 1)
t <-optresh[,2];t
cmx <- cmx(DAT, threshold = t, which.model = 1, na.rm = FALSE);cmx
pcc(cmx, st.dev = TRUE)
presence.absence.accuracy(DAT, threshold = t, find.auc = TRUE, st.dev = TRUE)

write.table (DAT,"D:\\DP\\DAT\\250m_5s1.csv",sep=";",row.names=FALSE)
detach(DATA)

-----

## Vypocet Pearsonova korelacniho koeficientu

library(data.table)

s1 <- read.table("D:\\DP\\DAT\\250m_5s1.csv", header=T, sep=";")
s2 <- read.table("D:\\DP\\DAT\\250m_5s2.csv", header=T, sep=";")
s3 <- read.table("D:\\DP\\DAT\\250m_5s3.csv", header=T, sep=";")
s4 <- read.table("D:\\DP\\DAT\\250m_5s4.csv", header=T, sep=";")
s5 <- read.table("D:\\DP\\DAT\\250m_5s5.csv", header=T, sep=";")
mean <- rbindlist(list(s1,s2,s3,s4,s5))[,lapply(.SD,mean), list(IDENT)]

write.table (mean,"D:\\DP\\DAT\\Korelace\\250m_5m.csv",sep=";",row.names=FALSE)

# Tabulka byla nasledne v ArcGIS propojena s hodnotami vhodnosti prostredi pochazejicimi z generovani
virtuálního druhu a nactena zpet do R pro vypocet korelacniho koeficientu

hodnoty <- read.table("D:\\DP\\DAT\\Korelace\\250m_5m_join.csv", header=T, sep=";")
cor(hodnoty$D_all,hodnoty$MEAN)

```

5 Výsledky

5.1 Metoda průměrných hodnot podmínek prostředí

Rozlišení environmentálních proměnných nemělo výraznější vliv na výkon jednotlivých modelů dle AUC (tab. č. 1). Pouze u 250m modelu je při zhoršování rozlišení vidět sestupný trend, avšak jen o 0,03 mezi 5m a 100m rozlišením. 500m a 1km modely nebyly snížením rozlišení téměř vůbec ovlivněny, u 1km modelu je dokonce vidět slabé zlepšení výkonu mezi rozlišením 5 m a 100 m. Ukázalo se však, že mnohem větší vliv má rozlišení samotných modelů, u kterých je pokles AUC zřetelný při všech rozlišeních environmentálních proměnných. Senzitivita (podíl správně předpovězených prezencí) nevykazuje žádný výrazný trend napříč všemi rozlišeními, zatímco specifická (podíl správně předpovězených absencí) se chová podobně jako AUC - nejvíce byla ovlivněna kvalitou samotných modelů než rozlišením prediktorů (tab. č. 2).

Tabulka č. 1 - Hodnoty AUC

	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,846	0,841	0,843	0,835	0,813
500 m	0,814	0,810	0,815	0,807	0,801
1 km	0,763	0,756	0,757	0,769	0,780

Tabulka č. 2 - Hodnoty senzitivity a specifacity

SENZITIVITA	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,835	0,857	0,849	0,829	0,819
500 m	0,847	0,812	0,853	0,814	0,738
1 km	0,854	0,843	0,814	0,849	0,831
SPECIFICITA	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,777	0,750	0,760	0,754	0,755
500 m	0,700	0,726	0,703	0,721	0,787
1 km	0,635	0,627	0,653	0,657	0,674

Pearsonův korelační koeficient udává vztah mezi hodnotami vhodnosti prostředí pocházející z modelu a z virtuální reality (tab. č. 3). Největší korelace je patrná u 250m modelu při použití 5m a 10m rozlišení prediktorů. U něj se také mírně projevuje zhoršování kvality prediktorů, zatímco 500m a 1km modely už rozlišení prediktorů nijak zvlášť neovlivňuje. I zde však má hlavní váhu rozlišení samotných modelů.

Tabulka č. 3 - Pearsonův korelační koeficient

	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,683	0,683	0,670	0,679	0,629
500 m	0,649	0,650	0,647	0,650	0,636
1 km	0,606	0,608	0,594	0,616	0,613

5.2 Metoda relativního zastoupení vyhovujících podmínek

Při použití této metody nebyl nalezen žádný vliv změny kvality modelů ani prediktorů na hodnoty AUC (tab. č. 4). Sice má 250m model nejlepší výkon při použití nejkvalitnějších prediktorů, avšak toto mohlo být i dílem náhody, jelikož v žádném směru není vidět jakýkoliv trend.

Tabulka č. 4 - Hodnoty AUC

	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,619	0,581	0,575	0,602	0,592
500 m	0,606	0,563	0,575	0,582	0,576
1 km	0,593	0,594	0,592	0,576	0,602

Stejně tak hodnoty senzitivity a specifity neukazují žádný vzor (tab. č. 5). Z Pearsonova korelačního koeficientu je i vidět, že kalibrace modelu postaveného na relativním zastoupení vyhovujících podmínek byla velmi špatná a nepodařilo se reflektovat virtuální realitu (tab. č. 6).

Tabulka č. 5 - Hodnoty senzitivity a specificity

SENZITIVITA	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,198	0,281	0,232	0,175	0,350
500 m	0,330	0,303	0,325	0,278	0,221
1 km	0,641	0,568	0,495	0,290	0,627
SPECIFICITA	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,908	0,858	0,893	0,917	0,722
500 m	0,861	0,848	0,785	0,881	0,905
1 km	0,476	0,554	0,628	0,774	0,482

Tabulka č. 6 - Pearsonův korelační koeficient

Pearsonův korelační koeficient	Rozlišení env. proměnných				
Rozlišení modelu	5 m	10 m	20 m	50 m	100 m
250 m	0,143	0,136	0,121	0,103	0,197
500 m	0,072	0,078	0,079	0,056	0,005
1 km	0,103	0,086	0,072	0,089	0,074

6 Diskuze

Zaměřím-li se nejprve na výsledky prvního přístupu - metodu průměrování hodnot podmínek prostředí v kvadrátech, pokles hodnot AUC při zhoršování rozlišení modelu koresponduje s výsledky několika podobně zaměřených studií. Nezer et al. (2017) porovnávali predikční schopnost modelu v rozlišeních 10 m, 100 m a 1 km a objevili výrazný pokles hodnoty AUC při zhoršování rozlišení (rozdíl 0,3 mezi 10m a 1km modelem). V další studii objevili taktéž strmý pokles mezi rozlišením 200 m a 5 km (Guisan et al., 2007 ex. Ferrier et Watson, 1997). Ve své práci jsem zaznamenal pouze slabší pokles v řádu setin, nicméně jedni autoři provedli až 100násobné zhoršení rozlišení a druzí 25násobné, zatímco v mém případě bylo jen 4 násobné, což by mohlo být jedním z důvodů, proč byl pokles AUC tak malý. Pradervand et al. (2013) při postupném snižování rozlišení z 2 m do 100 m u modelování distribuce 239 druhů bylin zaznamenali také pouze slabé snížení predikční schopnosti modelu, a to jen u zhruba poloviny zkoumaných druhů. U zbytku žádný trend patrný nebyl, což autoři vysvětlují výběrem prediktorů, které neměly rovnocenný význam pro všechny byliny. Díky použití virtuálního druhu se výskyt tohoto druhu chyb v mé práci projevit nemohl. Stejně tak Sheeren et al. (2014) ve své ptačí studii, kde použili jako prediktor index NDVI v různém rozlišení, objevili velmi malý pokles výkonu modelu při zhoršování rozlišení. Jejich studie je však výjimečná v tom, že pro zkoumání vlivu rozlišení nevycházeli z velmi kvalitních environmentálních dat, která by následně pomocí GIS softwaru zhoršili do požadované kvality, ale použili výstupy z různých senzorů. Toto věrněji napodobuje situaci, se kterou se při modelování můžeme opravdu setkat, tedy že máme pro část zájmové oblasti k dispozici velmi kvalitní výškový model (LiDAR), zatímco pro zbytek je dostupný například pouze SRTM model (Shuttle Radar Topography Mission) v rozlišení 30 m, což jsou zcela odlišné technologie. Nicméně výsledky této studie korespondují s ostatními autory a mojí prací i přes to, že environmentální vrstvy v horší kvalitě byly vytvořeny uměle pomocí *resamplingu*. Podobně negativní vliv nižšího rozlišení na výkon modelu objevili též Song et al. (2013) u savců a Wan et al. (2016) u invazních stromů. Velkou srovnávací studii provedli také Guisan et al. (2007), kteří zkoumali vliv desetinasobného zhoršení rozlišení (ze 100 m do 1 km) a objevili také pouze slabé snížení hodnot AUC (v řádu setin, stejně jako v mé práci), a to zejména u modelů, které měly i při nižším rozlišení dobrou predikční schopnost, což potvrzují i Dyer et al. (2013). Zajímavé je srovnání z hlediska množství záznamů v kalibračním vzorku dat. Modely mající k dispozici větší kalibrační vzorek byly více citlivé na zhoršení rozlišení. Naopak modely, které byly kalibrovány malým vzorkem dat, tak

výrazný pokles nevykazovaly (jejich špatná kalibrace mohla efekt zhoršeného rozlišení zakrýt). Ve své práci jsem použil 500 prezenčně-absenčních záznamů, jejichž počet ještě poklesl úpravou do jednotlivých rozlišení modelu (a tím poklesl i jejich počet v kalibračním vzorku), což může být další z příčin malých rozdílů v AUC. Gottschalk et al. (2011) ve své studii SDM na ptačích druzích objevili také slabší pokles predikční schopnosti modelu při zhoršování rozlišení z 1 m na 1000 m. Jejich data o výskytech byla v rozlišení 5 m a z velké části tak odpovídala rozlišením environmentálních proměnných, a proto autoři podotýkají, že vysoké rozlišení prediktorů nezlepší výkon modelu, pokud nemáme i druhová data v odpovídající kvalitě. V mé práci byla rozlišení všech prediktorů vyšší než rozlišení modelu (druhových dat), nicméně jsem u 250m modelu přesto pozoroval mírný pozitivní vliv vyššího rozlišení prediktorů na jeho výkon. Rozdíl v AUC mezi 5m a 100m kvalitou prediktorů činil 0,03, u 500m modelu jen 0,01 a u 1km modelu už nebyl zaznamenán žádný pozitivní vliv. Volba nejkvalitnějších vrstev prediktorů tedy může mít mírný pozitivní vliv na výkon modelu, i když nejsou druhová data v odpovídající kvalitě, avšak toto tvrzení bych nezobecňoval, jelikož zrovna u 1km modelu byla hodnota AUC nejvyšší při použití 100m (nejméně kvalitních) prediktorů. Je tedy možné, že kalibrace 1km modelu byla tak špatná, že žádné rozlišení prediktorů již nemohlo jeho výkon zlepšit. Na druhou stranu v případě 250m modelu mohly průměrné hodnoty env. proměnných pocházejících z 5m vrstev lépe popsat podmínky panující v daném kvadrátu a přispět tak ke zvýšení predikční schopnosti modelu. Každopádně rozdíl 0,03 je příliš malý a tato otázka by si zasloužila další pozornost v širší paletě rozlišení jak prediktorů, tak druhových dat.

Vrátím-li se zpět ke studiím zkoumajícím vliv rozlišení druhových dat na výkon SDM, ne vždy byl pozorován pozitivní vztah. Např. Xu et al. (2016) ve své práci zabývající se slony nezjistili žádnou vazbu mezi použitým rozlišením a predikční schopností modelu. Dokonce lze najít i případ, kdy se horší rozlišení odrazilo ve zlepšeném výkonu modelu (Guisan et al., 2007 ex. Tobalske, 2002). Nekonzistentnost závěrů ukazuje, že do hry vstupují i jiné jevy, například velikost zájmové lokality, vzácnost studovaného druhu či jeho vlastnosti (alokace zdrojů, sesilní/mobilní, generalista/specialista) (Guisan et al., 2007). Ve své práci jsem využil konceptu virtuálního druhu, díky kterému nebezpečí výskytu chyb pramenících z nepřesných druhových dat či špatně zvolených prediktorů. Navíc jsem použil prezenčně-absenční vstupní data, která jsou, co se týče vlivu na výkon modelů, nejlepší možnou variantou. Nicméně při interpretaci výsledků je třeba vzít v potaz i prevalenci, tedy podíl prezencí v celkovém počtu záznamů. Bylo dokázáno, že

threshold dependent evaluační metriky (senzitivita/specifická) jsou velmi citlivé právě na prevalenci v kalibračním a evaluačním vzorku (McPherson et al., 2004), a to zejména u prezenčně-absenčních dat (Guissan et Thuiller, 2005). V této práci byla prevalence relativně nízká u všech rozlišení modelu (přibližně 0,25), což může být důvod, proč senzitivita nevykazuje žádný trend napříč rozlišeními (prezenčních záznamů je příliš málo).

Ve druhém metodickém přístupu jsem namísto průměrných hodnot environmentálních proměnných v každém kvadrátu počítal relativní zastoupení vyhovujícího prostředí. Idea, že vysoké zastoupení vyhovujících podmínek v mapovacím kvadrátu determinuje výskyt druhu, se v případě tohoto konkrétního virtuálního druhu bohužel neosvědčila. Kamenem úrazu mohla být i nevhodně zvolená reklasifikace, podle které jsem vrstvy prediktorů převáděl do binární podoby. Dokonce i samotná binární podoba vrstev, nehledě na zvolené intervaly či povahu zkoumaného druhu, mohla při zvolených velikostech kvadrátů ohrozit kalibraci modelu, navíc ruku v ruce s nízkou prevalencí. Tento přístup by se možná lépe uplatnil při modelování druhové distribuce mobilních organismů, pro které může být podíl určitého typu povrchu v jejich okolí rozhodující. To už však vyžaduje využití jiného druhu jako předmětu studie a také jiných prediktorů (např. *landuse*, *landcover*). Žádná studie, která by porovnávala výkon SDM postaveném na tomto přístupu a na metodě průměrování v různých rozlišeních, však dosud nebyla provedena.

7 Závěr

V této práci jsem za pomoci virtuálního druhu porovnával vliv rozlišení environmentálních proměnných na modely druhové distribuce v různé kvalitě. K dosažení tohoto cíle jsem zvolil dva metodické přístupy. V prvním přístupu byly každému mapovacímu kvadrátu přiřazeny průměrné hodnoty environmentálních proměnných, zatímco druhý přístup počítal s relativním zastoupením vyhovujících podmínek v kvadrátu. Výsledky prvního přístupu ukázaly, že více než rozlišení environmentálních prediktorů má na výkon modelů vliv spíše rozlišení samotného modelu. Pokles hodnot AUC od 250m, přes 500m až do 1km modelu byl zřetelný při všech použitých rozlišeních prediktorů, avšak tento pokles nebyl příliš strmý (v řádu setin), což koresponduje s výsledky většiny příbuzných studií. Vliv rozlišení prediktorů se ukázal jako marginální, a to ještě jen u 250m modelu, u kterého činil pokles hodnoty AUC mezi 5m a 100m kvalitou prediktorů pouze 0,03. Ostatní dva modely nebyly rozlišením prediktorů ovlivněny. U druhého metodického přístupu jsem neobjevil žádný vliv rozlišení prediktorů ani druhových dat na kvalitu modelů, což mohlo být způsobeno nízkou prevalencí, vlastnostmi virtuálního druhu nebo výběrem prediktorů, pro které nebyl převod do binární podoby vhodný. Co se týče hypotézy definované v cílech práce, ta byla potvrzena jen částečně. Vyšší rozlišení vstupních druhových dat opravdu má pozitivní vliv na predikční schopnost SDM, naproti tomu rozlišení vstupních vrstev prediktorů hrálo už jen minimální roli.

Další možné rozšíření této práce by mohlo zahrnout více kategorií rozlišení druhových dat (např. od 250 m až do 25 m) a zjistit, zda se AUC bude zvyšovat stejně pozvolně, jako mezi 1km a 250m modelem. Spolu s rozlišením by bylo zajímavé za pomoci virtuálního druhu manipulovat i s velikostí vzorku druhových dat a prevalencí, protože tyto faktory kalibraci modelu ovlivňují a otázkou je do jaké míry a jak je tento vliv silný v různých rozlišeních.

8 Přehled literatury a použitých zdrojů

Aguirre-Gutiérrez, J., Carvalheiro, L., Polce, C., van Loon, E., Raes, N., Reemer, M., Biesmeijer, J., 2013. Fit-for-Purpose: Species Distribution Model Performance Depends on Evaluation Criteria – Dutch Hoverflies as a Case Study. *Plos One* 8.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33, 1677–1688.

Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species–climate impact models under climate change. *Global Change Biology* 44, 53.

Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology* 89, 2712–2724.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3, 327–338.

Bombi, P., D’Amen, M., 2012. Scaling down distribution maps from atlas data: a test of different approaches with virtual species. *Journal of Biogeography* 39, 640–651.

Dolanský, T., 2004. Lidary a letecké laserové skenování.

Duputié, A., Zimmermann, N.E., Chuine, I., 2014. Where are the wild things? Why we need better data on species distribution. *Global Ecology and Biogeography* 23, 457–467.

Dyer, J.J., Brewer, S.K., Worthington, T.A., Bergey, E.A., 2013. The influence of coarse-scale environmental features on current and predicted future distributions of narrow-range endemic crayfish populations. *Freshwater Biology* 58, 1071–1088.

Elith, Catherine H. Graham, R.P.A., Dudík, M., Ferrier, M., Guisan, S., Hijmans, R.J., Huetman, F., Leathwick, J.R., Lehmann, A., Li, J., Lohman, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species’ distributions from occurrence data. *Ecography* 29, 129–151.

Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? on finding reasons for differing performances of species distribution models. *Ecography* 32, 66–77.

Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *The Annual Review of Ecology, Evolution, and Systematics* 40, 677–97.

Ferrier, S., Watson, G., 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. *Environment Australia, Canberra*.

- Ferrier, S., Watson, G., Pearce, J., Drielsma, M., 2002.** Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation* 11, 2275–2307.
- Gottschalk, T.K., Aue, B., Hotes, S., Ekschmitt, K., 2011.** Influence of grain size on species-habitat models. *Ecological Modelling* 222, 3403–3412.
- Graham, C.H., Ferrier, S., Huettman, F., Moritz, C., Peterson, A.T., 2004.** New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19.
- Guillera-Aroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Mccarthy, M.A., Tingley, R., Wintle, B.A., 2015.** Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 24, 276–292.
- Guisan, A., Graham, C.H., Elith, J., Huettmann, F., NCEAS Species Distribution Modelling Group, 2007.** Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13, 332–340.
- Guisan, A., Thuiller, W., 2005.** Predicting species distribution: Offering more than simple habitat models. *Ecology Letters* 8, 993–1009.
- Gutiérrez, D., Fernández, P., Seymour, A.S., Jordano, D., 2005.** Habitat Distribution Models: Are Mutualist Distributions Good Predictors of Their Associates? *Ecological Applications* 15, 3–18.
- Hernandez, P.A., Graham, C., Master, L.L., Albert, D.L., 2006.** The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773–785.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005.** Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978.
- Hirzel, A.H., Helfer, V., Metral, F., 2001.** Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145, 111–121.
- Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morissette, J.T., Holcombe, T.R., 2015.** Caveats for correlative species distribution modeling. *Ecological Informatics* 29, 6–15.
- KRNAP, 2017.** Přírodní poměry. *Cit. 1.11. 2017, online:* <http://www.krnep.cz/prirodni-pomery/>
- Jetz, W., McPherson, J.M., Guralnick, R.P., 2011.** Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology and Evolution* 27.
- Leroy, B., 2015.** The virtualspecies R package: a complete tutorial. *Cit. 2.11. 2017, online:* <http://borisleroy.com/files/virtualspecies-tutorial.html>

McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41, 811–823.

Meynard, C.N., Kaplan, D.M., 2013. Using virtual species to study species distributions and model performance. *Journal of Biogeography* 40, 1–8.

Miller, J., 2010. Species distribution modeling. *Geography Compass* 4, 490–509.

Miller, J., 2014. Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Progress in Physical Geography* 38, 117–128.

Miller, J., Holloway, P., 2015. Incorporating movement in species distribution models. *Progress in Physical Geography* 39, 837–849.

Mitchell, P.J., Monk, J., Laurenson, L., 2017. Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution* 8, 12–21.

Moudrý, V., 2015. Modelling species distributions with simulated virtual species, *Journal of Biogeography*.

Nezer, O., Tomer, S.B.-D., Gueta, T., Carmel, Y., 2017. High-resolution species-distribution model based on systematic sampling and indirect observations. *Biodiversity and Conservation* 26, 421–437.

Parra, J.L., Graham, C.C., Freile, J.F., 2004. Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography* 27, 350–360.

Pearson, R.G., 2010. Species' distribution modeling for conservation educators and practitioners. *Lessons in Conservation* 3, 54–89.

Pradervand, J., Dubuis, A., Guisan, A., Randin, C., 2013. Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography* 38, 79–96.

Prosdij, A.S.J. Van, Sosef, M.S.M., Wieringa, J.J., Raes, N., 2015. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39, 542–552.

Sheeren, D., Bonthoux, S., Balent, G., 2014. Modeling bird communities using unclassified remote sensing imagery: Effects of the spatial resolution and data period. *Ecological Indicators* 43, 69–82.

Song, W., Kim, E., Lee, D., Lee, M., Jeon, S.W., 2013. The sensitivity of species distribution modeling to scale differences. *Ecological Modelling* 248, 113–118.

Tobalske, C., 2002. Effects of spatial scale on the predictive ability of habitat models for the green woodpecker in Switzerland. *Predicting species occurrences: issues of accuracy and scale*, pp. 197–204.

van Gils, H., Conti, F., Ciaschetti, G., Westinga, E., 2012. Fine resolution distribution modelling of endemics in Majella National Park, Central Italy. *Plant Biosystems* 146, 276–287.

Varela, S., Anderson, R.P., García-valdés, R., Fernández-gonzález, F., 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37, 1084–1091.

Wan, J.Z., Wang, C.J., Yu, F.H., 2016. Impacts of the spatial scale of climate data on the modeled distribution probabilities of invasive tree species throughout the world. *Ecological Informatics* 36, 42–49.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Zimmermann, N.E., 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14, 763–773.

Xu, W., Hays, B., Fayrer-Hosken, R., Presotto, A., 2016. Modeling the distribution of African savanna elephants in Kruger National Park: an application of multi-scale Globeland30 data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B8*, 1327–1334.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modeling using MAXENT: When can we trust the inferences? *Methods in Ecology and Evolution* 4, 236–243.