

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE



Fakulta životního
prostředí

KATEDRA APLIKOVANÉ GEOINFORMATIKY A ÚZEMNÍHO
PLÁNOVÁNÍ

VLIV NEPŘESNÉHO URČENÍ VÝSKYTU DRUHŮ NA KVALITU
MODELŮ DRUHOVÉ DISTRIBUCE

EFFECT OF SPECIES POSITIONAL ERRORS ON THE
PERFORMANCE OF SPECIES DISTRIBUTION MODELS

Diplomová práce

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

Diplomant: Bc. Michaela Patáková

2017

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Michaela Patáková

Aplikovaná ekologie

Název práce

Vliv nepřesného určení výskytu druhů na kvalitu modelů druhové distribuce

Název anglicky

Effect of species positional errors on the performance of species distribution models

Cíle práce

Cílem práce je posoudit vliv nepřesného určení výskytu druhů na kvalitu modelů druhové distribuce.

- Generování virtuálního druhu
- Vytvoření modelů druhové distribuce s daty různé přesnosti výskytu druhu
- 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 nepřesného určení výskytu druhů na kvalitu modelu.

Doporučený rozsah práce

40 – 50 stran

Klíčová slova

Formulace klíčových slov je úkolem autora.

Doporučené zdroje informací

- GUILLERA-ARROITA G., LAHOZ-MONFORT J. J., ASCELIN GORDON J. E., KUJALA H., LENTINI P. E., MCCARTHY M. A., TINGLEY R., WITNLE A., 2015: Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 24: 276 – 292.
- 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.
- MOUDRÝ V., 2015: Modelling species distributions with simulated virtual species. *Journal of Biogeography* 42: 1365 – 1366.
- POCHEVILLE A., 2015: The Ecological Niche: History and Recent Controversies. In: HEAMS TH., HUNEMAN P. LECOINTRE G., SIBERSTEIN M. (Eds.): *Handbook of Evolutionary Thinking in the Science*. Springer, Netherlands: 547 – 586.
- SOBERÓN J., NAKAMURA M., 2009: Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19644 – 1965.

Předběžný termín obhajoby

2017/18 ZS – FŽP

Vedoucí práce

Ing. Vítězslav Moudrý, Ph.D.

Garantující pracoviště

Katedra aplikované geoinformatiky a územního plánování

Elektronicky schváleno dne 8. 11. 2017

doc. Ing. Petra Šímová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 8. 11. 2017

prof. RNDr. Vladimír Bejček, CSc.

Děkan

V Praze dne 20. 11. 2017

Prohlášení

Prohlašuji, že jsem tuto práci na téma „Vliv nepřesného určení výskytu druhů na kvalitu modelů druhové distribuce“ vypracovala samostatně a všechny zdroje v této práci uvedené jsou řádně ocitovány a uvedeny v příloženém seznamu použité literatury.

Ve Lhotě dne 14. 12. 2017
Patáková Michaela

.....

Poděkování

Ráda bych poděkovala Ing. Vítězslavu Moudrému, Ph.D. za odborné vedení, velkou trpělivost a užitečné připomínky, které mi umožnily dokončit tuto práci. Dále bych chtěla poděkovat Ing. Kryštofu Dobrovolnému za odbornou pomoc při technologickém zpracování dat. V neposlední řadě děkuji své rodině za podporu v průběhu celého studia.

Abstrakt

Modelování druhové distribuce je efektivní nástroj používaný v různých přírodovědných disciplínách, jako je konzervační biologie, ekologie či biogeografie. Rozvoj technologií v oblasti DPZ a GPS umožňuje získat poměrně podrobná a přesná vstupní data, jejichž kvalita představuje základní limitující faktor ovlivňující prediktivní schopnost modelu. Pokud je předmětem zkoumání reakce modelu za určitých podmínek, je vhodné namísto reálných dat zatížených jistou mírou nahodilosti definovat virtuální druh s předem známými reakcemi na prostředí.

V této diplomové práci byl zkoumán vliv chybné lokalizace druhu při použití environmentálních proměnných s velmi vysokým rozlišením. Jako environmentální proměnné byly použity rastry nadmořské výšky a vegetace na území KRNAP s rozlišením buňky 1m^2 a dále byly odvozeny rastry sklonu a orientace vůči světovým stranám. Data o výskytu druhu byla získána definováním virtuálního druhu a jeho konverzí na prezenčně – absenční data pravděpodobnostní metodou. Chyba v lokalizaci druhu byla simulována ve třech krocích od 0 do 1, do 10 a do 50 m. Pomocí generalizovaného lineárního modelu byly vytvořeny modely druhové distribuce pro varianty s různým stupněm chyby v lokalizaci druhu. Efekt posunu byl nejprve testován na vstupních datech, v dalším kroku byla pomocí validačních metrik posuzována predikční schopnost modelu pro jednotlivé varianty.

Modely dosahovaly hodnot AUC v rozmezí od 0.99 u dat bez posunu do 0.82 u maximální chyby 50 m v lokalizaci druhu. Ačkoliv srovnání modelů pomocí AUC ukazovalo na velmi vysokou diskriminační schopnost všech modelů, podíl vysvětlené variability, zpětně odvozené rastry pravděpodobnosti výskytu druhu, hodnota Kappa a shoda mezi modely testovaná Pearsonovým korelačním koeficientem prokázaly, že schopnost modelů predikovat druh na základě chybně lokalizovaných dat značně klesá. Navíc, chybná lokalizace druhu ve velmi heterogenním prostředí v tomto případě ovlivnila podobu křivky, která popisuje reakci druhu na environmentální podmínky. Problematika chybné lokalizace je však značně rozsáhlá a zasloužila by další bádání např. v oblasti vlivu chyby v prostředí s různým stupněm heterogenity.

Klíčová slova:

AUC, chyby v datech, Pearsonův korelační koeficient, sensitivita, specificita, nepřesná lokalizace, GLM, modelování druhové distribuce, virtuální druh

Abstract

Species distribution modelling is an effective tool in various science disciplines such as conservation biology, ecology or biogeography. Development of technology on the field remote sensing and global positional systems allows us to obtain relatively detailed and accurate input data, which quality is the basic limiting factor affecting the model predictive ability. In the case that the object of study is model reaction under certain conditions, is better use data from virtual species observation defined with known reaction on environmental variables, than real species data, which are loaded with a certain degree of chance.

In this thesis was investigated effect of error in localization of species using environmental variables represented by raster with very high resolution. Environmental variables used in this experiment were height of vegetation and elevation represented by raster with cell size 1m² then were derived raster of slope and aspect from elevation raster. Data about species occurrence was obtained by defining virtual species, than the raster of environment suitability was converted into presence-absence data by probabilistic approach. Error in localization was simulated in the three steps from 0 to 1, to 10 and to 50 m. Data was fitted by generalized linear model for three variant with different degree of error in localization. Both the input data and the prediction ability of the models were compared in the evaluation.

The models achieved AUC values ranging from 0.99 for data without shifting to 0.82 for the maximum error 50 meters in localization of species. Although the comparison of models by AUC showed a very high discrimination ability, retrospectively derived raster of environmental suitability for each step of error, Kappa, and the match between models tested by the Pearson's correlation coefficient showed that the ability of model to predict species distribution based on the shifted data drops considerably. In addition, error in localization of the species in a very heterogeneous environment in this case has influenced the shape of a curve describing the species' response to environmental conditions. However, the effect of error in localization is quite extensive and deserves further research. For example, effect of error localization in an environment varying by degree of heterogeneity.

Key words:

AUC, sampling error, error localization, Pearson's correlation coefficient, sensitivity, specificity, GLM, species distribution modeling, virtual species

Obsah

1. Úvod	13
2. Literární přehled	15
2.1. Proces modelování druhové distribuce	15
2.2. Ekologická nika	16
2.3. Biotické interakce a environmentální proměnné	17
2.4. Vstupní data	18
2.4.1. Data o výskytu druhu	19
2.4.2. Chyby v datech o výskytu druhu	21
2.4.3. Environmentální data	22
2.5. Vliv měřítka na SDM	23
2.6. Virtuální druh	24
2.7. Modelovací metody	26
2.7.1. GLM a logistická regrese	26
2.8. Evaluace modelů	27
3. Metodika	30
3.1. Charakteristika zájmového území	30
3.2. Charakteristika vstupních dat	31
3.3. Použitý software	31
3.4. Generování dat	32
3.4.1. Definování virtuálního druhu	32
3.4.2. Konverze dat na presenčně – absenční rastr	34
3.4.3. Vzorkování dat	36
3.5. Statistické vyhodnocení	37
4. Výsledky	38
4.1. Základní analýza environmentálních dat	38
4.2. AIC, vysvětlená variabilita a AUC	41
4.3. Metriky odvozené z chybové matice	42
4.4. Korelační koeficient	43
5. Diskuse	44
6. Závěr	48
7. Přehled literatury a použitých zdrojů	49
8. Přílohy	55
8.1. Grafické přílohy	55

8.2. Části kódu dle jednotlivých kroků v jazyku R.....	58
8.3. Tabulky	61

Seznam použitých zkratk

AIC	Akaikeho informační kritérium
AUC	plocha pod křivkou (Area under cover)
BRT	Boosted regression trees
CART	klasifikační a regresní stromy (Classification and regression trees)
ČUZK	Český úřad zeměměřičský a katastrální
D ²	vysvětlená variabilita (Deviance)
DET	data odhadu obsazenosti (Detection of occupancy)
DPZ	dálkový průzkum země
DTM	digitální model terénu
GAM	generalizovaný aditivní model
GBIF	Globální informační databáze biodiversity (Global Biodiversity Information Facility)
GDM	Generalised dissimilarity
GLM	generalizovaný lineární model
GPS	globální poziční systémy
IUCN	Mezinárodní svaz ochrany přírody (International Union for Conservation of Nature)
KRNAP	Krkonošský národní park
MARS	Multivariate regression splines
MAXENT	model maximální entropie (Maximum entropy model)
MNČ	metoda nejmenších čtverců
PA	presenčně-absenční data
PB	presenčně-pozadřová data (Presence background)
PO	čistě presenční data (Presence only)
PCC	poměr správně klasifikovaných výskytů (Percent correctly classified)
R ²	koeficient determinace
S-JTSK	systém jednotné trigonometrické sítě katastrální
SDM	Model druhové distribuce (Species distribution model)

Seznam obrázků

Obrázky zařazené v práci:

Obrázek 1: Schéma procesu modelování druhové distribuce.....	15
Obrázek 2: Grafické znázornění prostorů fundamentální, potenciální a realizované niky	17
Obrázek 3: Základní tvar křivky popisující vztah mezi fitness druhu a podmínkou prostředí.....	18
Obrázek 4: pravděpodobnostní konverze dle logistické funkce	25
Obrázek 5: Ortofotomapa KRNAP s vyznačením nejvyšších vrcholů v území	30
Obrázek 6: Výškový profil hřebenu KRNAP	31
Obrázek 7: Průběh funkcí vyjadřuje závislost pravděpodobnosti výskytu druhu na jednotlivých environmentálních složkách.....	33
Obrázek 8: A – detail „rastru vhodnosti prostředí“ rastru, B – vhodnost environmentálních podmínek pro nadefinovaný virtuální druh v celém území KRNAP	34
Obrázek 9: Křivka logistické funkce pro převod na presenčně absenční rastr z dat pravděpodobnosti výskytu,.....	35
Obrázek 10: A – detail presenčně absenčních dat, B – přehled presence virtuálního druhu na celém území KRNAP	36
Obrázek 11: Rozložení 400 presenčně – absenčních bodů na území KRNAP.....	37
Obrázek 12: Rozložení četnosti environmentálních proměnných dle jednotlivých variant posunu.....	38
Obrázek 13: Vliv posunu dat na výslednou podobu křivky popisující pravděpodobnost výskytu druhu v závislosti na environmentální proměnné	39
Obrázek 14: detail rastrů vhodnosti prostředí pro jednotlivé varianty	40
Obrázek 15: korelační matice	43

Obrázky zařazené v příloze:

Obrázek přílohy 1: Digitální model terénu KRNAP s nejvyššími vrcholy v území.....	55
Obrázek přílohy 2: 3D Model území KRNAP.....	55
Obrázek přílohy 3: Výška vegetace na území KRNAP	56
Obrázek přílohy 4: Sklonitost terénu ve stupních na území KRNAP	56
Obrázek přílohy 5: Detail rastru expozice terénu vůči světovým stranám	57
Obrázek přílohy 6: Detail posunu presenčně-absenčních dat.....	57
Obrázek přílohy 7: Definování virtuálního druhu.....	58
Obrázek přílohy 8: Konverze rastru vhodnosti prostředí na presenčně-absenční rastr	58
Obrázek přílohy 9: Vzorkování výskytu virtuálního druhu z presenčně-absenčního rastru a následné přiřazení hodnot nadmořské výšky, výšky vegetace, sklonu a expozice	58
Obrázek přílohy 10: Náhodný posun dat výskytu virtuálního druhu v rámci intervalů	58
Obrázek přílohy 11: Rozdělení dat o výskytu druhu na trénovací a testovací sadu.....	59
Obrázek přílohy 12: generalizovaný lineární model, výpočet vysvětlené variability, predikce vhodnosti prostředí na základě modelu, validace metodou AUC.....	59
Obrázek přílohy 13: zpětné odvození křivky popisující vztah mezi environmentální proměnnou a pravděpodobností výskytu v prostředí.....	60
Obrázek přílohy 14: chybová matice a odvození sensitivity, specificity a poměru správně klasifikovaných presencí a absencí (PCC)	60
Obrázek přílohy 15: Odvození rastrů vhodnosti prostředí dle koeficientů modelů, korelační matice, export odvozených rastrů do formátu ESRI	61

Seznam tabulek

Tabulky zařazené v textu:

Tabulka 1: Přehled zdrojů dat o výskytu druhu.....	20
Tabulka 2: Přehled zdrojů environmentálních dat	22
Tabulka 3: Přehled nedávných studií s uplatněním virtuálního druhu	24
Tabulka 4: Matice validačních technik	28
Tabulka 5: chybová matice	29
Tabulka 6: Základní parametry vstupních vrstev	31
Tabulka 7: Srovnání modelů dle AIC, vysvětlené variability a AUC	41
Tabulka 8: Metriky odvozené z chybové matice pro různé prahové hodnoty	42

Tabulky zařazené v příloze:

Tabulka přílohy 1: Vyhodnocení GLM bez přítomnosti kvadratického členu	61
Tabulka přílohy 2: Výsledky GLM s kvadratickým členem.....	62

1. Úvod

Porozumění vztahu mezi druhy a prostředím které obývají, je ve světě, jež podléhá globálním změnám, jako fragmentace biotopů, změna využití území či klimatické změny, klíčové nejen pro přežití druhu, ale i pro ochranu a stabilitu celého biotopu, jehož jsou součástí. Druhovému distribuční modely jsou považovány za základní nástroje v ekologii a konzervační biologii, které tyto vztahy matematicky popisují. Díky rozvoji geografických informačních systémů je možné matematicky vyjádřené vztahy závislosti výskytu na environmentálním prostředí převést na predikční mapy.

Nejvíce diskutované využití SDM v praxi je vliv klimatických změn na minulou či budoucí distribuci druhu a stanovení rezervace pro ochranné účely druhu (*Guisan et Thuiller, 2005*). Modely druhové distribuce byly v praxi dále použity pro sestavení map druhového bohatství (*Rissler et al., 2006*) nebo pro odhad chování potencionálně invazních druhů (*Václavík et Meentemeyer, 2012*).

Přesnost, s jakou dokáže model predikovat druh na určité lokalitě, a následnou interpretaci výsledků, ovlivňuje řada faktorů. Například *Guillera-Arroita et al. (2015)* se zabývala, jakým způsobem by měla být interpretována data o výskytu obsahující různou míru informace (presenční, nebo presenčně – absenční data. *Elith et al. (2006)* řešila, jaký modelovací algoritmus nejlépe fituje čistě presenční data. Častou zkoumaný parametr presenčně – absenčních dat je prevalence (poměr presence ku absenci). Vlivem prevalence na přesnost modelu se zabýval např. *Fukuda et Baets (2016)*. V závěru své studie uvádějí, že prevalence měla vliv jak na přesnost modelu, tak ovlivnila i křivku popisující vztah mezi druhem a prostředím. Vzácné druhy, které mají nižší prevalenci oproti druhům běžným, dosahují vyšší přesnosti v modelech druhové distribuce (*Jiménez-Valverde et al., 2008*). *Lawson et al. (2014)* popisuje, jak prevalence ovlivňuje následnou validaci modelu a zdůrazňuje, že pouze validační metriky závislé na prevalenci mohou spolehlivě posoudit kalibrační schopnosti modelu.

Rozvoj metod dálkového průzkumu země umožňuje pořízení vstupních dat ve vysoké kvalitě. V současné době je možné pořídit např. družicové snímky s velmi jemným rozlišením (QuickBird 0.6 m²/pixel, WorldView-2 0.5 m²/pixel) nebo získat data o využití území pomocí snímkování povrchu leteckými drony. S tím souvisí i potřeba přesnější lokalizace výskytu druhu. Jemnější rozlišení navíc umožňuje odhalit biotické interakce, které ve vysokém rozlišené ztrácí význam na úkor např. klimatických faktorů (*Thuiller et al., 2004*).

Z hlediska modelování druhové distribuce je zásadní přesnost, s jakou jsou jednotlivé výskyty zaznamenány. Vlivem chybné lokalizace druhu v terénu se již zabývala řada studií. Například *Osborne et Leitão (2009)* zkoumali tento fenomén na reálných datech. Do dat o výskytu ptačích druhů, která jsou obecně velmi málo přesná díky vysoké mobilitě těchto druhů, vnášeli náhodnou chybu v lokalizaci v rozmezí 1 – 5 km. V závěru své studie uvádějí, že schopnost modelu predikovat druh v určitém místě klesá se zvyšující se chybou v prostorové lokalizaci, avšak větší vliv na snížení přesnosti modelu měla vysoká variabilita v datech. *Johnson et Gillingham (2008)* ve své studii také použili reálná data vysoce mobilního druhu. Do presenčních dat o výskytu vnášeli chybu v jednotlivých krocích od 50 do 1000 m. Oproti studii *Osborne et Leitão* však došli k závěru, že chyba v lokalizaci druhu je hlavní příčinou snížení predikční schopnosti modelu. *Graham et al. (2008)* testovali schopnost modelu predikovat výskyt pro několik různých algoritmů. Ve své studii uvádí, že nejcitlivější vůči chybě v lokalizaci druhu jsou metody strojového učení MaxEnt a BRT, naopak metody založené na regresi dokáží dobře předikovat druh i při chybě v lokalizaci 50 krát větší než je rozlišení vstupních rastrů. *Duputié et al. (2014)* se zabýval kvalitou

dat o výskytu druhu odvozených z atlasů rozšíření evropských stromů. Porovnával čtyři zdroje dat několika modelovacími technikami. Výsledkem bylo, že modely vytvořené na základě těchto zdrojů se liší až o 74%. V závěru své práce proto doporučuje kombinovat více zdrojů vstupních dat. Výše zmíněné studie však využívaly jako vstupní data rastry s rozlišením v řádech několika desítek metrů až kilometrů. Teprve nedávná studie *Mitchella et al. (2017)* se zabývala vlivem nepřesné lokalizace při vysokém rozlišení vstupních dat na druhové distribuční modely, avšak vstupem byla data o výskytu reálného druhu.

Vzhledem ke skutečnosti, že v přírodě působí vysoká míra náhody, způsob sběru dat o výskytu druhu je zejména v případě presenčně-absenčních dat velmi náročný a jak bude uvedeno dále ovlivněn do jisté míry osobou, která data pořizuje, metodikou sběru a v neposlední řadě nelze s jistotou předpokládat v jaké fázi „nasycení“ se druh v daném prostředí nachází, nelze některé vlastnosti modelů druhové distribuce nestranně hodnotit. Data o výskytu virtuálních druhů tak v současné době představují jedinou možnost, jak tyto vlastnosti porovnávat (*Austin et al., 2006*). Hlavní výhodou virtuálního druhu je známá virtuální realita vycházející z předem definovaných vztahů mezi druhem a prostředím (*Miller, 2014*).

Cílem této práce je: (1) generování virtuálního druhu na základě environmentálních proměnných s velmi vysokým rozlišením; (2) simulovat presenčně-absenční sběr dat v terénu s náhodnou chybou v lokalizaci druhu; (3) porovnat, jaký vliv má velikost náhodné chyby v lokalizaci druhu na změnu jeho nároků vůči prostředí (4); posoudit přesnost modelu z hlediska pravděpodobnosti výskytu (5); a následně schopnosti rozpoznat presenci či absenci v daném místě.

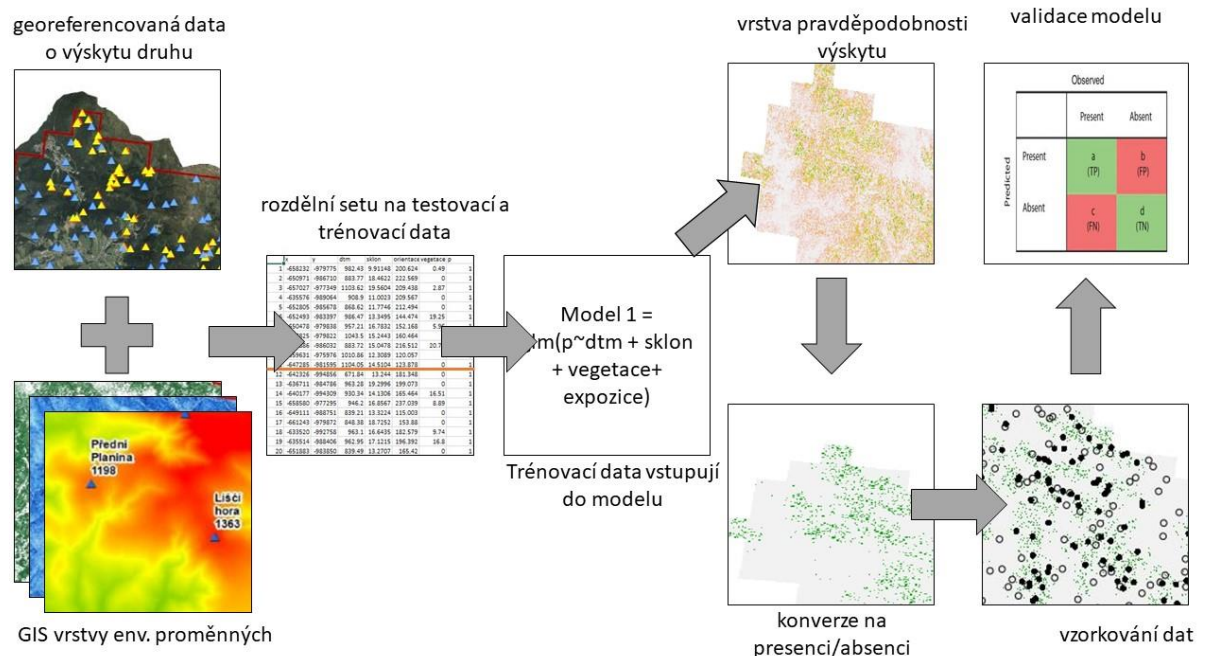
2. Literární přehled

2.1. Proces modelování druhové distribuce

Modelování druhové distribuce slouží k popisu vztahů mezi prostředím a výskytem rostlin či živočichů a k predikci jejich výskytu. Většina přístupů vychází z kvantifikování vztahu mezi environmentálními faktory prostředí a reakcí druhu na něj. Modelování druhové distribuce (Species Distribution Modelling) se vyvinulo z prostého vyjádření tohoto vztahu mimo prostorové určení do sofistikované metody využívající pokročilé GIS a statistické nástroje (Guisan et Thuiller, 2005). Právě rozvoj geografických informačních systémů a digitalizace vstupních dat urychlili vývoj v oblasti SDM. Výsledkem je široká variabilita metod aplikujících podobné přístupy, využívající různou kvalitu vstupních dat s více či méně rozdílnými výstupy v podobě mapy rozšíření druhu (Miller, 2010).

Guisan et Zimmerman (2000) uvádějí, že proces SDM by měl ideálně zahrnovat tyto kroky: promyšlení konceptu celé studie, příprava dat, modelování a následná validace modelu, predikce druhu v zájmovém území a posouzení aplikovatelnosti modelu. Znalost ekologické teorie by měla být zahrnuta v každé fázi procesu SDM (Austin, 2002), např. při volbě environmentálních proměnných, které nejvíce ovlivňují druh, nebo při hledání ekologicky relevantní křivky popisující vztah druhu a environmentální proměnné (Guisan et Thuiller, 2005).

Presenčně – absenční modely jsou založené na pozorovaném výskytu druhu v prostoru. Ke každému pozorování lze přiřadit environmentální podmínky v místě výskytu či absence. SDM pak využívá tohoto vztahu k predikování pravděpodobnosti, s jakou se bude druh v jednotlivých buňkách vyskytovat. Validací metriky pak zkoumají, jakou podobnost dosahují predikované výskytu na základě modelu s původními daty (Lawson et al., 2014).



Obrázek 1: Schéma procesu modelování druhové distribuce

(autor, 2017)

Je vhodné si uvědomit, že biologická i ekologická data jsou sbírána po určitý časový úsek. Modely založené na těchto datech tak odrážejí očekávané vztahy mezi druhem a prostředím pouze dočasně. Lze vycházet z předpokladu, že druh je s prostředím, které obývá v tzv. „pseudo-rovnováze“ (Guisan et al., 1998). Existuje tedy riziko, že model, který bude až příliš dobře predikovat reálná data, bude podhodnocený z hlediska potencionálního areálu rozšíření (Svenning et Skov, 2004). To platí např. v případě modelování druhové distribuce invazivních druhů. Většina invazivních druhů totiž není ve vztahu k prostředí, do něhož proniká, v rovnováze a proto by modelování jejich distribuce mělo probíhat v areálu jejich přirozeného rozšíření (Peterson, 2003).

2.2. Ekologická nika

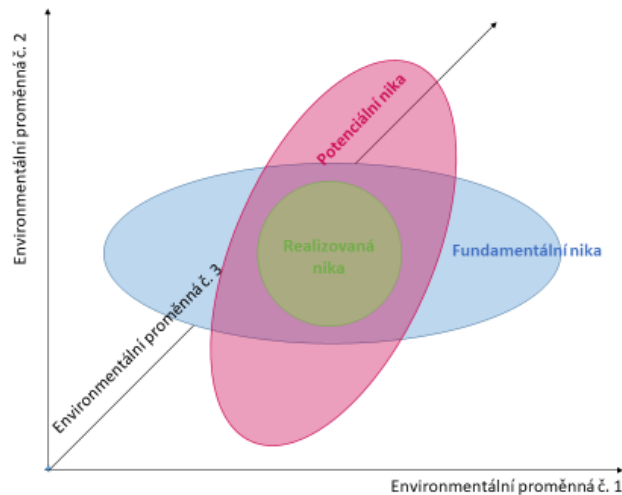
Modelování druhové distribuce patří mezi skupinu metod (podobně jako modelování potencionálního habitatu či modelování potencionální niky), které si kladou podobný úkol. Identifikovat místa vhodná pro přežití populace na základě předem známých nároků na prostředí daného druhu a právě tento vztah mezi organismem a podmínkami prostředí popisuje ekologická nika (Soberón et Nakamura, 2009). Existuje mnoho konceptů, jak ekologickou niku definovat.

První vědecká definice pojmu „ekologická nika“ pochází od Josepha Grinnella z roku 1913. Grinnell definoval niku jako soubor všech podmínek, které ovlivňují existenci druhu v daném místě. Mezi podmínky řadí jak abiotické faktory: teplota, vlhkost, úhrn srážek, ...) tak i biotické (dostupná potrava, interakce s ostatními organismy, ...). Dále je nutné podotknout, že Grinnellova definice ekologické niky částečně vychází z Darwinovy teorie O původu druhů, kdy koexistence dvou druhů v jedné nise vede k jejich vzájemnému ovlivnění. Naopak Charles Elton definoval ekologickou niku organismu z hlediska jeho postavení v potravním řetězci. Nika je soubor rolí a funkcí, kterou populace zastává vůči organismům, s nimiž přichází do kontaktu. Na rozdíl od Grinnella, Elton tvrdí, že různé organismy mohou obývat stejné potravní niky. V roce 1957 přišel Gorge E. Hutchinson s obráceným pojetím. Ekologickou niku chápe jako soubor nároků živočichů, nikoliv podmínek prostředí. Nika je pak tvořena n-rozměrným prostorem, z nichž každý rozměr tvoří právě jeden z nároků dané populace (Pocheville, 2015).

Také velikost plochy a rozlišení v jakém je druhová distribuce zkoumána určuje, jaký koncept je pro daný výzkum vhodnější. Grinnellův koncept je určený pro měřítka s nízkým rozlišením, protože nika je definována především základními podmínkami prostředí (teplota, nadmořská výška, úhrn srážek,...), jež jsou při malých měřítcích a vysokém rozlišení málo proměnlivé. Ekologická nika v pojetí Eltona je naopak založena na biologických interakcích mezi organismy, které jsou detekovány lokálně. Z toho jasně vyplývá, že tento koncept je určen pro výzkumy, kde se distribuce druhů řeší ve vysokém rozlišení (Soberón, 2007).

Hutchinson dále rozlišuje niku *fundamentální*, což je takový n-rozměrný prostor, který by populace obývala v případě, že by nebyl zohledněn vliv konkurentů, a niku *realizovanou*, která vzniká, odečte-li se od fundamentální niky prostor, ze kterého populaci vylučují konkurenti (Pocheville, 2015). Existuje však možnost, že některé kombinace nároků na prostředí v daném čase na Zemi neexistují. Proto byla definována *potencionální* nika, která představuje průnik niky fundamentální s existujícími kombinacemi nároků na prostředí v určitém čase (Jackson et Overpeck, 2000).

Hutchinsonovo dělení se projevuje ve vztahu vstupních dat a jejich interpretaci v SDM. Pokud do modelu vstupují reálná data, jsou zatížena chybou. Z tohoto důvodu je možné dosáhnout reálné velikosti fundamentální niky pouze s daty virtuálních druhů, nebo experimentálně odvozenými daty o chování a fyziologických vlastnostech zkoumaného druhu (Guisan et al., 2005)



Obrázek 2: Grafické znázornění prostorů fundamentální, potenciální a realizované niky

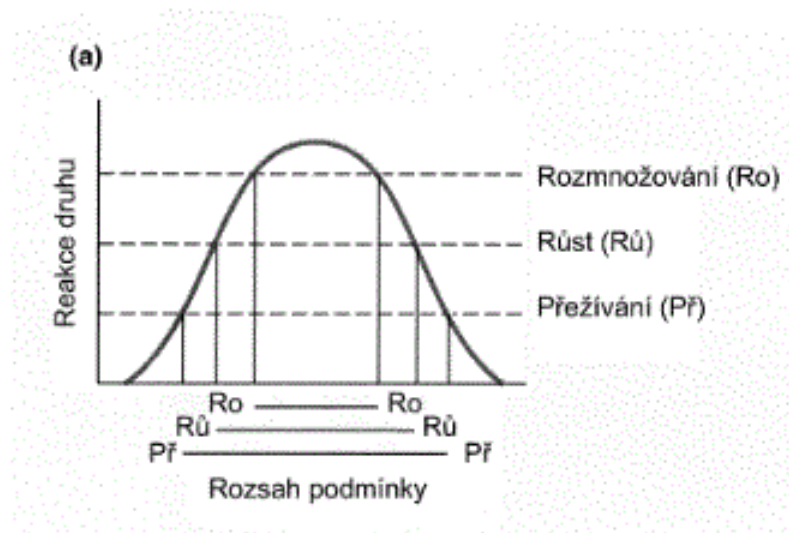
(autor, 2017).

Ačkoliv je Hutchinsonův koncept ekologické niky v současnosti nejvíce využívaným v souvislosti s modelováním druhové distribuce, existují výtky, že Hutchinson zohledňuje pouze negativní interakce, zatímco ty pozitivní jako mutualismus či symbióza nejsou zahrnuty, byť by mohly způsobit zvětšení prostoru fundamentální niky (Miller, 2010).

2.3. Biotické interakce a environmentální proměnné

Begon *et al.* (1990) definuje základní vztah mezi fitness druhu (tzv. biologickou zdatností) a environmentální proměnou (někdy označovanou jako environmentální gradient) jako symetrickou křivku zvonovitého tvaru, avšak za předpokladu, že tento vztah nebude ovlivněn např. konkurencí. Austin (1999) dále předpokládá, že tento průběh křivky platí jak pro fundamentální, tak i realizovanou niku avšak velmi silný konkurent dokáže vytlačit druh z jeho optima. Následkem toho pak dochází k deformaci křivky realizované niky od jemného zkosení po bimodální průběh.

Některé studie, zabývající se vlivem biotických interakcí ukázaly, že přidáním presenčně – absenčních dat známých konkurentů daného druhu jako další příčinná proměnná může výrazně posílit přesnost modelu. Nevýhodou však může být, že se zmenšujícím se rozlišením (např. na úrovni regionů či kontinentů) je vliv tohoto parametru zanedbatelný (Guisan *et al.*, 2005).



Obrázek 3: Základní tvar křivky popisující vztah mezi fitness druhu a podmínkou prostředí

(Begon et al., 1990)

Environmentální gradient lze chápat jako jeden rozměr v n-rozměrné nícě definované Hutchinsonem. Jedná se o vztah mezi výskytem druhu podél gradientu environmentální proměnné, který definuje distribuci druhu v environmentálním prostoru (Miller, 2010).

Existuje několik málo odlišných způsobů, které environmentální gradienty řadí do kategorií. Guisan et al. (2005) dělí gradienty na limitující faktory, disturbance (rušivé události), jež zahrnují i lidský faktor a zdroje. Avšak nejčastěji používané kategorie dle Austina (1980) dělí environmentální gradienty dle vlivu, který na daný druh mají na přímé, nepřímé a zdroje a řadí je dle hierarchie, kterou v procesu výsledného vlivu na daný druh zaujímají od blízkých po okrajové. Přímé gradienty jsou ty, které působí na druh z hlediska jeho fyziologických vlastností a mohou tak ovlivnit růst populace (např. teplota či pH půdy). Mezi nepřímé gradienty patří ty, kde environmentální proměnná nemá žádný fyziologický vliv na druh či biologické interakce s jiným druhem (např. nadmořská výška). Definice zdrojových gradientů je založena na jejich vyčerpitelnosti vzhledem k tomu, že jsou organismy konzumovány a stávají se tak hlavní příčinou mezidruhových interakcí (např. kořist pro lovce nebo obsah fosforu v půdě).

Z hlediska řazení environmentálních gradientů dle jejich vlivu platí, že všechny nepřímé gradienty mají v hierarchii vlivů určující distribuci v environmentálním prostoru spíše okrajovou. Modely založené na blízkých gradientech jsou velmi robustní a široce aplikovatelné, avšak v praxi jsou velmi špatně použitelné vzhledem k velkým nárokům na vstupní environmentální proměnné (Austin, 2002).

2.4. Vstupní data

Pro proces modelování druhové distribuce jsou klíčové dva typy vstupních parametrů. Biologická data, která popisují výskyt druhu v prostoru a environmentální data popisující prostředí, ve kterém byl jedinec příslušného druhu detekován.

Pro znázorňování jevu např. v mapách jsou obecně přijímány dvě metody vyobrazení. Metoda diskretních objektů jev klasifikuje do několika oddělených kategorií. Pro tyto objekty znázorněné bodem, linií či plochou platí, že se mohou přesouvat či překrývat, aniž by pozbyly svou integritu. Tento

typ dat je charakteristický pro klasifikaci ploch využití území, třídy ochrany půd, typy biotopů či fenoménu, jímž se tato práce zabývá, tedy výskytu organismu. Naopak některé jevy, např. geomorfologické, u nichž je proměnlivost tak vysoká, že není možné roztřídit je do přehledného množství kategorií, je vhodné vyjádřit metodou kontinuálních polí. Jev jako je např. nadmořská výška či průměrná teplota je zaznamenán určitou hodnotou v každé buňce rastru o daném rozsahu a rozlišení (Goodchild, 2010).

Data lze kategorizovat ve vztahu k nástrojům geografických informačních systémů vektorová, která se dále dělí na body. Jako příklad lze uvést data o výskytu druhů, či informace o pozici meteorologických stanic. Dalším typem vektorových dat jsou polygony, jež mohou reprezentovat některé environmentální proměnné s kategoriální stupnicí (např. půdní typy) či linie, které mohou znázorňovat např. vodní toky, nebo komunikace. Do druhé kategorie patří rastry, síť buněk, z nichž každá nese hodnotu daného gradientu (většina environmentálních proměnných) (Pearson, 2010). Pro potřeby SDM lze rozdělit vstupní data na informaci o výskytu druhu, která je nejčastěji reprezentována body, představujícími výskyt či absenci druhu, popřípadě jejich četnost v dané lokalitě, ale může být reprezentována i polygony navrženými odborníky (Fourcade, 2016), opět následně převedené na bodovou vrstvu. Environmentální data poskytující informaci o prostředí, ve kterém druh žije, vstupují do modelu nejčastěji v rastrové formě.

2.4.1. Data o výskytu druhu

Data o výskytu druhu zachycují rozložení druhu v geografickém prostoru. Mohou být založena například na náhodném (např. nálezové databáze GBIF, ebird, data z GPS lokátorů), či pravidelně rozmístěném sběru, nebo náhodně získaných datech z muzejních sbírek (Graham et al., 2004).

Některé studie potvrzují snahy o využití dat z muzejních sbírek (Elith et al., 2006) a nálezových databází (Deputié et al., 2014) jež jsou v poslední době stále více dostupná online. Autoři těchto studií porovnávají prediktivní schopnost různých modelů založených na těchto datech a v závěru své práce upozorňují na limity, které sebou predikce založená na takovýchto datech nese. Jedním z problémů těchto dat je skutečnost, že jsou pouze presenčního rázu, avšak velké množství modelů využívá presenčně absenční data. Další nevýhodou je, že jejich sběr je silně ovlivněn např. atraktivitou určité lokality, časovou epochou, ve které byla data pořízena či nepřesnou lokalizací. Tato data tak mohou být využita pouze v některých modelech jako je BIOCLIM, DOMAIN, LIVES či MaxEnt které jsou pro použití presenčních dat navrženy, za předpokladu jasné interpretace výsledku a zdůraznění omezení, jež použití těchto dat přináší (Elith et al., 2006).

Fourcade (2016) dále navrhuje v případě absence kvalitních dat o výskytu využití databáze Red List Spatial Data, která je tvořena hrubými polygony reprezentujícími areál rozšíření druhu, který byl navržený odborníky z IUCN. Z těchto polygonů je posléze možné odvozením získat pseudo-presenční data o výskytu druhu. Fourcade takto porovnával data získaná z veřejných databází biodiverzity (GBIF a iNaturalist) s presenčními daty získanými z IUCN a došel k závěru, že modely založené na těchto datech jsou stejně kvalitní jako ty založené na veřejných databázích. Pouze v případě vysoce mobilních organismů (např. ptáci) se predikční schopnost modelů lišila. Zdroje dat o výskytu druhu uvádí následující tabulka č. 1.

Tabulka 1: Přehled zdrojů dat o výskytu druhu

organizace	dostupné online	popis dat
Global Biodiversity Information Facility	https://www.gbif.org/species/search	data o výskyt druhů jednotlivých druhů, či
Delivering Alien Invasive Species Inventories for Europe	http://www.europe-alien.org/default.do	data o výskytu invazivních druhů na Evropském kontinentu
eBird	http://ebird.org/ebird/explore	data o výskytu ptáků po celém světě
International Union for Conservation of Nature	http://www.iucnredlist.org/	areály výskytu ohrožených druhů ve formátu shapefile
BioLib	https://mapovani.biolib.cz/	mapování výskytu druhů na území České republiky

(autor, 2017)

Z hlediska záznamu výskytů může být distribuce druhu měřena několika způsoby. V případě, že hodnotou 1 označíme výskyt a hodnotou 0 absenci se jedná o nominální způsob. Pokud se bude na dané lokalitě vyskytovat druh opakovaně, můžeme lokalitě přiřadit váhu dle počtu výskytů a získáme tak data ordinálního typu, která nesou informaci o abundanci. Posledním typem jsou data poměrová, kdy lze výskyt druhu porovnávat vůči např. výskytu jiného druhu čímž lze získat informaci o druhovém složení na lokalitě (Miller, 2010).

Guillera-Arroita et al. (2015) dělí data dle způsobu sběru a množství informací, která obsahují od čistě presenčních dat po tzv. odhad obsazenosti. Čistě presenční data (PO) představují soubor bodů, kde byl druh detekován, aniž by bylo bráno v potaz širší okolí. „Presenčně-pozadová“ data (PB) jsou odvozena z lokalit, kde byl druh pozorován a doplněna o širší okolí, kde by se dle zjištěných nároků z pozorování měl druh vyskytovat. Nejpoužívanějším typem dat jsou presenčně-absenční data (PA). Výhodou těchto dat je, že kromě informace kde byl druh s jistotou detekován, nesou i informaci, kde se druh určitě vyskytovat nebude. Nejkomplexnější informaci o výskytu a distribuci druhu poskytují data o odhadu obsazenosti (DET). Tato data se od předchozích liší tím, že obsahují informaci o četnosti výskytu za danou dobu a tím pádem i pravděpodobnosti výskytu druhu na lokalitě. Nevýhodou těchto dat je však vysoká náročnost sběru.

Důležitý parametr, který ovlivňuje přesnost modelování druhové distribuce je velikost vzorku (počet pozorování). Vaughan a Ormerod (2003) uvádějí, že na každou vysvětlující proměnou by mělo připadnout alespoň 10 pozorování. Při použití krokové lineární regrese se doporučení pohybují v rozmezí od 20 pozorování po 40 pozorování ku 1 vysvětlující proměnné (Franklin, 2010). Wisz et al. (2008) ve svém výzkumu uvádí, že žádný z testovaných algoritmů nebyl spolehlivý s velikostí vzorku menší než 30 pozorování. Nejvíce náchylné k velikosti vzorku byly modely GLM, GAM a MARSINT, jež zahrnují interakce mezi prediktory, naopak jako nejvíce robustní vůči malému počtu vzorků se jevil model MAXENT. Obecně lze říci, že se zvyšující se velikostí vzorku a snižující se variabilitou napříč pozorování se zvyšuje přesnost modelu (Wisz et al., 2008).

V případě použití presenčně-absenčních dat dále ovlivňuje přesnost modelu prevalence, tedy poměr presenčních ku absenčním bodům. *McPearson et al. (2004)* ve své studii uvádí, že k nejvyšší přesnosti modelů docházelo při hodnotě prevalence blízké 0.5. V některých případech však vyrovnané prevalence nelze dosáhnout (u vzácných druhů či specialistů). V rozporu s tímto tvrzením jsou výsledkem studií, které se zabývaly modelováním druhové distribuce vzácných druhů, či druhů s úzkou ekologickou valencí, distribuční modely těchto druhů dosahovaly vyšší přesnosti, než v případě běžných druhů (*McPearsn et Jetz, 2007; Tsoar et al., 2007; Franklin et al., 2009*).

2.4.2. Chyby v datech o výskytu druhu

Pravděpodobnost detekování druhu na dané lokalitě lze rozložit na tři dílčí složky. První z nich je pravděpodobnost, že druh na dané lokalitě skutečně žije, dále pravděpodobnost, že lokalita bude zahrnuta do průzkumu, třetím komponentem je pravděpodobnost navštívení lokality právě v okamžiku, kdy se živočich (v případě rostlin lze zaměnit za vegetační období, kdy je rostlina detekovatelná) na lokalitě nalézá (*Yackulic et al., 2013*). Pravděpodobnost s jakou se druh na dané lokalitě nachází, je nejčastěji řešená otázka ve výzkumech modelování druhové distribuce. Pravděpodobnost navštívení konkrétní lokality nejvíce podléhá ovlivnění systematickou chybou, jež bude vysvětlena v dalších odstavcích. Třetí komponenta, tedy pravděpodobnost že druh bude v daný čas na konkrétní lokalitě detekován, pak nejvíce ovlivňují environmentální proměnné (*Guillera-Arroita et al., 2015*).

Existuje několik způsobů, které mohou reálná data zatížit chybou. Významným faktorem je detekovatelnost druhu. Druh se může na lokalitě vyskytovat, avšak nebude zaznamenán. U živočišných druhů může špatná detekovatelnost souviset s kryptickým zbarvením nebo vysokou mobilitou (*Franklin, 2010*). U rostlin jsou např. obtížně detekovatelné přisedlé druhy (*Chen et al., 2013*). Dalšími faktory mohou být nízká populační hustota či omezené úsilí na sběr vzorků (*Gu et Swihart, 2004*). Pokud je chybná detekce zároveň funkce environmentální proměnné, může zásadním způsobem ovlivnit tvar výsledného předpokládaného areálu rozšíření. PA a PB data jsou touto chybou ovlivněna nejvíce (*Guillera-Arroita et al., 2015*).

Data o výskytu druhu jsou nejčastěji zaznamenávána pomocí přístrojů GPS. Ty podle svých možností zaznamenávají přesnost s určitou odchylkou od místa výskytu (cca 0.5 až několik desítek metrů) (*Guillera-Arroita et al., 2013*). Jak moc se chyba projeví, závisí na heterogenitě prostředí (*Johnson et Gillingham, 2008*). Se zaznamenáváním dat o výskytu do GPS mohou být dále spojeny další chyby jako je špatné nastavení přístroje GPS v datu a čase, což vede ke zkreslení tvaru Země a následné ztrátě orientace v souřadnicové síti. U dat získaných z muzejních sbírek dále dochází ke zkreslení při jejich zpětné georeferenci ze slovního popisu místa nálezů (*Graham et al., 2008*). Stejně tak u mapování výskytu zpěvavých ptáků nestačí pouhý zvukový záznam a přibližné určení místa odkud pochází (*Osborne et Leitão, 2009*).

Vlivem nepřesnosti biologických dat se zabývá řada studií. Např. *Johnson a Gillingham (2008)* ve své práci tvrdí, že chybně určená pozice organismu má na přesnost modelu mnohem větší dopad než systematická chyba, či špatně klasifikované environmentální proměnné. Model MAXENT byl vůči těmto chybám nejvíce citlivý a to jak v případě studie *Graham et al. (2008)* tak i v případě srovnání modelovacích metod využívajících čistě presenčních nebo presenčně-pozadových dat (*Elith et al, 2006*).

Zkreslení biologických dat může být dále způsobeno systematickou chybou. Ta je dána vychýlením dat dle určitého trendu. Systematická chyba může vzniknout při sběru dat v terénu, kdy jsou mnohem častěji vzorkovány dobře dostupné lokality (např. podél komunikací, velkých sídel) nebo jsou vzorkovány pouze v určitém časovém období (např. pouze v určitou denní dobu) (Guillera-Arroita *et al.*, 2015). Systematická chyba může také vzniknout odstíněním GPS lokátorů umístěných na obojcích zvířat díky přerušení signálu v hustém porostu nebo hornatém terénu (Johnson *et al.*, 2009). Výsledkem systematické chyby může být změna odezvy druhu na prostředí, pokud bude jako environmentální proměnná použit gradient, který se systematickou chybou souvisí (např. zvýšené množství živin v půdě v blízkosti sídel díky aplikaci hnojiv.) (Guillera-Arroita *et al.*, 2009).

Také vliv chybné detekce, na přesnost modelu založeném na GLM, byl zkoumán. Model byl i v případě velmi nízké chyby v detekci druhu výrazně ovlivněn. Vliv na přesnost modelu mělo také to, zda environmentální proměnná, která způsobuje nízkou detekci, ovlivňovala počet organismů na lokalitě (Gu *et Swihart*, 2003).

2.4.3. Environmentální data

Environmentální data popisují kombinaci přímých, zdrojových a nepřímých environmentálních proměnných (Miller, 2010). Z některých dat, jako je nadmořská výška, či průměrná teplota lze odvozením získat rastry, které mají přímý vliv na fyziologické vlastnosti či chování druhu (Pearson, 2010). Environmentální data mohou vznikat odvozením interpolačními metodami z bodových dat, nebo z dálkového průzkumu země. V poslední době stoupá využití dat z DPZ díky zvyšující se kvalitě satelitních snímků (Bradley *et al.*, 2012). Data odvozená z DPZ mohou podat věrnější obraz o environmentální proměnné v zájmovém území než interpolovaná data z několika málo bodových měření (Deblauwe *et al.*, 2016). Existuje nepřeborné množství zdrojů, z nichž je možné environmentální data čerpat, příklad dostupných dat pro prostředí ČR shrnuje následující tabulka č. 2.

Tabulka 2: Přehled zdrojů environmentálních dat

organizace	dostupné online	popis dat
Český úřad zeměměřičský a katastrální	https://geoportal.cuzk.cz	ZABAGED - detailní výškopis ČR ve vrstevnicích či digitální model reliéfu detailní ortofoto mapa ČR, včetně archivních snímků
Národní geoportál	https://geoportal.gov	klimatické oblasti ČR vymezení chráněných území, biogeografických regionů či přírodních biotopů
Agentura ochrany přírody a krajiny	mapy.nature.cz	klimatické regiony ČR, expozice svahů, skeletovitost a hloubka půdy, ohrožení větrnou a vodní erozí
Výzkumný ústav meliorací a ochrany půdy	http://geoportal.vumop.cz	geologická mapa ČR, vlivy důlní činnosti, radonový index,

(autor, 2017)

Schopnost vybrat vhodná environmentální data pro vystihnutí nároků zkoumaného druhu na prostředí je důležitým krokem z hlediska optimalizace výzkumu (Guisan et al., 2007). Miller (2010) označuje klimatická a topografická data jako dvě nejpoužívanější kategorie v modelování druhové distribuce. Klimatická data popisují řadu faktorů jako je teplota, úhrn srážek či počet slunečných dní. Guisan et al. (2007) se ve své studii testovali vliv 13 environmentálních proměnných na přesnost výskytu několika druhů listnatých a jehličnatých stromů. Rychlost růstu odvozená z klimatických dat měla nejvyšší podíl na přesnosti modelu.

Topografická data mají v SDM široké využití. Lze je využít ve fázi úprav vstupních vrstev (georeferencování dat z dálkového průzkumu země (DPZ) či zpřesňující vrstva při interpolaci klimatických dat), lze z nich odvodit komplexnější environmentální proměnné, které lépe vystihují nároky druhu, nebo mohou být rovnou využita jako vstupní vrstva (Guisan et Zimmerman, 2000). Data typu nadmořská výška, sklon či orientace ke světovým stranám patří mezi nepřímé environmentální proměnné a v případě, že jsou k dispozici pouze v hrubém rozlišení, mohou v kombinaci přímých environmentálních proměnných (např.: potenciační oslunění) vést ke snížení predikční schopnosti modelu. (Guisan et Zimmerman, 2000).

Mezi další data, která mohou sloužit jako podklad při procesu modelování druhové distribuce, patří mapy půdních typů, ze kterých lze získat informace o dostupnosti živin pro rostliny (Guisan et al., 2007). Osborne et al. (2001) ve svém výzkumu zabývajícím se modelováním druhové distribuce dropa velkého využívá údajů o silniční síti, zástavbě či vodních zdrojích k odvození rastru vzdálenosti od jednotlivých zdrojů. V neposlední řadě lze pomocí DPZ získat informaci o habitatu, který druh obývá avšak je vhodnější pro jemná rozlišení vstupních vrstev a na malé ploše.

Je důležité vybrat takové environmentální proměnné, které mají přímý vliv na výskyt druhu v prostředí. V opačném případě mohou environmentální proměnné, které reflektují nároky druhu nepřímo způsobit při modelování výskytu v jiném regionu či pod jiným klimatickým scénářem výrazně chybně predikovaný areál rozšíření (Guisan et Thuiller, 2005). Osborne et Leitão (2009) se mimo jiné zabývali vlivem posunutí celé rastrové vrstvy na přesnost SDM. Výsledkem jejich studie bylo velmi nízké ovlivnění modelu při posunu až o 60 km. Vysvětlením je velmi nízká korelace mezi vybranými environmentálními proměnnými a zkoumaným druhem.

2.5. Vliv měřítka na SDM

Jednou z hlavních otázek při tvorbě SDM je volba vhodného měřítka. Samotné slovo „měřítka“ má však v oblasti GIS dva významy. Lze jím označit rozsah nebo rozlišení zájmové oblasti - tedy určitou míru podrobnosti s jakou je území mapováno (Goodchild, 2010). Teoreticky je možné mapovat zemský povrch s nekonečně velkým rozlišením, avšak v praxi je takový přístup limitován vysokým objemem dat, jehož zpracování je časově a finančně natolik náročné, že nezbývá než přistoupit na jistou míru generalizace (McMaster et Shea, 1992). Na rozdíl od vektorových dat jsou pro vědecké účely preferována data rastrová, která mají jednoznačně definované rozlišení (Goodchild, 2010).

Volba vhodného rozlišení vstupních environmentálních vrstev může záviset na velikosti domovského okrsku druhu a na způsobu, jakým druh využívá zdrojové environmentální proměnné v krajině. Naopak volba vhodného rozsahu závisí na hloubce orientace v hierarchii environmentálních gradientů (Guisan et Thuiller, 2005), na části zkoumané populace, či na sezónní fluktuaci jedinců (Jaberg et Guisan, 2001). Při modelování druhové distribuce u vysoce mobilních organismů je vhodnější zvolit rastr s větší velikostí buňky, která obsáhne několik různých habitatů (Jaberg et Guisan,

2001). Areál málo pohyblivých či přisedlých druhů je lépe modelován na environmentálních datech s vysokým rozlišením (Guisan et Thuiller, 2005). V případě modelování druhové distribuce vzácných druhů může nízké rozlišení způsobit systematickou chybu (Huston, 2002).

Chyba, s jakou jsou biologická data v terénu zaznamenávána, by neměla být větší, než je rozlišení rastru environmentálních dat. Obecně platí, že jemnější rozlišení environmentálních dat vede k vyšší prediktivní schopnosti SDM za předpokladu adekvátní kvality biologických dat (Huston, 2002). Moudrý et Šímová (2012) uvádějí, že rozlišení rastrů vstupních dat pohybující se 1 – 100 km² je ideální pro nalezení environmentálních limitů druhu a je dostačující pro modelování druhové distribuce při použití v současnosti dostupných dat.

2.6. Virtuální druh

Principem modelování distribuce virtuálního druhu je simulace pravděpodobnosti výskytu druhu (či jinak environmentální vhodnosti prostředí) dle jednoho nebo několika environmentálních gradientů za předpokladu předem definované reakce druhu na environmentální proměnnou (Meynard et Kaplan, 2013) a následné převedení těchto hodnot z environmentálního do geografického prostoru (Leroy et al., 2015).

Důsledkem stále se zvyšující oblíbenosti SDM je jejich neustálá modifikace. Aby bylo možné posoudit, zda a jak daná modifikace zdokonalila model, je nutné jeho validování. Tedy schopnost s jakou přesností dokáže model zpětně predikovat výskyt nezávislé sady biologických dat (Miller, 2014). Jak již bylo zmíněno v kapitole 2.4.2, biologická data mohou obsahovat řadu chyb a interpretace takových výsledků by mohla být zavádějící. Simulovaná data o výskytu s přesně definovanými parametry tak v mnoha případech představují jedinou možnost, jak nestranně porovnávat různé vlastnosti SDM (Austin et al., 2006, Moudrý, 2015). Následující tabulka č. 3 uvádí příklad několika studií s využitím virtuálního druhu.

Tabulka 3: Přehled nedávných studií s uplatněním virtuálního druhu

testované téma	autor
Generování predikce s vysokým rozlišením z dat o výskytu s nízkým rozlišením	Bombi et D'Amen, 2012
Efekt filtrovacích metod na redukci systematické chyby v biologických datech	Varela et al., 2014
Vliv prevalence na vztah druh - prostředí	Fukuda et Baets, 2016
Vliv prevalence a velikosti vzorku na přesnost modelu	Jiménez-Valverde et al., 2009
Vliv invazní fáze na přesnost SDM	Václavík et Meentemeyer, 2012
Porovnání prahové a pravděpodobnostní konverze	Meynard et Kaplan, 2013

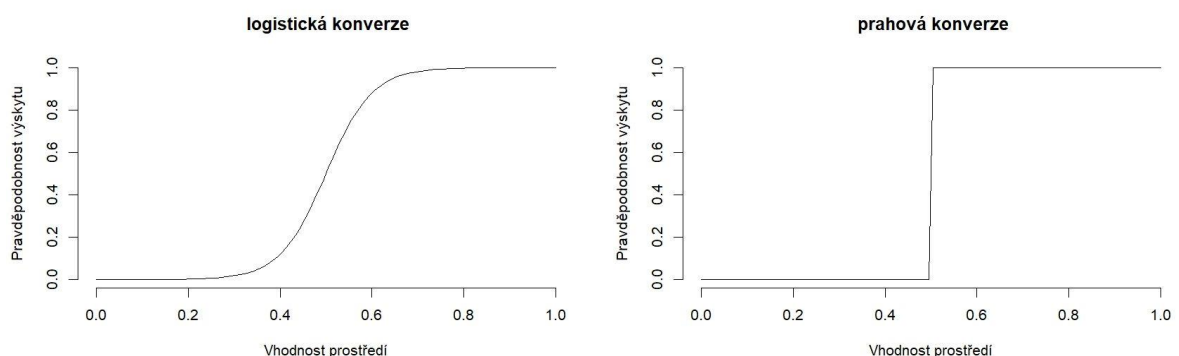
(Miller, 2014; upraveno)

Vlivem prevalence virtuálního druhu na přesnost modelů SDM se zabývali např. Jiménez-Valverde et al. (2009), či Fukuda et Baets (2016). Zatímco první studie testovala poměr presenčně – absenčních dat spolu s velikostí vzorku pomocí GLM druhá testovala prevalenci z hlediska interakce jednotlivých environmentálních proměnných. Výsledek obou studií však potvrzuje vliv prevalence na přesnost modelů druhové distribuce. Václavík et Meentemeyer (2012) využili virtuálního druhu při simulaci

jednotlivých fází šíření virtuálního druhu. Zjistili, že model kalibrovaný pro druh v počáteční fázi invaze výrazně podhodnocuje potenciální distribuci a proto by fáze invaze měla být brána při formulování modelu v úvahu. *Bombi et D'Amen (2012)* ověřovali pomocí virtuálního druhu přesnost modelů druhové distribuce s vysokým rozlišením, generovaných na datech s nízkým rozlišením. Závěrem jejich studie je, že vztah mezi environmentální proměnnou a reakcí druhu na ní je průkazný až při dvanáctinásobně vyšším rozlišení dat.

Pro simulování virtuálního druhu bylo vyvinuto několik platform. RangeShifter (*Bocedi et al., 2014*) a HexSim (*Schumaker, 2015*) byly navrženy pro modelování populační dynamiky, avšak neumožňují definovat vztah reakce druhu – prostředí. Ve statistickém programu R byly dále vyvinuty dva balíčky *SDMvspecies (Kong et al., 2015)* a *virtualspecies (Leroy et al., 2015)*. Zásadní rozdíl mezi těmito dvěma balíčky je především ve způsobu konverze na presenčně absenční data. Zatímco *SMDvspecies* využívá pouze prahové konverze tak *virtualspecies* umožňuje i pravděpodobnostní přístup. Další nevýhodou balíčku *SMDvspecies* je limitované množství křivek, popisujících odpověď druhu na prostředí (*Leroy et al., 2015*).

Prahová konverze vhodnosti prostředí na presenčně-absenční data je založena na stanovení hranice, nad kterou jsou všechny hodnoty převedeny na presenci a pod ní na absenci, avšak tento způsob je označován za nevhodný (*Meynard et Kaplan, 2013; Leroy et al., 2015; Moudrý, 2015*). Zvláště faktory, které se objevují na malých měřítcích (biotické interakce, náhodné události, mikrohabitáty, ...) mohou způsobit, že druh se nebude vyskytovat v územích s vysokou hodnotou vhodnosti prostředí a naopak (*Leroy, 2015*). *Meynard et Kaplan (2013)* prahový přístup kritizují, protože může za určitých podmínek způsobit nadhodnocení modelu. Pravděpodobnostní konverze umožňuje naopak výskyt presence či absence podél celého environmentálního gradientu. V případě, že pravděpodobnostní funkce bude mít stejný průběh, jako na obrázku č. 4, bude mít buňka s vhodností prostředí 0,3 v 1 případě z 10 přidělenou presenci. Opakovaná konverze vhodnosti prostředí pravděpodobnostním přístupem se nazývá filtrování a poskytuje informaci, která místa jsou druhem preferována (*Moudrý, 2015*). *Varela et al., 2014* se ve své studii využila virtuální druh k testování efektu geografických a klimatických filtrů na zmírnění dopadů systematické chyby v datech o výskytu. Závěr její studie potvrdil účinnost klimatického filtrování, naopak geografické filtry nezvýšili predikční schopnost modelu a v některých případech ji i mírně zhoršily.



Obrázek 4: pravděpodobnostní konverze dle logistické funkce ($\alpha = -0,05$) a prahová konverze ($\alpha = -0,000001$)

(*autor, 2017*)

2.7. Modelovací metody

Guissan et Zimmermann (2000) rozlišují tři žádoucí vlastnosti, kterých by měl model dosahovat: skutečnost, preciznost a obecnost. V praxi však model dosahuje kombinace dvou na úkor třetí vlastnosti. Žádný model nemůže současně dosahovat vysoké přesnosti (preciznosti) na základě skutečných fyziologických procesů a mechanismů (skutečnosti) a být obecně aplikovatelný na široké množství různých podmínek (obecnost). Mnoho studií, které porovnávaly různé algoritmy, se žádná z metod neprojevila jako konzistentně nezávislá na prostředí druhu a vlastnostech biologických dat (*Elith et al., 2006; Segurado and Araújo, 2004*).

V současné době existuje řada modelovacích metod, které jsou dostupné buď jako součást GIS nebo jako samostatný software. Typ algoritmu by měl být volen dle objemu, typu či rozdělení biologických dat, s ohledem na rozlišení, objem a komplexnost environmentálních dat a dle požadovaného mapového výstupu (např. aktuální či potenciální distribuce) (*Miller, 2010*).

Elith et al. (2006) dělí metody modelování dle typu vstupních dat na metody pro čistě presenční data a metody pro presenčně – absenční data. Do první skupiny lze zařadit metody modelující environmentální „obálku“ druhu (BIOCLIM), metody založené na posuzování vzdálenosti v environmentálním prostoru mezi potencionálním a pozorovaným výskytem (DOMAIN a LIVES) či MAXENT (model maximální entropie) což je metoda strojového učení porovnávající data na základě maximální uspořádanosti.

Druhou skupinu metod, využívající presenčně – absenční data dělí *Miller (2010)* na regresní metody a některé metody strojového učení, které nejsou na rozdíl od regresních metod založeny na specifické funkci, která distribuci druhu určuje. Metody strojového učení jsou vhodnější pro území malých rozsahů v protikladu k regresním metodám, které jsou vhodnější pro globální využití. *Elith et al. (2006)* uvádí jako příklad metod strojového učení pro presenčně – absenční data BRT (boosted regression trees) využívající klasifikaci dat do podobných shluků a GDM (generalized dissimilarity model), který porovnává dvě lokality na základě environmentální rozdílnosti.

Jako příklad nejvíce používaných regresních metod lze uvést GLM, GAM, CART, MARS a jeho modifikace. Na rozdíl od GLM, využívá GAM neparametrické funkce pro popis vztahu mezi druhem a prostředím. Díky vyšší flexibilitě je GAM vhodnější pro modelování s více komplexními proměnnými než GLM (*Elith et al., 2006*), a je vhodnější pro nalezení realistického tvaru křivky druh – prostředí (*Leathwick et al., 2005*), avšak na rozdíl od GLM neumí zahrnout interakce mezi proměnnými (*Kienast et al., 2012*). MARS je alternativou ke GAM, jeho výhodou je snazší převedení na predikční mapu pomocí GIS, avšak za předpokladu normálního rozdělení dat, což ho činí nepoužitelným v případě presenčně – absenčních dat (*Franklin et Miller, 2010*). Klasifikační a regresní stromy (CART) jsou další možností v SDM. Opakovaně klasifikují data na základě dichotomního rozhodujícího klíče čímž vysvětlují variabilitu mezi skupinami závislou na environmentálních proměnných. Stejně jako GAM, nevyžadují jakoukoliv formu rozdělení dat, avšak složité stromy se obtížně ekologicky interpretují. *Kienast et al. (2012)* proto doporučuje využití metody CART pro vstupní analýzu dat.

2.7.1. GLM a logistická regrese

Generalizovaný lineární model vychází ze statistické metody lineární regrese, která se využívá ke stanovení závislosti dvou či více proměnných a tu vyjadřuje přímkou ve tvaru:

$$y = k * x + q$$

Jedním z předpokladů pro použití lineární regrese je normální rozdělení proměnných, avšak prezenčně – absenční data nabývají pouze binárních hodnot 0 a 1. Generalizovaný lineární model transformuje binární hodnoty dle následující link-funkce:

$$y = \log\left(\frac{\mu}{1-\mu}\right)$$

a převádí tak hodnoty na poměr šancí, tedy pravděpodobnost kdy jev nastane (1) či nenastane (0) (Zvára, 2008). Díky transformaci lze dosáhnout žadoucích vlastností lineárního regresního modelu. Pravděpodobnost μ se nachází v intervalu (0, 1) a hodnoty výsledné šance leží v intervalu (0, ∞) (Grafarend, 2006). Po transformaci šance logitovou funkcí dojde k rozšíření intervalu na celou množinu reálných čísel ($-\infty, +\infty$) v závislosti na prediktorech, stejně jako u lineárního regresního modelu (Zvára, 2008).

Pro odhad parametrů logistické regrese slouží metoda maximální věrohodnosti definovaná jako:

$$L(\beta) = \ln(l(\beta)) = \sum_{i=1}^n (y_i \ln(\pi(x_i)) + (1 - y_i) \ln(1 - \pi(x_i)))$$

kde $\pi(x_i)$ představuje příspěvek věrohodnostní funkce pokud je y_i 1 a $1-\pi(x_i)$ pokud je y_i rovno 0 (Hosmer et Lemeshow, 2000).

2.8. Evaluace modelů

Validační metody hodnotí, do jaké míry se predikovaná data podobají těm pozorovaným. Existuje několik důvodů, proč predikční schopnost modelu testovat. V případě reálných dat je možné odhalit, jaká environmentální proměnná má největší vliv na distribuci. Díky validaci lze vybrat model, který nejlépe distribuci druhu predikuje, posoudíme spolehlivost predikce a můžeme odhadnout ve kterém kroku je třeba model vylepšit (Lawson et al., 2014).

Validační techniky lze rozdělit na kalibrační a diskriminační. Zatímco kalibrační metriky určují rozsah, v jakém se predikce shoduje s realitou v absolutní hodnotě (Murphy et Winkler, 1987), diskriminační metody měří schopnost rozpoznat presenci od absence na základě modelované predikce (Vaughan et Ormerod, 2005). Z toho vyplývá, že při posuzování diskriminační schopnosti jsou hodnoty prediktorů nepodstatné, model, který má vysokou diskriminační schopnost může být zároveň špatně kalibrován (Pearce et Ferrier, 2000). V rámci těchto dvou skupin mohou být hodnoceny binární (prezenčně – absenční data) nebo kontinuální predikované hodnoty (vhodnost prostředí). Validace založená na binárních datech bude vždy hodnotit diskriminační schopnost modelu, neboť se předpokládá, že binární data budou dosahovat vysoké míry kalibrace (Pearce et Ferrier, 2000). Následující tabulka č. 4 uvádí příklady validačních metod na základě výše uvedeného dělení.

Tabulka 4: Matice validačních technik

	kontinuální	binární
kalibrace	MNČ	KAPPA
	R^2	
	D^2	
	AIC	
diskriminace	AUC	Sensitivita
	Pearsonův korelační koeficient	Specificita

(Lawson et al., 2014; upraveno)

Nejjednoduššími ukazateli, které vypovídají o rozdílnosti modelů, jsou AIC, D^2 a R^2 . Jejich nevýhodou je, že nemohou být použity při srovnávání různých modelovacích technik (Miller, 2010). Hodnota Akaikeho informačního kritéria je tvořena součtem logaritmu residuální sumy čtverců a koeficientem složitosti modelu. Modely s nižší hodnotou AIC jsou preferovány, protože dokáží s menším počtem koeficientů lépe vystihnout realitu (Šmilauer, 2007). Koeficient determinance (R^2) vyjadřuje jaký podíl z celkových dat je vysvětlen modelem. Hodnoty se pohybují od 0 do 100%, avšak v případě ekologických modelů se považuje za dobrý výsledek i 30%, vzhledem k vysoké nahodilosti procesů v prostředí. Deviance (D^2) popisuje, jaké množství variability v datech bylo modelem popsáno (Kienast et al., 2012).

Robustní ověřování přesnosti modelu zahrnuje použití nezávislých testovacích dat pořízených na jiné lokalitě nebo jinou metodikou sběru (Guisan et Zimmermann, 2000). Pokud taková data nejsou dostupná (jako v případě virtuálního druhu), může být datová sada rozdělena na trénovací a testovací sadu (Miller, 2010; Miller, 2014). Pearson et al. (2007) uvádějí, že ideální poměr trénovací a testovací sady je 75: 25 s více než deseti environmentálními proměnnými. Na podobném poměru se shoduje i Kienast et al. (2012), který uvádí poměr 3:1.

Chybová matice tvoří základ mnoha binárních validačních technik, jejichž výsledky jsou závislé na zvolené prahové hodnotě (Lawson et al., 2014). Kappa (výpočet Miller, 2010) měří do jaké míry je model lepší, než náhodná predikce založená na prevalenci. Kienast et al. (2012) uvádějí, že modely dosahující hodnoty Kappa 0.8 jsou výborné, do hodnoty 0.6 jsou dostačující a modely nabývající záporných hodnot Kappa jsou systematicky špatné. Prahová hodnota blízká prevalenci dosahuje nejlepších výsledků Kappa. Sensitivita a Specificita jsou diskriminační metriky odvozené opět z chybové matice. Sensitivita vyjadřuje poměr správně určené presence ku celkové presence ve vzorku (část presence modelem vyhodnocená opět jako presence a část presence modelem klasifikovaná jako absence). Specificita vyjadřuje poměr správně klasifikované absence ku celkové absenci ve vzorku. Nejlepší diskriminační schopnosti dosahuje model s vyváženým poměrem obou koeficientů. Citlivost Sensitivity se projeví při nízké prahové hodnotě, kdy může dojít k nadhodnocení modelu. Binární metriky upřednostňují modely se sníženou schopností diskriminace, což vede k pod nebo nadhodnocení predikce (Lawson et al., 2014).

Tabulka 5: chybová matice

		Pozorovaná data	
		P	A
Predikovaná data	P	skutečná presence (A)	falešná presence (B)
	A	falešná absence (C)	skutečná absence (D)

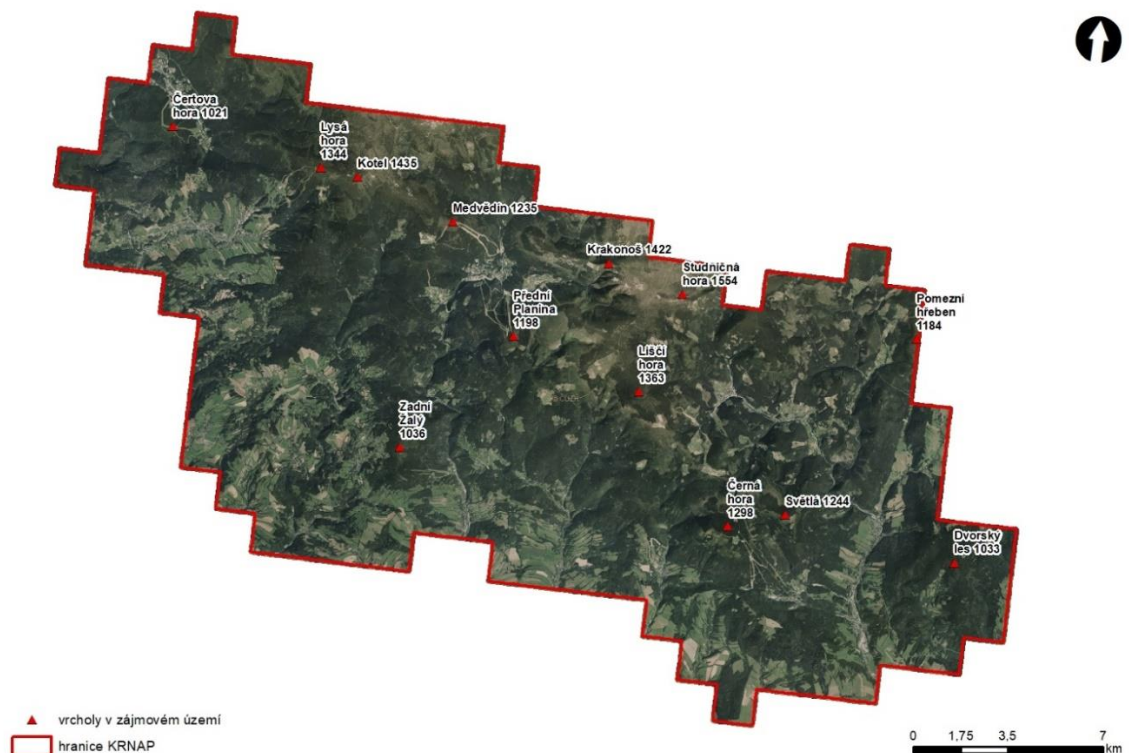
(Kienast et al., 2012)

Plocha pod křivkou (AUC) je jedna z nejvyužívanějších validačních metod vůbec (Lawson et al., 2014). AUC hodnotí poměr správně klasifikované presence a absence na všech možných hladinách prahové hodnoty a není tedy závislá na prevalenci. Výstupem je graf závislosti poměru skutečné presence ku poměru skutečné absence pro dostupné prahové hodnoty. Modely s hodnotou AUC do 0.8 jsou považovány za dobré, od 0.7 se považují za slabé a od 0.6 lze říci, že model má nulovou vypovídací hodnotu (Kienast et al., 2012). Lawson et al. (2014) doporučuje využívání validační metriky závislé na prevalenci vždy, když záleží na absolutních hodnotách výskytu druhu a dále tvrdí, že by metody využívající kontinuální data měly být upřednostňovány před binárními z důvodu ztráty části informace při konverzi na binární data.

3. Metodika

3.1. Charakteristika zájmového území

Krkonošský národní park (KRNAP) byl vyhlášen v roce 1963 za účelem ochrany „uchování a zlepšení přírodního prostředí, zejména ochrana či obnova samořídících funkcí přírodních systémů...“. Se svou rozlohou cca 550 km² zasahuje do dvou krajů Královehradeckého a Libereckého (KRNAP, 2010). Z hlediska geomorfologického členění se celé území KRNAP nachází v Krkonoško-jesenické soustavě, celku Krkonoše. Celým územím prochází horský hřeben začínající na západě Novosvětským sedlem, který končí po 35 km sedlem Královeckým (Demek et al., 2006).

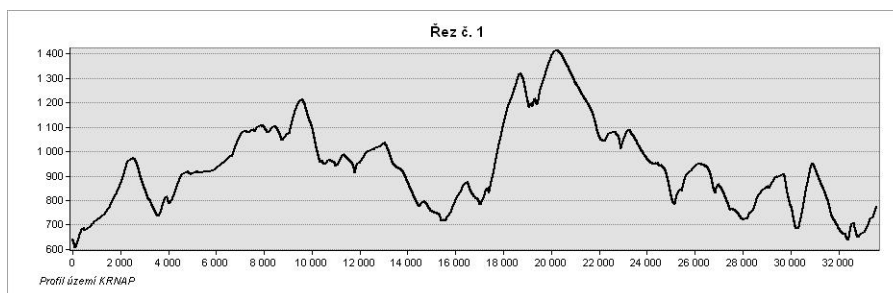


Obrázek 5: Ortofotomapa KRNAP s vyznačením nejvyšších vrcholů v území

(Ortofoto podklad: ČUZK, 2017;

vrcholy v zájmovém území: ARCDATA PRAHA s. r. o., 2017)

Na následujícím obrázku č. 6 je znázorněn výškový profil území KRNAP, který prochází zmiňovaným hřebenem.



Obrázek 6: Výškový profil hřebenu KRNP

(autor, 2017)

Z Grafu je patrné, že se jedná o území geomorfologicky značně proměnlivé. Na cca 32 km, ze kterých byl průběh výškového profilu odečten, se nachází cca 32 výraznějších lokálních extrémů, které mezi sebou mají prudké přechody. Toto rychlé střídání nadmořské výšky se může projevit při nepřesné lokalizaci druhu.

3.2. Charakteristika vstupních dat

Jako vstupní data byly použity dva základní parametry a to výškový digitální model terénu (DTM) a výška vegetace. Z digitálního modelu terénu byly následně odvozeny v programu ArcMap ver. 10.4. další dva parametry a to sklon terénu ve stupních a orientace vůči světovým stranám (tzv. aspekt). Všechny vstupní rastry jsou v souřadnicovém systému S-JTSK, a ve výškovém systému Baltském, pro potřeby této práce nebyla data autorem této práce transformována. Původní data pocházejí z projektu Snímkování a skenování území KRNP a jeho ochranného pásma z roku 2012 a byla pořízena firmou Geodis Brno, spol. s. r. o. Snímkování probíhalo během letních měsíců od července do srpna, tedy v hlavní vegetační fázi, kdy všechny stromy (listnaté, tak i jehličnaté) dosahovaly maximálního olistění. V průběhu letecké akce byla nasnímaná plocha o rozloze 550 km² z letové výšky 700 m a pod skenovacím úhlem ±30° s minimálním 20 % překrytím sousedních snímků a minimální průměrovou hustotou laserových paprsků 5 bodů/ m². Následující tabulka shrnuje některé základní parametry vstupních vrstev jako je rozpětí hodnot, rozloha nasnímané plochy, která se liší s údaji o rozloze Krkonošského národního parku s jinými zdroji, či původ vstupní vrstvy.

Tabulka 6: Základní parametry vstupních vrstev

název vrstvy	rozlišení (m ²)	rozloha (km ²)	minimum	maximum	průměr	jednotky	původ
dtm			393,7	1555,9	853,2	m n. m.	
vegetace	1	472,05*	0	52,8	9,2	m	
sklon			0	85,8	16,4	°	dtm
aspekt			0	360	179,5	°	dtm
poznámky	* rozloha vrstvy, rozloha parku dle správy KRNP činí 550 m ² včetně ochranného pásma						

(autor, 2017)

Obrázky zařazené v příloze v kapitole 8.1 Grafické přílohy pod č. 1 a č. 3 až č. 5 představují upravené mapové výstupy vstupních vrstev nasnímaného území KRNP.

3.3. Použitý software

Jak již bylo uvedeno výše, pro účely této diplomové práce byla některá vstupní data (sklon a aspekt) nejprve odvozena v programu ArcMap 10.4. Aby bylo možné data nahrát do jednoho objektu tzv. „raster stacku“ musela splnit podmínku stejné rozlohy. Z tohoto důvodu byl rastr vegetace „oříznut“

v programu ArcMap 10.4 funkcí *Extract by mask*. Pro další výpočty byl použit statistický software R Studio (použitá verze softwaru: R Studio 1.0.136), který využívá programovacího jazyku R (obdobu open source programovacího jazyka S), který je volně dostupný na tomto odkazu: www.rstudio.com. Pro provedení další analýzy bylo nutné implementovat několik extenzí a to: *rgdal* – pro práci se souřadnicovými systémy, *raster* – pro manipulaci s rastrovými daty, *spatial* – pro bodové analýzy, *sp* – pro převod dat do map, zjišťování souřadnic, *rJava* – rozhraní Java, umožňuje vytvoření objektů, jejich volání a přístup k polím, *virtualspecies* (Leroy et al., 2015) – modelování virtuálního druhu, a převod na presenčně-absenční data. Výstupy v podobě csv tabulky byly následně opět vyhodnoceny v programu R Studio, k čemuž byl použit balíček *PresenceAbsence* – pro statistické hodnocení vygenerovaných dat a balíčky *corrplot* a *ggplot2* – pro znázornění výstupů v podobě grafů. Jednotlivé kroky, které popisují následující kapitoly, provedené pomocí příkazů v softwaru R jsou zařazeny v příloze v kapitole 8. 2.

3.4. Generování dat

Aby bylo možné zkoumat přesnost vstupních dat na vliv druhové distribuce, je nejprve nutné mít taková kontrolní data, u nichž existuje předpoklad velmi vysoké přesnosti záznamu jednotlivých výskytů. To s reálnými daty není v praxi možné a v současné době to představuje jeden z hlavních limitujících faktorů pro praktické využití SDM (Moudrý, 2015) a z tohoto důvodu je nutné vytvořit tzv. „virtuální druh“.

3.4.1. Definování virtuálního druhu

Virtuální druh je matematický model skládající se z několika nedefinovaných odezvových funkcí (lineární, kvadratická, exponenciální, Gaussova, logaritmická, etc.) na předem zjištěné podmínky prostředí (v tomto případě rastry nadmořské výšky, vegetace, sklonu a aspektu), které spolu mohou být v různých kombinacích. Sčítání odezvových funkcí vyjadřuje skutečnost, že funkce jsou na sobě nezávislé, avšak jejich hodnoty v určitém místě mají za následek zvýšení pravděpodobnosti výskytu druhu, násobení naopak znamená, že hodnota pravděpodobnosti výskytu je závislá na všech proměnných, které jsou v součinu. K vygenerování virtuálního druhu v programu R Studio slouží příkaz *generateSpFromFun*. Výsledkem je funkce, která vyjadřuje pravděpodobnost výskytu organismu na základě environmetálních proměnných v prostředí.

Pro popis vztahu výskyt virtuálního druhu – environmentální prostředí byly v této práci použity dva typy funkcí a to Gaussova křivka v případě vrstvy DTM, sklon a aspekt a lineární funkce v případě vrstvy vegetace.

Obecný předpis tvaru Gaussovy křivky je definován jako:

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

Kde μ vyjadřuje průměrnou hodnotu, σ standartní odchylku a $f(x)$ vrací hodnotu funkce. Lineární funkce je pak definována tímto vzorcem:

$$f(x) = ax + b$$

- 1) Pro vrstvu vyjadřující hodnoty nadmořské výšky (DTM) vypadá předpis Gaussovy křivky následovně:

$$f(x_1) = \frac{1}{200\sqrt{2\pi}} e^{-\frac{(x-950)^2}{2 \times 200^2}}$$

Jedná se tedy o takový druh, jehož optimum je v nadmořské výšce 950 m n. m., avšak začíná se vyskytovat již od 750 m n. m. a pravděpodobnost jeho výskytu klesá až k horní hranici 1150 m n. m.

- 2) Pro vztah mezi výskytem virtuálního druhu a výškou vegetace byla použita lineární funkce, dána předpisem:

$$f(x_2) = -0.05x + 8$$

To znamená, že pravděpodobnost výskytu druhu se bude v místech s vyšší vegetací snižovat.

- 3) Vztah mezi sklonem a přítomností virtuálního druhu opět popisuje Gaussova křivka:

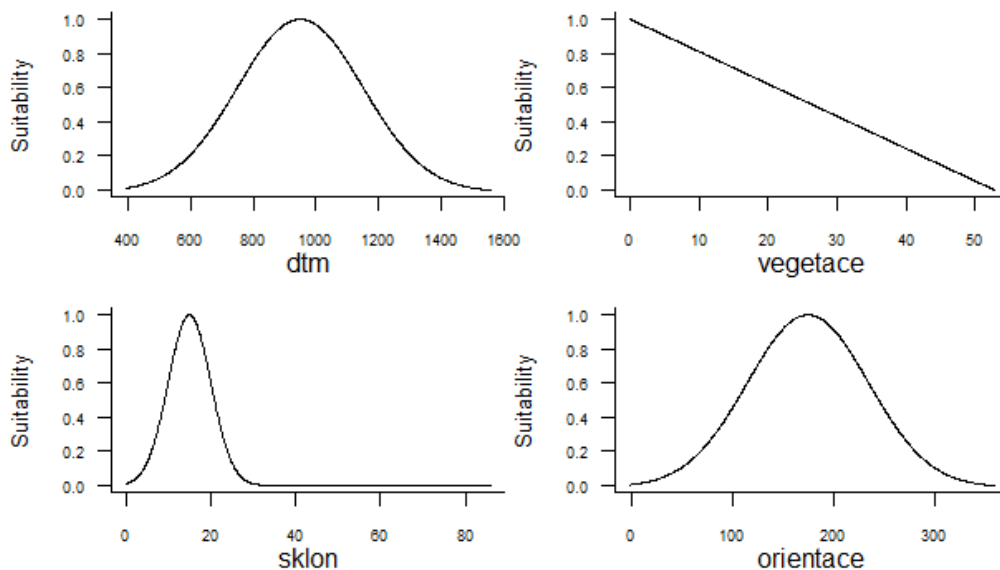
$$f(x_3) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{(x-15)^2}{2 \times 5^2}}$$

- 4) Stejně je tomu i u poslední environmentální proměnné aspektu.

$$f(x_4) = \frac{1}{60\sqrt{2\pi}} e^{-\frac{(x-175)^2}{2 \times 60^2}}$$

V tomto případě je optimální orientace svahu na západ. Pravděpodobnost výskytu druhu pak klesá na obě strany v intervalu 60 °.

Na následujícím obrázku č. 7 jsou znázorněny průběhy funkcí pravděpodobnosti výskytu virtuálního druhu na jednotlivé složky prostředí. Například v případě environmentální proměnné vegetace je patrné, že nejvhodnější prostředí je s velmi nízkou vegetací, naopak tam, kde výška vegetace dosahuje více než 50 m se pravděpodobnost výskytu druhu blíží nule. K vyvolání grafů slouží v programu R funkce *plotResponse*.



Obrázek 7: Průběh funkcí vyjadřuje závislost pravděpodobnosti výskytu druhu na jednotlivých environmentálních složkách

(autor, 2017)

Výše byly naznačeny dva způsoby, kterým lze rovnice kombinovat. V tomto případě se funkce mezi sebou násobí:

$$\begin{aligned} & \text{pravděpodobnost výskytu virtuálního druhu v prostředí} \\ & = f(x_1) * f(x_2) * f(x_3) * f(x_4) \end{aligned}$$

Což znamená, že pravděpodobnost výskytu druhu je závislá na všech environmentálních proměnných v jednom místě současně. Výsledná složená funkce je vizualizovaná do tzv. „rastru vhodnosti prostředí“, který na škále od 0 do 1 vyjadřuje pravděpodobnost výskytu organismu v daném místě. Následující obrázek č. 8 reprezentuje výslednou odezвовou funkci na prostředí, složenou z jednotlivých odpovědí organismu na složky prostředí v rozlišení 1 x 1 m².



Obrázek 8: A – detail „rastru vhodnosti prostředí“ rastru, B – vhodnost environmentálních podmínek pro nadefinovaný virtuální druh v celém území KRNP

(autor, 2017)

3.4.2. Konverze dat na presenčně – absenční rastr

Samotný „rastr vhodnosti prostředí“ vyjadřuje pouze pravděpodobnost, s jakou bude daný druh na lokalitě detekován, avšak neposkytuje informaci, zda byl druh na lokalitě skutečně přítomný. Dalším krokem je převedení pravděpodobnostních dat na data o výskytu (data je možné vizualizovat formou presenčně – absenčního rastru), k tomuto účelu slouží funkce *convertToPA*.

Jak již bylo uvedeno v kapitole 2.6. Virtuální druh, existují dva způsoby jak převést data o pravděpodobnosti výskytu druhu na presenčně absenční data. Práhový přístup, který buňkám přiřazuje presenci či absenci na základě stanovení prahové hodnoty a pravděpodobnostní přístup, který převádí pravděpodobnostní data podle logistické funkce.

Obecný předpis tvaru logistické křivky je definován takto:

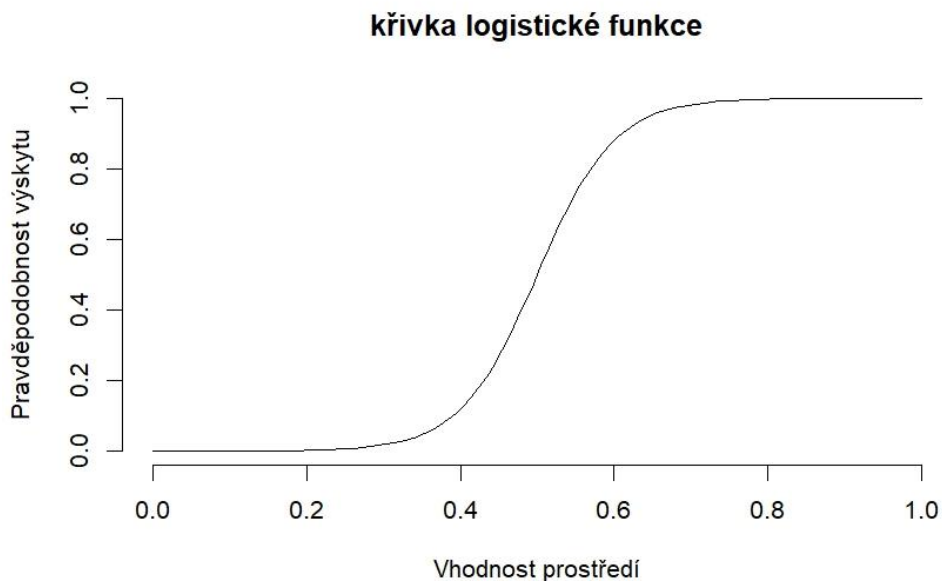
$$f(x) = \frac{1}{1 + e^{-\frac{x-\beta}{\alpha}}}$$

Kde parametr α – definuje strmost průběhu funkce a β – vyjadřuje hodnotu inflexního bodu. V tomto případě byly pro převod na presenčně – absenční data ponechány defaultní hodnoty.

Průběh logistické funkce má pak následující tvar:

$$f(x) = \frac{1}{1 + e^{\frac{x-0,5}{-0,05}}}$$

Což znamená, že buňka s pravděpodobností výskytu druhu 50% bude v polovině případů označena jako presenční a v polovině případů jako absenční. Na následujícím obrázku č. 9 je znázorněn průběh logistické funkce, která vyjadřuje pravděpodobnost přiřazení presence či absence na základě pravděpodobnosti výskytu druhu v daném místě.

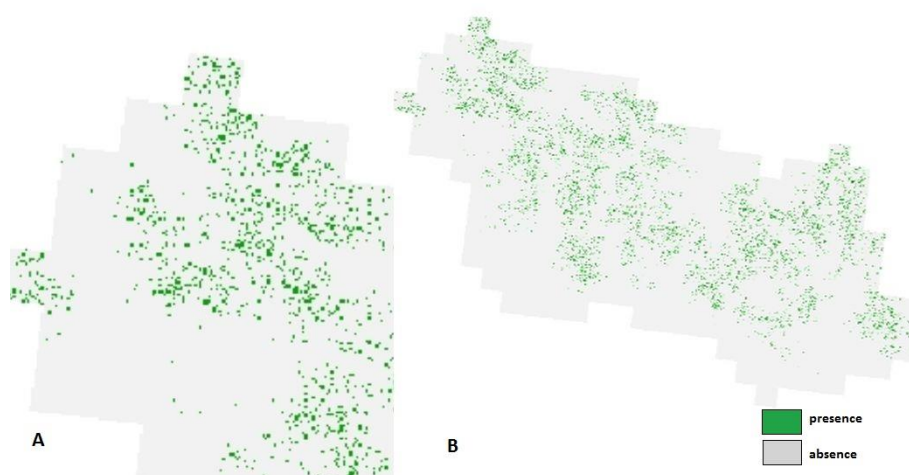


Obrázek 9: Křivka logistické funkce pro převod na presenčně absenční rastr z dat pravděpodobnosti výskytu,

$$\alpha = -0,05; \beta = 0,5$$

(autor, 2017)

V tomto kroku byla presenčně absenční data generována ve 100 opakováních ze stále stejných dat pravděpodobnosti výskytu. Na následujícím obrázku č. 10 je znázorněn detail presenčně-absenčního rastru při jedné z konverzí.



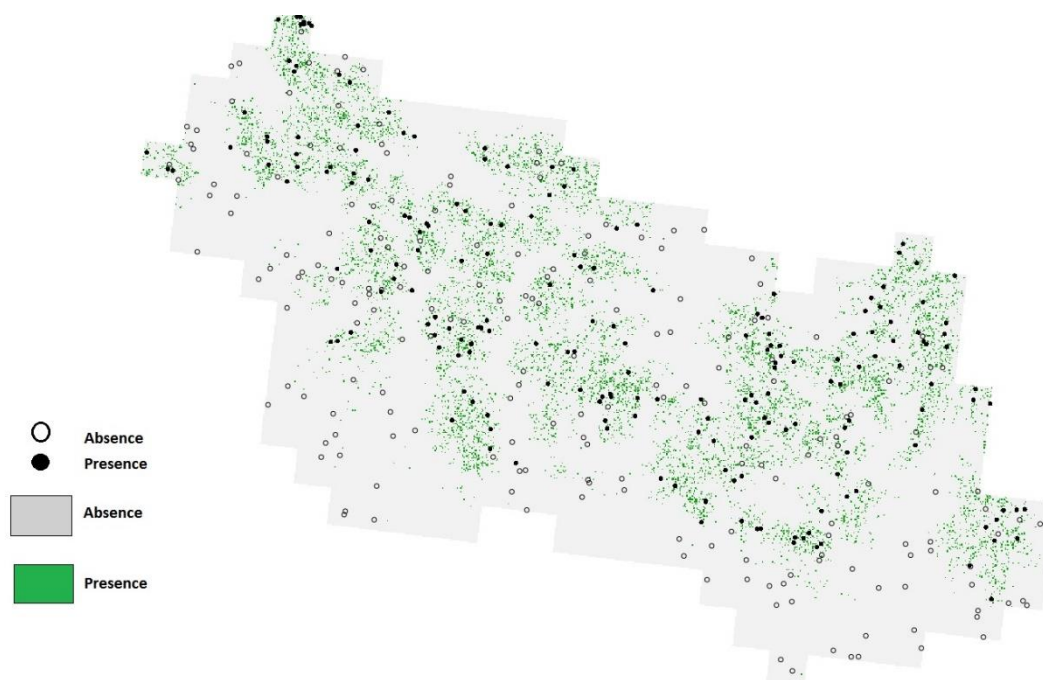
Obrázek 10: A – detail presenčně absenčních dat, B – přehled presence virtuálního druhu na celém území KRNAP

(autor, 2017)

3.4.3. Vzorkování dat

K tomu, aby bylo možné otestovat, jakým způsobem náhodný posun dat ovlivní výsledný model druhové distribuce je potřeba získat data o výskytu druhu v zájmovém území. Toho lze docílit použitím funkce *sampleOccurrences*.

Díky skutečnosti, že v této studii se využívá virtuálního druhu, bylo možné použít presenčně-absenční typ dat (PA), kvůli tomu že odpadá jejich náročný sběr v terénu a máme stoprocentní jistotu, že druh se na lokalitě s absencí skutečně nevyskytuje. Na každém presenčně – absenčním rastru bylo náhodně vybráno 400 buněk tak, aby 50 % z nich připadlo na presenci a 50% na absenci (datová prevalence 0.5) (varianta 0). Dále byla simulována chyba detekce druhu ve třech krocích o 1, 10 a 50 m tak, že se v každém kroku k souřadnicím x a y přičetly náhodné hodnoty v intervalu od -1 do 1 (varianta A), -10 do 10 (varianta B) a -50 do 50 metrů (varianta C). Následující obrázek č. 11 představuje rozložení presenčně – absenčních dat na území Krkonošského národního parku v průběhu 52. cyklu.



Obrázek 11: Rozložení 400 presenčně – absenčních bodů na území KRMAP

(autor, 2017)

3.5. Statistické vyhodnocení

Nejprve byla provedena základní analýza celého souboru dat. Jednotlivé proměnné byly zobrazeny pomocí histogramu v rámci variant pro srovnání vlivu posunu. Dále byly odečteny základní parametry charakterizující proměnnou v rámci každé varianty. V dalším kroku byla modelována křivka popisující vztah mezi druhem a pravděpodobností výskytu v prostředí pro každou proměnnou v rámci jednotlivých variant posunu.

Na základě naměřených dat byly vytvořeny dva soubory: trénovací a testovací set v poměru 7:3. Trénovací set byl použit k vytvoření modelu druhové distribuce pomocí generalizovaného lineárního modelu (GLM) opět v softwaru R, který popisuje vztah mezi presenčně – absenčními daty a environmentálními proměnnými. Každá varianta posunu byla hodnocena zvlášť, v případech, kdy data nespĺňovala podmínku linearity, byl do modelu přidán kvadratický člen (nadmořská výška, sklon a expozice). Z výsledku modelu byla dále dopočítána vysvětlená variabilita.

Koeficienty β jednotlivých prediktorů a testovací set byly použity jako vstup do funkce *predict*, jejímž výstupem byly pravděpodobnosti výskytu pro vzorkované buňky. Křivka AUC pak byla vyhodnocena na základě těchto pravděpodobností pro presenčně – absenční data v testovací sadě. Druhým výstupem byla chybová matice pro zvolenou prahovou hodnotu, ze které byly odvozeny hodnoty KAPPA, sensitivita, specificita či PCC.

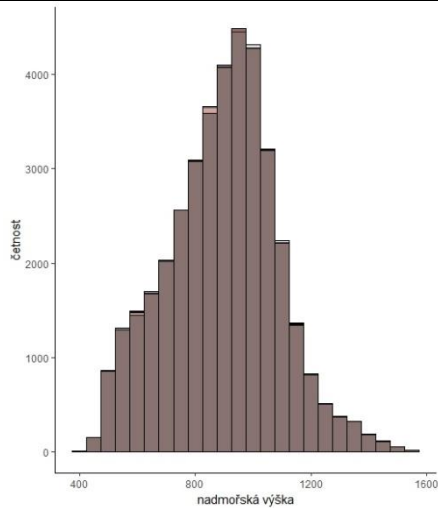
Pro vyhodnocení Pearsonova korelačního koeficientu bylo nutné odvodit opět pomocí funkce *predict* pravděpodobnosti výskytu druhu pro celé zájmové území. Poté byly rastry vhodnosti prostředí pro virtuální realitu a varianty 0 – C porovnány pomocí funkce *corrplot*.

4. Výsledky

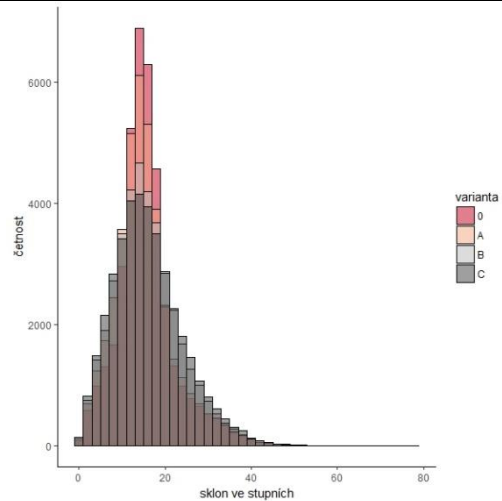
4.1. Základní analýza environmentálních dat

Z hlediska dopadu, jaký měl náhodný posun dat o výskytu druhu na odečtené hodnoty environmentálních proměnných v místě detekce, lze konstatovat, že sklon a orientace vůči světovým stranám byly posunem ovlivněny nejvíce. Z toho vyplývá, že tyto dvě proměnné budou mít na změnu predikční schopnosti modelu největší vliv. Nejméně byla oproti původnímu očekávání ovlivněna nadmořská výška.

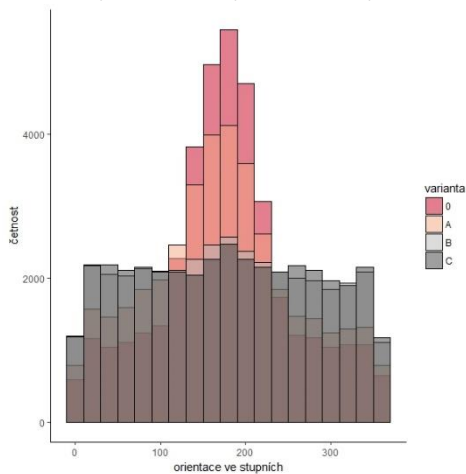
Histogram četnosti jevu



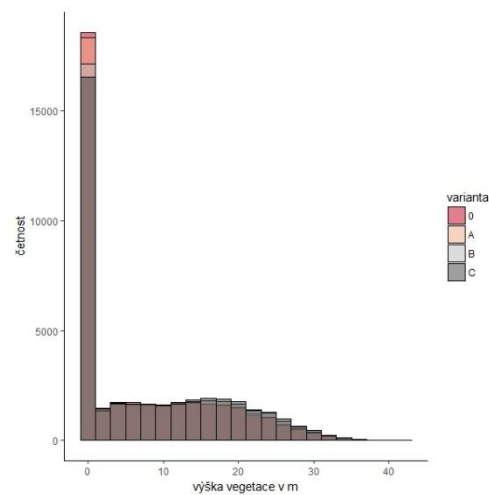
$E_0 = 895.7$; $E_A = 895.7$; $E_B = 897.7$; $E_C = 895.6$



$E_0 = 15.71$; $E_A = 15.30$; $E_B = 15.97$; $E_C = 16.13$



$E_0 = 177.6$; $E_A = 173.3$; $E_B = 177.4$; $E_C = 178.5$



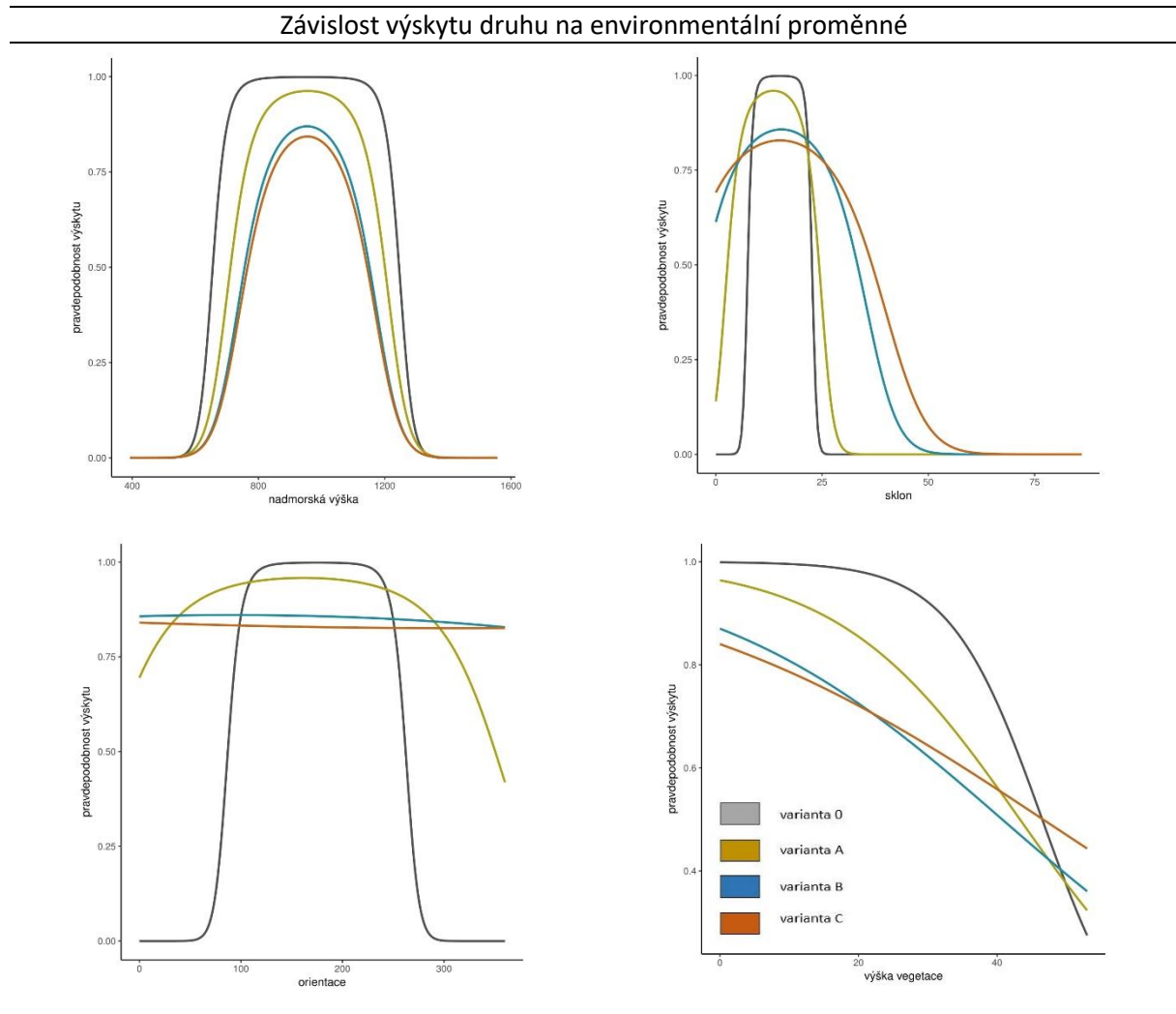
$E_0 = 7.28$; $E_A = 7.43$; $E_B = 8.12$; $E_C = 8.48$

$n = 38832$

Obrázek 12: Rozložení četnosti environmentálních proměnných dle jednotlivých variant posunu, E je střední hodnota proměnné, n = velikost vzorku

(autor, 2017)

Z následujícího obrázku č. 13 je patrné, že chybnou lokalizací druhu byla do jisté míry ovlivněna reakce na všechny environmentální proměnné. Vliv posunu dat se nejvíce projevil u průběhu křivky závislosti pravděpodobnosti výskytu druhu na environmentální proměnné podle očekávání u orientace. Chybná lokalizace v rozmezí od 0 do 10 a do 50 metrů způsobila, že druh byl mylně považován za euryvalentní ve vztahu k orientaci svahu vůči světovým stranám. Vyšší tolerance v rozmezí hodnot výskytu virtuálního druhu byla pozorována i u sklonu. Naopak u nadmořské výšky a vegetace došlo díky chybné lokalizaci ke snížení tolerance virtuálního druhu vůči jeho prostředí.

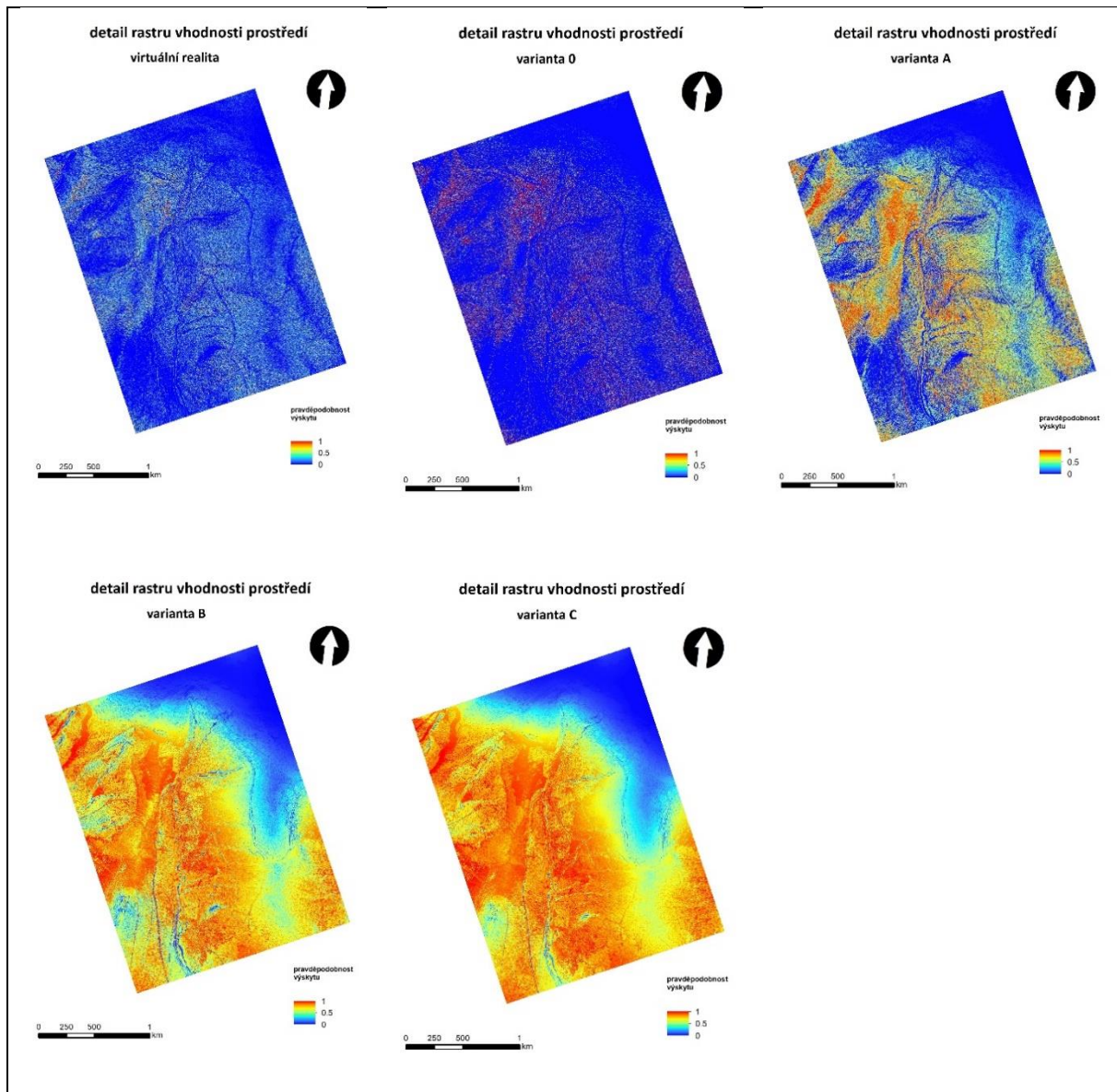


n = 38832

Obrázek 13: Vliv posunu dat na výslednou podobu křivky popisující pravděpodobnost výskytu druhu v závislosti na environmentální proměnné

(autor, 2017)

Z detailu zpětně odvozených rastrů vhodnosti prostředí na základě koeficientů determinace β vyplývá, že s rostoucí chybou v lokalizaci virtuálního druhu došlo k nadhodnocování pravděpodobnosti výskytu virtuálního druhu. Na variantě 0 je patrné, do jaké míry dokázal zvolený model zpětně reflektovat virtuální realitu z náhodného vzorku presenčně-absenčních dat. V případě variant B a C je z pouhého pohledu patrné, že model silně nadhodnocuje pravděpodobnost výskytu i v lokalitách, které jsou v případě rastru virtuální reality klasifikovány jako velmi málo pravděpodobné. Nadhodnocená predikce je pravděpodobně důsledkem chybně odvozené funkce reflektující vztah mezi druhem a environmentální proměnnou, umocněná interakcemi mezi nimi.



Obrázek 14: detail rastrů vhodnosti prostředí pro jednotlivé varianty

(autor, 2017)

4.2. AIC, vysvětlená variabilita a AUC

Jak již bylo uvedeno výše, jedním ze základních kritérií při výběru modelu by měla být hodnota AIC a vysvětlené variability. Vzhledem k tomu, že distribuce druhu byla modelována vždy ve stejných podmínkách, je toto porovnání relevantní. Nejnižší hodnoty AIC dosahuje varianta 0 (data bez posunu). Naproti tomu varianta A, kdy byla data posunuta náhodně o 1 m, dosahuje více než trojnásobně vyšší hodnoty AIC. Překvapivě nízký rozdíl hodnot AIC nastal mezi variantami B a C, navzdory skutečnosti, že informace o environmentálním prostředí, kde byl druh zaznamenán, byla odečtena až s pětinásobnou chybou oproti variantě B.

Podíl vysvětlené variability modelem klesal dle předpokladu od varianty 0 po variantu C. Navzdory vysoké proměnlivosti prostředí a skutečnosti že posun dat u variant B byl až o 10 buněk a u varianty C až o 50 buněk není rozdíl mezi vysvětlenou variabilitou těchto dvou variant výrazný.

Hodnota AUC klesá dle předpokladu od varianty 0 po variantu C. Na základě klasifikační tabulky modelů dle hodnot AUC a Kappa, kterou uvádí *Kienast et al. (2012)* lze tvrdit, že modely variant 0 a A dosahují velmi vysoké přesnosti predikce presenčně – absenčních dat. Modely variant B a C vykazují navzdory posunům dobrou diskriminační schopnost.

Tabulka 7: Srovnání modelů dle AIC, vysvětlené variability a AUC

GLM				
Varianta	AIC	DEVIANCE	AUC	AUC sd
0) bez pos	5591	0.8520	0.9927	0.0006
A) ± 1 m	18040	0.5216	0.9316	0.0023
B) ± 10 m	25328	0.3281	0.8434	0.0036
C) ± 50	26226	0.3043	0.8218	0.0039
n = 27182				

(autor, 2017)

4.3. Metriky odvozené z chybové matice

Z chybové matice odvozené pro konkrétní prahové hodnoty byla vypočítána Kappa, vyjadřující kalibrační schopnost modelu a hodnoty sensitivity a specificity vyjadřující diskriminační schopnost modelu. Tyto metriky jsou závislé na prahové hodnotě při konverzi na presenčně absenční data.

Z tabulky č. 8. je patrné, že ve všech případech volené prahové hodnoty dosahoval model varianty 0 jak nejvyšších hodnot Kappa, tak i sensitivity a specificity. Dle *Kienasta et al. (2012)* jsou níže uvedené hodnoty Kappa odpovídající hodnotám AUC uvedených v tabulce č. 7. Dále je možné pozorovat, že čím je prahová hodnota nižší, tím vyšší je sensitivity, avšak důsledkem může být nadhodnocení predikce virtuálního druhu. Varianty B a C vykazovaly nižší procento správně klasifikovaných výskytů v důsledku nerovnoměrného poměru mezi sensitivitou a specificitou.

Tabulka 8: Metriky odvozené z chybové matice pro různé prahové hodnoty

		Sensitivita a Specificita			
		prahová hodnota = 0.5			
		Kappa	Sensitivita	Specificita	PCC
0		0.9313	0.9776	0.9541	0.9657
A		0.7145	0.8908	0.8244	0.8572
B		0.5517	0.8668	0.6862	0.7754
C		0.5205	0.8592	0.6627	0.7597
t		prahová hodnota = maximální PCC			
0	0.6	0.9320	0.9715	0.9607	0.9660
A	0.55	0.7170	0.8719	0.8454	0.8585
B	0.44	0.5549	0.9044	0.6523	0.7767
C	0.45	0.5257	0.8957	0.6318	0.7621
		prahová hodnota = sensitivity > 85%			
0	0.91	0.8480	0.8593	0.9873	0.9241
A	0.59	0.7126	0.8529	0.8596	0.8563
B	0.52	0.5495	0.8513	0.6693	0.7743
C	0.51	0.5206	0.8519	0.6699	0.7597
		n = 11650			

(autor, 2017)

4.4. Korelační koeficient

Pearsonův korelační koeficient je další validační metoda hodnotící diskriminační schopnosti modelu na škále kontinuálních dat, na rozdíl od AUC však zohledňuje pravděpodobnost výskytu, která byla v jednotlivých buňkách dosažena. Hodnoty Pearsonova korelačního koeficientu ukazují na významnou shodu mezi virtuální realitou a variantou 0. Hodnota korelačního koeficientu nadále klesá, se zvyšující se hodnotou posunu vysvětlujících proměnných. Nejnižší shody s virtuální realitou dosahuje varianta C. Nejsilnější kladně korelovaný vztah se projevil mezi variantou B a C navzdory skutečnosti, že chyba v lokalizaci u varianty C je až pětinasobně vyšší.

Symbole nad hlavní diagonálou vyjadřují, na jaké hladině významnosti je statistická shoda Pearsonova korelačního koeficientu významná. Ve všech případech dosahuje hladina významnosti nižší hodnoty než 0.001.

Pearsonův korelační koeficient



n = 472051332

Obrázek 15: korelační matice, hladina významnosti testu: 0.05 = *; 0.001 = **; <0.001 = ***

(autor, 2017)

5. Diskuse

Histogramy, pomocí kterých byly environmentální proměnné znázorněny dle jednotlivých posunů v lokalizaci druhu, ukázaly, jak se změnilo rozložení četnosti jevu v rámci prediktoru, a lze z nich odhadnout, která environmentální proměnná byla vůči posunu nejcitlivější a ovlivnila tak následně křivku popisující reakci druhu na prostředí. Jako nejcitlivější se vůči posunům projevila orientace buňky vůči světovým stranám a sklon svahu. To může být způsobeno především dvěma faktory. Zaprvé způsobem, jakým byl sklon a orientace odvozen z původního rastru nadmořské výšky a dále skutečností, že obě proměnné byly použity ve stupních a nikoliv v radiánech.

Funkce odvození sklonu a orientace byla vypočítána pro konkrétní buňku ze sousedních osmi buněk (*Borrough et McDonell, 1998*), což je důvodem proč sousední buňky obou rastrů nemusí vykazovat lineární trend. I malým posunem bodu o několik buněk lze v případě rastru s vysokým rozlišením odečíst úplně odlišné hodnoty, s rostoucí velikostí buňky pak klesá spád sklonu (*Thompson et al., 2001*) a dojde ke snížení variability v rastru, což má za následek odečtení méně odlišné hodnoty v případě nepřesné lokalizace. Tímto fenoménem se zabýval např. *Smith et al. (2006)*, který řešil vliv rozlišení digitálního výškového modelu na heterogenitu odvozených rastrů sklonu či orientace a jejich ekologickou relevanci ve vztahu k půdním průzkumům. Pro každé zájmové území lze nalézt určité množství sousedních buněk, ze kterých je možné sklon nebo orientaci vypočítat tak, aby byla heterogenita rastru ve vztahu k experimentu relevantní. Pokud je tento fakt při odvozování rastrů vyjadřujících vlastnosti terénu přehlédnut, může ovlivnit následnou ekologickou interpretaci. Nicméně funkce Aspect a Slope v programu ArcGIS neumožňují zvolit množství sousedních buněk, ze kterých je výsledná hodnota buňky vypočtena. Tomuto problému je možné v dalších studiích předejít odvozením rastrů v software, který umožňuje nastavení počtu sousedních buněk vstupujících do výpočtu, jako je např. SoLIM (*Burt et Zhu, 2004*).

Vyjádřením orientace ve stupních mohlo dojít k situaci, kdy buňky orientované na západ mohly nabývat jak hodnot blízkých nule, tak hodnot blízkých 360°, model by tedy takové buňky vyhodnotil jako naprosto odlišné, ačkoliv jejich orientace je stejná. Navzdory tomuto jevu je orientace ve stupních v modelování druhové distribuce běžně používána (*Kosicki, 2017; Li et al., 2017*). Řešením může být převedením orientace ve stupních na metriky Northness (severnost), která se vypočítá jako cosinus orientace ve stupních a Eastness (východnost) kterou lze vypočítat podobně za pomoci funkce sinus. Buňky vyjadřující orientaci vůči světovým stranám díky této transformaci nabývají hodnot od 1 reprezentující sever (východ) po -1 vyjadřující jih (západ) (*Wilson et al., 2007*).

Chybná lokalizace druhu může v případě velmi heterogenního prostředí způsobit zcela mylné předpoklady o jeho nárocích na prostředí, což bylo demonstrováno zpětným odvozením křivek popisujících závislost pravděpodobnosti výskytu druhu na environmentální proměnné pro jednotlivé varianty posunu. Chyba v lokalizaci virtuálního druhu způsobila zejména u proměnných orientace a sklon zvětšení intervalu hodnot, které druh toleruje. K podobnému výsledku došel i *Mitchell et al. (2017)*, který obdobný experiment provedl s daty reálného druhu na rastrech s rozlišením buňky 1m² pro environmentální proměnné v rámci kterých byl zahrnuta i orientace vůči světovým stranám odvozena stejně jako v této studii pomocí software ArcGIS.

Výsledná hodnota pravděpodobnosti výskytu druhu v konkrétní buňce je dána vztahem mezi environmentálními proměnnými. Vzhledem k tomu, že pravděpodobnost výskytu druhu v buňce byla dána interakcí všech environmentálních proměnných, došlo především v případě variant B a C k tomu,

že posunutá data environmentálních proměnných model silně nadhodnotily a výskyt druhu byl tak predikován i v méně pravděpodobných lokalitách. Ke stejnému závěru došel i *Fukuda et Baets (2016)*, méně strmá křivka druh – prostředí a interakce mezi proměnnými způsobila nadhodnocení zpětně odvozeného rastru vhodnosti prostředí i v jejich studii.

Tvar křivky je částečně ovlivněn i zvoleným modelovacím algoritmem. Metody jako BRT nebo RF mohou sice v některých případech data fitovat lépe, ale výsledná křivka složená z polynomu může být obtížně interpretovatelná oproti GLM (nebo MaxEnt), který hodnoty prokládá křivkou s hladkým průběhem (*Elith et Graham, 2009*). Námětem pro navazující výzkum by tak mohlo být využití dat o výskytu virtuálního druhu použitých v této práci k porovnání vlivu chybné detekce na průběh křivek druh - prostředí různými modelovacími algoritmy, jako to ve své studii provedla např. *Graham et al. (2008)* s reálnými daty o výskytu druhů. Potvrdil by se tak závěr její práce, že ne všechny modely reagují na chyby v lokalizaci druhu stejně, např. modely BRT a MaxEnt vykazovaly větší pokles prediktivní schopnosti než regresní modely, jsou tedy k chybné lokalizaci citlivější než GLM.

Základní analýza dat, v rámci které byly environmentální proměnné vyneseny do histogramu znázorňující četnosti jevu, potvrdila, že kromě proměnné vegetace, nesplňují ostatní proměnné podmínku lineárního trendu v rozdělení dat. Z toho důvodu byla data fitována generalizovaným lineárním modelem s kvadratickým členem pro nadmořskou výšku, sklon a orientaci. Model s kvadratickým členem vykazuje oproti základnímu modelu mnohem více vysvětlené variability ve všech variantách a zároveň dosahuje vyšších hodnot AUC, což bylo důvodem pro volbu tohoto modelu při dalších validačních testech.

Vysvětlená variabilita modelem klesala dle očekávání od varianty 0 po variantu C, důvodem pro tento jev může být zvýšení variability v datech a skutečnost, že pro všechny varianty byl použit GLM ve stejném tvaru, ačkoliv např. u orientace došlo u varianty C k lineárnímu trendu v rozložení dat a tedy jednoduchý lineární člen by data vystihl lépe než kvadratický. V souvislosti s vypočítanou hodnotou AUC, která se pohybovala v rozmezí od 0.99 u varianty 0 po 0.82 u varianty C, lze konstatovat, že i model, který poměrně špatně fituje data dosahuje velmi dobré diskriminační schopnosti. *Pearce et Ferrier (2000)* uvádějí, že modely dosahující hodnot AUC vyšších než 0.75 jsou obecně přijímány jako dostačující pro potřeby konzervační biologie. V tomto případě způsobila chybná lokalizace druhu až o 50 metrů pokles AUC o 18% na hodnotu 0.81, model pro variantu C by tedy byl brán jako aplikovatelný v praxi navzdory skutečnosti, že křivky popisující vztah mezi druhem a environmentální proměnnou či zpětně odvozený rastr pravděpodobnosti výskytu druhu se značně liší od virtuální reality. Ke stejnému závěru došli i *Osborne et Leitão (2009)*. Ačkoliv jejich modely vytvořené na základě dat s chybou v lokalizaci byly ze statistického hlediska dobré, jejich ekologická interpretace by byla chybná díky změně v rozložení četnosti environmentálních proměnných, které byly vybrány k modelování druhu.

Graham et al. (2009) dále upozorňuje na relativní změnu hodnoty AUC v případě modelů s původními a posunutými daty. V případě že model s původními daty dosahuje nízké hodnoty AUC (blíží se 0.5) rozdíl mezi tímto a modelem s posunutými daty se projeví nižším poklesem AUC než v případě velmi přesného modelu s vysokou hodnotou AUC. To lze potvrdit i z výsledků této práce. Rozdíl v poklesu hodnot v případě modelu s kvadratickým členem, který dosahoval u varianty 0 hodnoty AUC 0.99, byl oproti modelu bez kvadratického členu, který vykazoval v případě varianty 0 AUC 0.69, téměř desetinásobný.

Diskriminační schopnost modelu byla dále porovnána Pearsonovým korelačním koeficientem. Stejně jako v případě AUC i zde byl splněn předpoklad a nejvyšší podobnosti s virtuální realitou dosáhla varianta 0. Nejsilněji korelovaný vztah nastal mezi variantou B a C a zároveň jsou tyto dvě varianty velmi málo korelované s virtuální realitou. Velmi nízké hodnoty Pearsonova korelačního koeficientu oproti hodnotám AUC v případě porovnání variant B a C s virtuální realitou lze vysvětlit tak, že pro výpočet hodnot AUC byla použita vzorkovaná data, oproti tomu byl korelační koeficient vypočítán pro celé území zpětně odvozených rastrů. Z toho lze usuzovat, že provedené vzorkování virtuálního druhu nebylo dostatečně reprezentativní.

Navzdory velmi častému využití AUC jakožto validační metody SDM např. *Lobo et al. (2007)* nedoporučuje její použití, protože AUC nebere v potaz, do jaké míry dokáže model data vystihnout. *Lawson et al. (2014)* proto dále doporučují zahrnout do procesu validace více metrik závislých na prevalenci jako je posouzení modelu z hlediska AIC nebo právě vysvětlené variability.

Kalibrační schopnost modelu byla testována pomocí ukazatele Kappa, což je metoda závislá na prahové hodnotě. Vysoká hodnota Kappa pro variantu 0 pro tři zvolené prahové hodnoty značí, že model v tomto případě velmi dobře vystihl virtuální realitu. Se zvyšující se prahovou hodnotou však kalibrační schopnost modelu klesá. To může být způsobeno tím, že i na lokalitách s vysokou pravděpodobností výskytu bude zaznamenána absence. *Miller (2010)* proto doporučuje zvolit prahovou hodnotu blízkou prevalenci druhu. Pokud bychom vztáhli hodnoty Kappa k senzitivitě, což je poměr správně klasifikované presence, je patrné že i při nízkých hodnotách Kappa v případě variant B a C dosahuje model poměrně dobré diskriminační schopnosti, která se jen mírně snižuje s rostoucí prevalencí. *Lawson et al. (2014)* však tvrdí, že validace založená na binárních metrikách často zvýhodní modely, které jsou buď nad, nebo podhodnocené a snižují na rozdíl od kontinuálních metrik diskriminační schopnost modelu.

Porovnání výsledků této práce s nastudovanou literaturou, která řešila vliv chybné lokalizace na přesnost modelů druhové distribuce, bylo obtížné z několika důvodů. Zaprvé existuje velká variabilita vstupních dat a modelovacích algoritmů i možností, jak prediktivní schopnost modelu posoudit. Jak bylo uvedeno v kapitole 2.4.1 „Data o výskytu druhu“, mohou být např. presenční či presenčně – absenční. *Graham et al. (2008)* uvádí, že vliv nepřesné lokalizace má větší dopad na modely, které využívají pouze presenční data, která na rozdíl od presenčně – absenčních dat nenesou informaci o svém okolí. Některé studie zkoumaly vliv chybné lokalizace s reálnými daty, u kterých existuje předpoklad zatížení chybou již v průběhu jejich pořizování z důvodů uvedených v kapitole 2.4.2. „Chyby v datech o výskytu druhu“. Studie se také lišily v rozlišení vstupních vrstev, velikosti posunu, použitých environmentálních proměnných či heterogenitě prostředí. Přesto se v některých obecných závěrech s touto prací shodují: (1) s rostoucí chybou v lokalizaci druhu klesá predikční schopnost modelu; (2) ačkoliv model založený na datech s chybnou lokalizací dosahuje vysokých hodnot AUC, jeho schopnost shody s realitou může být velmi nízká (3) a na chybu způsobenou nepřesnou lokalizací má vliv heterogenita prostředí.

Tato studie ukázala, jaký dopad může mít chybná lokalizace druhu, zejména v případě velmi heterogenního prostředí s vysokým rozlišením vstupních vrstev, kdy dojde k ovlivnění křivky popisující nároky druhu na environmentální proměnnou. Ačkoliv všechny varianty dosahují vysoké diskriminační přesnosti, jejich kalibrační schopnost demonstrována v této práci AIC, vysvětlenou variabilitou a hodnotou Kappa s rostoucí chybou v lokalizaci značně klesá. Toto téma by bylo vhodné podrobit

dalšímu zkoumání, např. testovat vliv chybné lokalizace pro různě heterogenní prostředí, či fitovat data jiným modelovacím algoritmem. Prostor pro další bádání také nabízí testování na reálných druzích, např. porovnání chyby jaké se lze dopustit při špatné lokalizaci euryvalentního druhu oproti specialistovi.

6. Závěr

Tato práce se zabývá problematikou nepřesné lokalizace druhu a jejími dopady na modely druhové distribuce při použití vstupních vrstev s vysokým rozlišením.

Z výsledků této práce je možné vyvodit několik závěrů. Podařilo se prokázat, že chybná lokalizace ve výskytu druhu má vliv na přesnost modelů druhové distribuce. Chybná lokalizace může vést ke zcela mylné představě o nárocích druhu na prostředí, což může hrát velkou roli např. při stanovení managementu chráněných druhů nebo při odhadování chování invazního druhu. Vzhledem ke skutečnosti, že chyba v lokalizaci druhu byla mezi variantou B a C až pětinasobná, avšak jejich diskriminační a kalibrační schopnost podobná, lze tvrdit, že v tomto případě nezpůsobila absolutní hodnota vzdálenosti mezi skutečným místem výskytu a chybnou lokalizací úměrný pokles prediktivní schopnosti modelu. Výsledky práce dále potvrdily, že nejčastěji používaná validační metrika AUC je pro posouzení výkonosti modelu nedostačující, neboť i model založený na datech s maximální chybou 50 m dosahoval hodnoty AUC vyšší, než je obecně přijímaná hranice pro použití v konzervační biologii, ačkoliv analýza vstupních dat ukázala, že křivka popisující vztah mezi druhem a environmentální proměnnou má zcela jiný průběh.

Téma této práce by bylo vhodné dále rozvést a ověřit tak závěry z této práce plynoucí. Lze otestovat, jak se chyba v lokalizaci druhu projeví u prostředí s různým stupněm heterogenity, jaký vliv bude mít na výsledný model velikost vzorku a hodnota prevalence či jak se model zachová při použití dat o výskytu reálného druhu. Další prostor pro zkoumání nabízí rovněž porovnání prediktivní schopnosti modelů založených na datech s chybnou lokalizací napříč různými modelovacími algoritmy.

7. Přehled literatury a použitých zdrojů

- AUSTIN M. P.**, 1980: Searching for a model for use in vegetation analysis. *Vegetatio* 42: 11 – 21.
- AUSTIN M. P.**, 1999: A silent clash of paradigms: some inconsistencies in community ecology. *Oikos* 86: 170 – 178.
- AUSTIN M. P.**, 2002: Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101 – 118.
- AUSTIN M. P., BELBIN L., MEYERS J. A., DOHERTY M. D., LUOTO M.**, 2006: Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling* 199: 197 – 216.
- BEGON M., HARPER J. L., TOWNSEND C. R.**, 1990: *Ecology: individuals, populations and communities*. Blackwell Scientific Publications. 949 s.
- BOCEDI G., PALMER S. C. F., PEER G., HEIKKINEN R. K., MATSINOS Y. G., WATTS K., TRAVIS J. M. J.**, 2014: RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Methods in Ecology and Evolution* 5: 388 – 396.
- 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.
- BRADLEY B. A., OLSSON A. D., WANG O., DICKSON B. G., PELECH L., SENSIE S. E., ZACHMANN L. J.**, 2012: Species detection vs. Habitat suitability: are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling* 244: 57 – 64.
- BURROUGH P. A., MCDONELL R. A.**, 1998: *Principles of Geographical Information Systems*. Oxford University Press: 190 s.
- BURT J. E., ZHU A. X.**, 2004: 3dMapper 4.02. online: www.terrainanalytics.com
- DEBLAUWE V., DROISSART V., BOSE R., SONKÉ B., BLACH-OVERGAARD A., SVENNING J. C., WIERINGA J. J., RAMESH B. R., STÉVART T., COUVREUR T. L. P.**, 2016: Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography* 25: 443 – 454.
- DEMEK J., MACKOVČIN P., BLATKA B., BUČEK A., CIBULKOVÁ P., CULEK M., ČERMÁK P., DOBIÁŠ D., HAVLÍČEK M., HRÁDEK M., KIRCHNER K., LACINA J., PÁNEK T., SLAVÍK P., VAŠÁTKO J.**, 2006: Hory a nížiny. *Zeměpisný lexikon ČR. 2. upravené vydání. MŽP ČR*: 582 s.
- 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.
- ELITH J., GRAHAM C. H.**, 2009: Do they? How they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66 – 77.
- ELITH J., GRAHAM C. H., ANDERSON R. P., DUDÍK M., FERRIER S., GUISAN A., HIJMANS R. J., HUETTMANN F., LAETHWICK J. R., LEHMANN A., LI J., LOHMANN L. G., LOISELLE B. A., MANION G., MORITZ C., NAKAMURA M., NAKAZAWA Y., OVERTON J McC., PETERSON A. T., PHILLIPS S. J.**,

RICHARDSON K., SCACHETTI-PEREIRA R., SCHAPIRE R. E., SOBERÓN J., WILLIAMS S., WISZ M. S., ZIMMERMANN N. E., 2006: Novel methods improve prediction of species' distribution from occurrence data. *Ecography* 29: 129 – 151.

FRANKLIN J., WEJNERT K. E., HATHAWAY S. A., ROCHESTER C. J., FISHER R. N., 2009: Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. *Diversity and Distributions* 15: 167 – 177.

FRANKLIN J., 2010: Mapping species distribution spatial inference and prediction. Cambridge University Press. 340 s.

FRANKLIN J., MILLER J., 2010: Statistical methods – modern regression. In: **FRANKLIN J.** (Ed.): Mapping species distribution: Spatial inference and prediction. Cambridge University Press: 340 s.

FOURCADE Y., 2016: Comparing species distribution modelled from occurrence data and from expert – based range maps. Implication for predicting range shifts with climate change. *Ecological Informatics* 36: 8 – 14.

FUKUDA S., DE BAETS B., 2016: Data prevalence matters when assessing species' responses using data-driven species distribution models. *Ecological Informatics* 32: 69 – 78.

GOODCHILD M. F., 2010: Scale in GIS. *Geomorphology* 130: 5 – 9.

GRAFAREND W. E., 2006: Linear and nonlinear models fixed effect, random effect and mixed models. De Gruyter: 752 s.

GRAHAM C. H., ELITH J., HIJMANS R. J., GUISAN A., PETERSON A. T., LOISELLE B. A., 2008: The influence of spatial errors in species occurrence data in distribution models. *Journal of Applied Ecology* 45: 239 – 247.

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 & Evolution* 19: 497 – 503.

GU W., SWIHART R. K., 2003: Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation* 116: 195 – 203.

GUILLERA-ARROITA G., LAHOZ-MONFORT J. J., ASCELIN GORDON J. E., KUJALA H., LENTINI P. E., MCCARTHY M. A., TINGLEY R., WITNLE 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., THEURILLAT J. P., KIENAST F., 1998: Predicting the potential distribution of plant species in an Alpine environment. *Journal of Vegetation Science* 9: 65 – 74.

GUISAN A., THUILLER W., 2005: Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993 – 1009.

GUISAN A., ZIMMERMANN N., 2000: Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147 – 186.

- GUISAN A. N., ZIMMERMANN N. E., ELITH J., GRAHAM C. H., PHILLIPS S., PETERSON A. T., 2007:** What matters for predicting the occurrences of trees: techniques, data or species' characteristic? *Ecological Monographs* 77: 615 – 630.
- HOSMER D. W., LEMENSHAW W. S., 2000:** Applied Logistic Regression: Second Edition. John Wiley & Sons, Inc: 357 s.
- HUSTON M. A., 2002:** Introductory essay: critical issues for improving predictions. In: **SCOTT J. M., HEGLUNG P. J., SAMSON F., HAUFLE J., MORRISON M., RAPHAEL M., WALL B.** (Eds.): Predicting species occurrences: issues of accuracy and scale. Island Press: 7 – 21.
- CHEN G., KÉRY M., PLATTNER M., MA K., GARDNER B., 2013:** Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology* 101: 183 – 191.
- JABERG C., GUISAN A., 2001:** Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology* 38: 1169 – 1181.
- JACKSON ST., OVERPECK JT., 2000:** Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology* 26: 194 – 220.
- JIMÉNEZ-VALVERDE A., LOBO J. M., HORTAL J., 2009:** Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distribution* 14: 885 – 890.
- JIMÉNEZ-VALVERDE A., LOBO J. M., HORTAL J., 2009:** The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology* 10: 196 – 2005.
- JOHNSON J. CH., GILLINGHAM M. P., 2008:** Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection function for woodland caribou. *Ecological Modelling* 213: 143 – 155.
- KIENAST F., BOLLINGER J., ZIMMERMANN N. E., 2012:** Advanced Landscape Ecology: Lecture Notes. online:www.researchgate.net/file.PostFileLoader.html?id=5730d34496b7e4bb46356f31&assetKey=AS%3A359897272209408%401462817604512, 16 s.
- KOSICKI J., 2017:** Should topographic metrics be considered when predicting species density of birds on a large geographical scale? A case of Random Forest approach. *Ecological Modelling* 10: 76 – 85.
- LAWSON R. C., HODGSON J. A., WILSON R. J., RICHARDS S. A., 2014:** Prevalence, thresholds and performance of presence – absence models. *Methods in Ecology and Evolution* 5: 54 – 64.
- LEATHWICK J. R., ROWE D., RICHARSON J., ELITH J., HASTIE T., 2005:** Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology* 50: 2034 – 2052.
- LEROY B., MEYNARD C. N., BELLARD C., COURCHAMP R., 2015:** virtualspecies: an R package to generate virtual species distributions. *Ecography* 39: 599 – 607
- LOBO J. M., JIMÉNEZ-VALVERDE A., REAL R., 2007:** AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145 – 151.

- McMASTER R. B., SHEA K. S.,** 1992: Generalization in digital cartography. *Association of American Geographers*, 137 s.
- McPEARSON J. M., WALTER J.,** 2007: Effects of species' ecology on the accuracy of distribution models. *Ecography* 30: 135 – 151.
- MEYNARD CH. N., KAPLAN D. M.,** 2013: Using virtual species to study species distributions and model performance. *Journal of Biogeography* 40: 1- 8.
- MILLER J. A.,** 2010: Species Distribution Modeling. *Geography Compass* 4: 490 – 509.
- MILLER J. A.,** 2014: Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Progress in Physical Geography* 38: 117 – 128.
- MITCHELL P. J., MONK J., LAURENSEN L.,** 2017: Sensitivity of fine-scale species distribution models to location uncertainty in occurrence data across multiple sample sizes. *Methods in Ecology and Evolution* 8: 12 – 21.
- MURPHY A., WINKLER R.,** 1987: A general Framework for forecast verification. *Monthly Weather Review* 115: 1330 – 1338.
- MOUDRÝ V.,** 2015: Modelling species distributions with simulated virtual species. *Journal of Biogeography* 42: 1365 – 1366.
- MOUDRÝ V., ŠÍMOVÁ P.,** 2012: Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science* 26: 2083 – 2095.
- OSBORNE P. E., ALONSO J. C., BRYANT R. G.,** 2001: Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology* 38: 458 – 471.
- OSBORNE P. E., LEITAO P. J.,** 2009: Effect of species and habitat positional errors on the performance and interpretation of species distribution models. *Biodiversity research* 15: 671 – 681.
- PEARCE J., FERRIER S.,** 2000: Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133: 225 – 245.
- PETERSON A. T.,** 2003: Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology* 78: 419 – 433.
- POCHEVILLE A.,** 2015: The Ecological Niche: History and Recent Controversies. In: **HEAMS TH., HUNEMAN P. LECOINTRE G., SIBERSTEIN M.** (Eds.): *Handbook of Evolutionary Thinking in the Science*. Springer, Netherlands: 547 – 586.
- RISSLER L. J., HIJMANS R. J., GRAHAM C. H., MORITZ C., WAKE D. B.,** 2006: Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *American Naturalist* 167: 655 – 666.
- SEGURADO P., ARAÚJO M.,** 2004: An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31: 1555 – 1568.

- SCHUMAKER N. H.**, 2015: HexSim version 3.0. Environmental Protection Agency, online: www.hexsim.net
- SCHWARTZ O. (Ed.)**, 2010: Plán péče – Krkonošský národní park a jeho ochranné pásmo (2010 – 2020), online: <http://www.krnap.cz/plan-pece>, 179 s.
- SMITH M. P., ZHU A. X., BURT J. E., STILES C.**, 2006: The Effect of DEM resolution and neighborhood size on digital soil survey. *Geoderma* 137: 58 – 69.
- SOBERÓN J.**, 2007: Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115 – 1123.
- SOBERÓN J., NAKAMURA M.**, 2009: Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19644 – 1965.
- SVENNING J. C., SKOV F.**, 2004: Limited filling of the potential range in European tree species. *Ecology Letters* 7: 565 – 573.
- ŠMILAUER P.**, 2007: Moderní regresní metody. Biologická fakulta Jihočeské university: 168 s.
- THOMPSON J. A., BELL J. C., BUTLER C. A.**, 2001: DEM resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma* 100: 67 – 89.
- THUILLER W., ARAÚJO M., LAVOREL S.**, 2004: Do we need land-cover data to model species distribution in Europe? *Journal of Biogeography* 31: 353 – 361.
- TSOAR A., ALLOUCHE O., STEINITZ O., ROTEM D., KADMON R.**, 2007: A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions* 13: 397 – 405.
- VÁCLAVÍK T., MEENTENMEYEER R. K.**, 2012: Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distribution* 18: 73 – 83.
- VARELA S., ANDERSON R. P., GARCÍA-VALDÉS R., FERNÁNDEZ-GONZÁLES F.**, 2014: Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37: 1084 – 1091.
- VAUGHAM I. P., ORMEROD S. J.**, 2003: Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conservation Biology* 17: 1601 – 1611.
- VAUGHAM I. P., ORMEROD S. J.**, 2005: The continuing challenges of testing species distribution models. *Journal of Applied Ecology* 42: 720 – 730.
- WILSON M. F. J., O'CONNEL B., BROWN C., GUINAN J. C., GREHAN A.**, 2007: Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. *Marine Geodesy* 30: 3 – 35.
- WISZ M. S., HIJMANS R. J., LI J., PETERSON A. T., GRAHAM C. H., GUISAN A.**, 2008: Effects of sample size on the performance of species distribution models. *Diversity and Distribution* 14: 763 – 773.

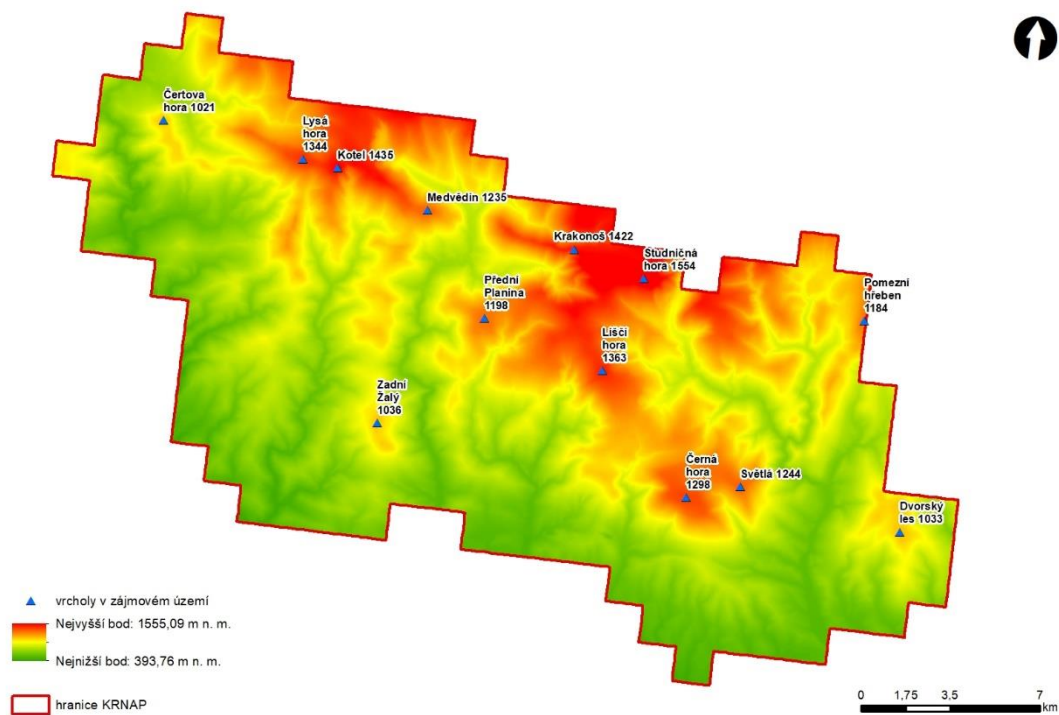
LI X., SI Y., JI L., GONG P., 2017: Dynamic response of East Asian Greater White-fronted Geese to changes of environmental during migration: Use of multitemporal species distribution model. *Ecological Modelling* 24: 70 – 79.

YACKULIC C. B., CHANDLER R., ZIPKIN E. F., ROYLE J. A., NICHOLS J. D., GRANT E. H. C., VERAN S., 2013: Presence-only modelin using MAXENT: when can we trust the infereces? *Methods in Ecology and Evolution* 4: 236 – 243.

ZVÁRA K., 2008: Regrese. *Matematicko-fyzikální fakulta University Karlovy*: 253 s.

8. Přílohy

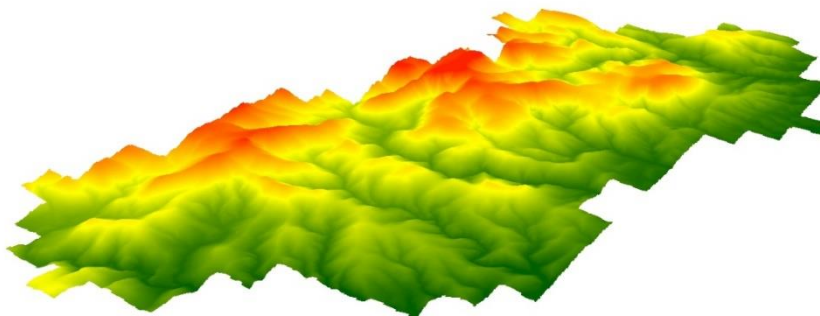
8.1. Grafické přílohy



Obrázek přílohy 1: Digitální model terénu KRNP s nejvyššími vrcholy v území

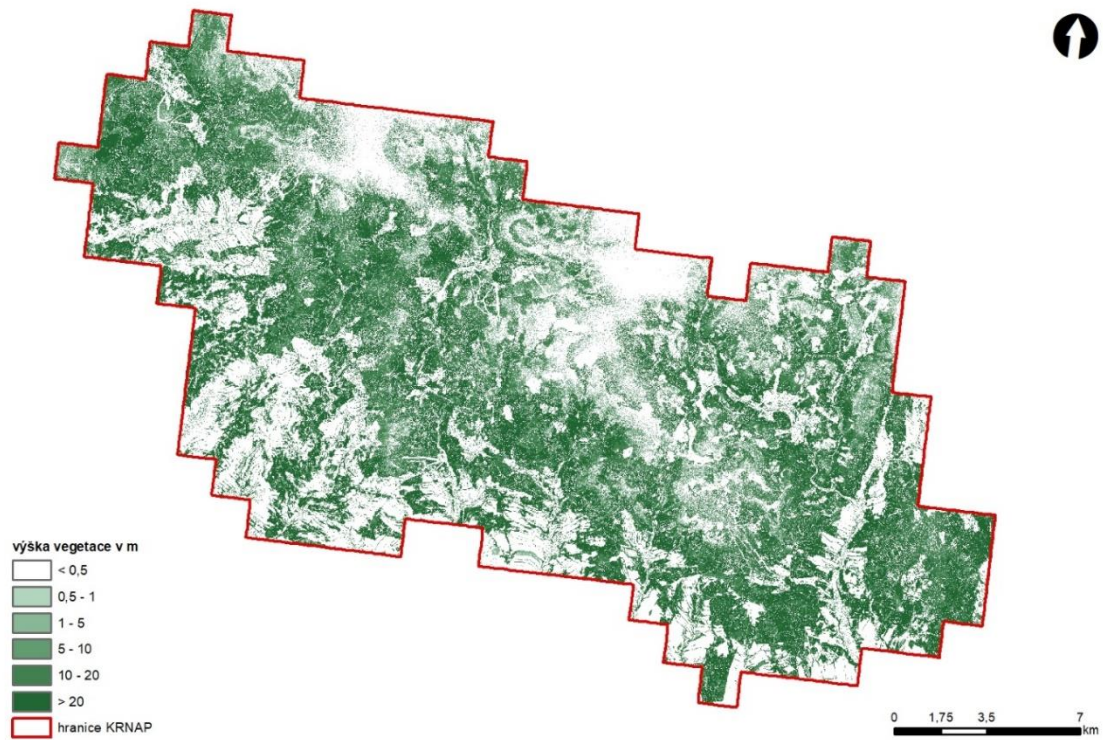
(DTM: autor, 2017;

vrcholy v zájmovém území: ARCDATA PRAHA s. r. o., 2017)



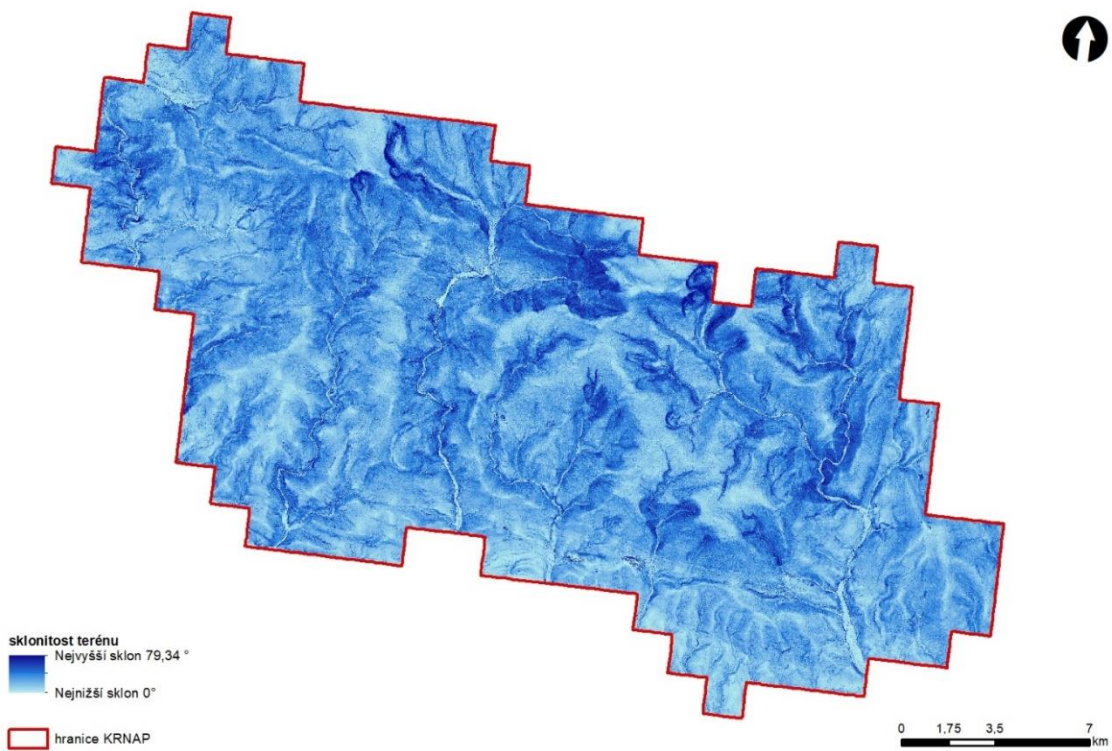
Obrázek přílohy 2: 3D Model území KRNP, pro lepší názornost byly výšky vynásobeny 1,5 krát.

(3D model KRNP: autor, 2017)



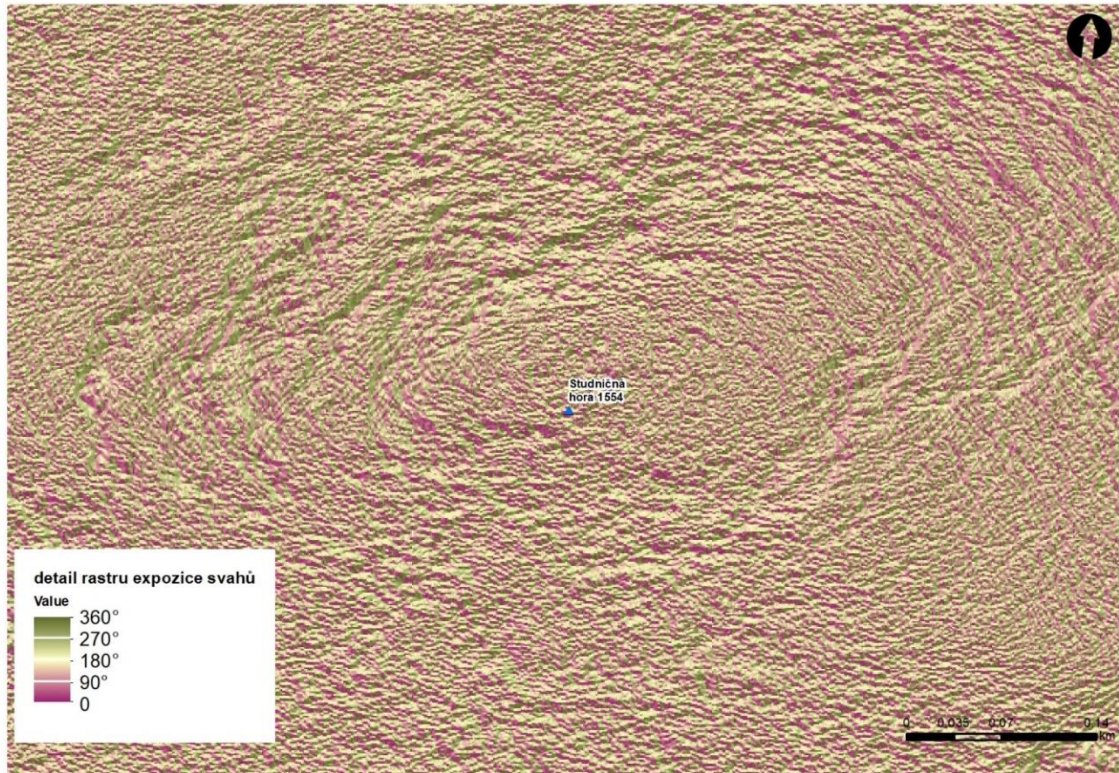
Obrázek přílohy 3: Výška vegetace na území KRNP

(výška vegetace: autor, 2017)



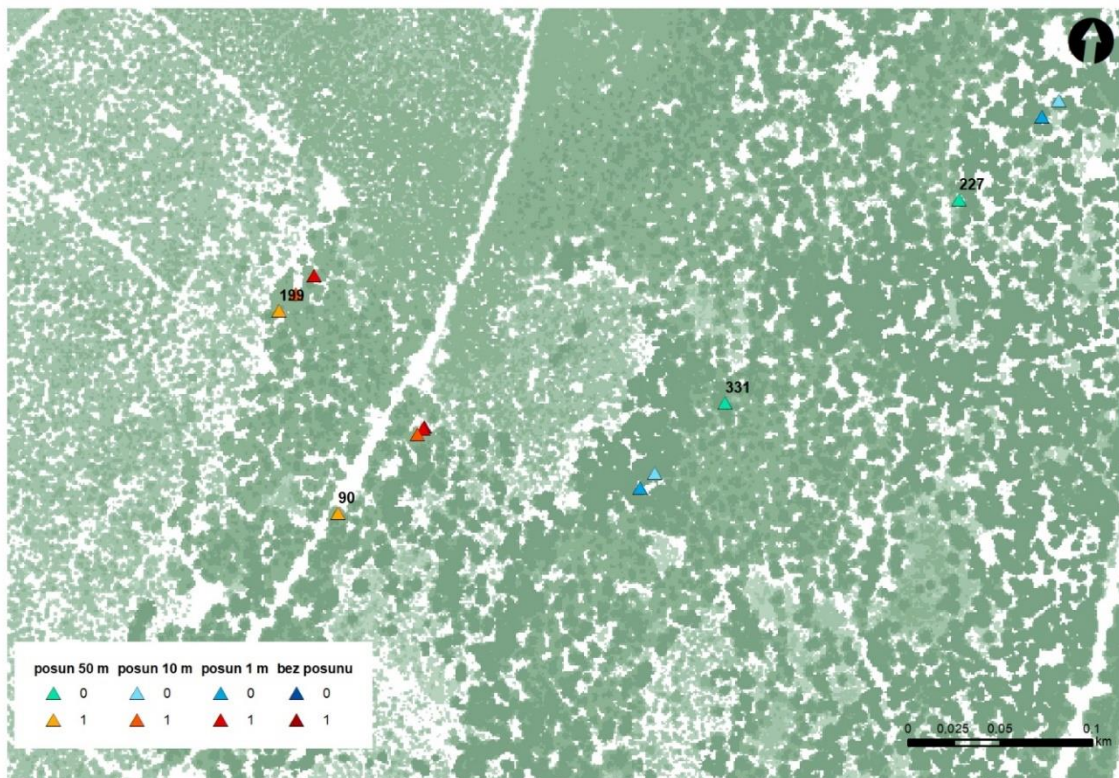
Obrázek přílohy 4: Sklonitost terénu ve stupních na území KRNP

(sklonitost terénu: autor, 2017)



Obrázek přílohy 5: Detail rastru expozice terénu vůči světovým stranám

(expozice terénu: autor, 2017)



Obrázek přílohy 6: Detail posunu presenčně-absenčních dat

(autor, 2017)

8.2. Části kódu dle jednotlivých kroků v jazyku R

```
3 library(raster)
4 library(sp)
5 library(virtualspecies)
6 library(rgdal)
7 setwd("c:\\Users\\sko1a\\Desktop\\dp\\data.tif")
8 dtm = "krnap_dtm_1m_clip.tif"
9 vegetace = "vegetace_clip.tif"
10 sklon = "sklon_stupne1.tif"
11 orientace = "aspekt.tif"
12 prostredi = stack(dtm, sklon, orientace, vegetace)
13 names(prostredi) = c("dtm", "sklon", "orientace", "vegetace")
14 parametry = formatFunctions(dtm = c(fun = "dnorm", mean = 950, sd = 200),
15                             vegetace = c(fun = "linearFun", a = -0.05, b = 8),
16                             sklon = c(fun = "dnorm", mean = 15, sd = 5),
17                             orientace = c(fun = "dnorm", mean = 175, sd = 60))
18 virtualdruh = generateSpFromFun(raster.stack = prostredi, parameters = parametry, plot = TRUE)
```

Obrázek přílohy 7: Definování virtuálního druhu

(autor, 2017)

```
22 a = (1:100)
23 for(i in a){
24   if (i == 1)
25     {print(i)}
26   PARAstr = convertToPA(virtualdruh, beta = 0.5, alpha = -0.05, plot = TRUE)
27 }
```

Obrázek přílohy 8: Konverze rastru vhodnosti prostředí na presenčně-absenční rastr dle pravděpodobnostního přístupu logistické funkce

(autor, 2017)

```
28 vzorkovani = sampleOccurrences(PARAstr, n = 400,
29                               type = "presence-absence",
30                               sample.prevalence = 0.5, plot = TRUE)
31 x = vzorkovani$sample.points$x
32 y = vzorkovani$sample.points$y
33 p = vzorkovani$sample.points$observed
34 bodyXY = SpatialPoints(cbind(x,y))
35 hodnoty=extract(prostredi, bodyXY)
```

Obrázek přílohy 9: Vzorkování výskytu virtuálního druhu z presenčně-absenčního rastru a následné přiřazení hodnot nadmořské výšky, výšky vegetace, sklonu a expozice

(autor, 2017)

```
37 a = floor(runif(100, min = -1, max = +1))
38 pax = x+a
39 pay = y+a
40 b = floor(runif(100, min = -10, max = +10))
41 pbx = x+b
42 pby = y+b
43 c = floor(runif(100, min = -50, max = +50))
44 pcx = x+c
45 pcy = y+c
46 write.table(cbind(x,y,hodoty,p, "\t", pax, pay, posuna, "\t", pbx, pby, posunb, "\t",
47                 pcx, pcy, posunc), "hodnoty.csv", col.names = TRUE, sep = ";", dec = ".", append = FALSE)
```

Obrázek přílohy 10: Náhodný posun dat výskytu virtuálního druhu v rámci intervalů

(autor, 2017)

```

1 datB = read.delim("clipboard", header = T)
2 datB
3 attach(datB)
4 names(datB)
5 #ROZDĚLENÍ DAT 70% TRAINING 30% TESTING (pro všechna data)
6 splitdf = function(dataframe, seed = NULL){
7   if(!is.null(seed)) set.seed(seed)
8   index = 1:nrow(dataframe)
9   trainindex = sample(index, trunc(length(index)*7/10))
10  trainset = dataframe[(trainindex), ]
11  testset = dataframe[-(trainindex), ]
12  list(trainset = trainset, testset=testset)
13 }
14 splits=(splitdf(datB))
15 str(splits)
16 Train=splits$trainset
17 Test=splits$testset

```

Obrázek přílohy 11: Rozdělení dat o výskytu druhu na trénovací a testovací sadu

(autor, 2017)

```

1 #GLM BETA RESPONSE
2 Model1 = glm(p~dtm+I(dtm^2)+sklon+I(sklon^2)+orientace+I(orientace^2)+
3           vegetace, family = binomial, data = Train)
4 summary(Model1)
5 #DEVIANCE
6 D1=(Model1$null.deviance-Model1$deviance)/Model1$null.deviance
7 n1=length(Model1$fitted.values)
8 p1=length(Model1$coefficients)
9 D11=1-((n1-1)/(n1-p1))*(1-D1);D11
10 #PREDICKE
11 predict1=predict(Model1, newdata=Test,type="response")
12 predicted1=as.vector(predict1)
13 #AUC
14 library(PresenceAbsence)
15 SET1=data.frame(Test$id, Test$p, predicted1)
16 AUC1=auc(SET1, st.dev=TRUE, which.model=1, na.rm=FALSE);AUC1

```

Obrázek přílohy 12: generalizovaný lineární model, výpočet vysvětlené variability, predikce vhodnosti prostředí na základě modelu, validace metodou AUC

(autor, 2017)

```

1 sequenceForvegetace = seq(0,53,0.2)
2 for(i in a){
3   if (i == 1){
4     NEW_Test_dat=data.frame(cbind(rep(907.4), rep(14.89), rep(176.81),sequenceForvegetace))
5     NEW_Test_dat=setNames(NEW_Test_dat, c("dtm", "sklon", "orientace", "vegetace"))
6     predicti=predict(Model1, newdata =NEW_Test_dat, type = "response" )
7     predicted=as.vector(predicti)
8     NEW_Test_dat_EVAL=data.frame(predictedi)
9   }
10  else{
11    NEW_Test_dat=data.frame(cbind(rep(907.4), rep(14.89), rep(176.81),sequenceForvegetace))
12    NEW_Test_dat=setNames(NEW_Test_dat, c("dtm", "sklon", "orientace", "vegetace"))
13    predicti=predict(Model1, newdata =NEW_Test_dat, type = "response" )
14    predicted=as.vector(predicti)
15    NEW_Test_dat_EVAL=cbind(NEW_Test_dat_EVAL, predictedi)
16  }
17 }
18 MEDIAN=apply(NEW_Test_dat_EVAL,1,quantile, probs=c(.50));MEDIAN
19 UPPER=apply(NEW_Test_dat_EVAL,1,quantile, probs=c(.95));UPPER
20 LOWER=apply(NEW_Test_dat_EVAL,1,quantile, probs=c(.05));LOWER
21
22 train_dar1=data.frame(Train$vegetace, Train$suitability_raster)
23 train_dar1=setNames(train_dar1, c("vegetace", "suitability"))
24 Test_dar1_model1=data.frame(NEW_Test_dat$vegetace,MEDIAN, LOWER, UPPER)
25 Test_dar1_model1=setNames(Test_dar1_model1, c("vegetace", "fit", "lower", "upper"))
26 Test_dar1_model1=transform(Test_dar1_model1, vegetace=as.numeric(vegetace),
27                             fit=as.numeric(fit),lower=as.numeric(lower),upper=as.numeric(upper))
28
29 h <- ggplot(train_dar1, aes(x=vegetace,y=suitability))
30 h <- h + geom_line(colour="#4E5152", size = 1, data=Test_dar1_model1, aes(y=fit, x=vegetace))
31 h <- h + geom_line(colour="#ADA115", size = 1, data=Test_dar1_model2, aes(y=fit.1, x=vegetace.1))
32 h <- h + geom_line(colour="#218BA3", size = 1, data=Test_dar1_model3, aes(y=fit.2, x=vegetace.2))
33 h <- h + geom_line(colour="#C4661A", size = 1, data=Test_dar1_model4, aes(y=fit.3, x=vegetace.3))
34 h <- h + labs(y= "pravděpodobnost výskytu", x = "výška vegetace")
35 h <- h + theme_classic()
36 h

```

Obrázek přílohy 13: zpětné odvození křivky popisující vztah mezi environmentální proměnnou a pravděpodobností výskytu v prostředí

(autor, 2017)

```

4 sens1=data.frame(Test$id, Test$p, predicted2)
5 optresh1 <- optimal.thresholds(sens1, opt.methods="Default", which.model = 1)
6 t <- optresh1[,2];t
7 cmx <- cmx(sens1, threshold = t, which.model = 1, na.rm = FALSE);cmx
8 pcc(cmx, st.dev = TRUE)
9 presence.absence.accuracy(sens1, threshold = t, find.auc = TRUE, st.dev = TRUE)

```

Obrázek přílohy 14: chybová matice a odvození sensitivity, specificity a poměru správně klasifikovaných presencí a absencí (PCC)

(autor, 2017)

```

20 # převedení rastru na list hodnot
21 vr_p= rasterToPoints(prostredi$virtual_reality, fun = NULL, spatial = TRUE)
22 TEST3 = data.frame(dtm_p$dtm, sklon_p$sklon, orientace_p$orientace, vegetace_p$vegetace)
23 names(TEST3)=c("dtm", "sklon", "orientace", "vegetace")
24 #odvození pravděpodobnosti vyskytu v kazde bunce
25 predict5=predict(Model1, newdata=TEST3,type="response")
26 predicted5=as.vector(predict5)
27 #korelační matice
28 install.packages("corrplot")
29 library(corrplot)
30 corrdat = data.frame(vr0, predicted5, predicted6, predicted7, predicted8)
31 C = cor(corrdat)
32 p.mat = cor.mtest(corrdat);p.mat
33 res1 = cor.mtest(corrdat, conf.level = .95);res1
34 colnames(C) <- c("VR", "0", "A", "B", "C")
35 rownames(C) <- c("VR", "0", "A", "B", "C")
36 corrplot.mixed(C,p.mat = res1$p,lower.col = "black", upper = "color",
37               insig = "label_sig",sig.level = c(.001, .01,.05), tl.col = "black",
38               tl.srt = 360, pch.cex = 0.9, pch.col = "white" )
39 #export rastru
40 s1 = data.frame(x, y, z)
41 suitab1=rasterFromXYZ(s1)
42 writeRaster(suitab1, filename="suitab_p_0", format="ascii", overwrite=TRUE)

```

Obrázek přílohy 15: Odvození rastrů vhodnosti prostředí dle koeficientů modelů, korelační matice, export odvozených rastrů do formátu ESRI

(autor, 2017)

8.3. Tabulky

Tabulka přílohy 1: Vyhodnocení GLM bez přítomnosti kvadratického členu

GLM				
Varianta	AIC	DEVIANCE	AUC	AUC sd
0) bez pos	34873	0.0746	0.6949	0.0051
A) ± 1 m	34440	0.0861	0.7094	0.0050
B) ± 10 m	35789	0.0503	0.6778	0.0053
C) ± 50	35939	0.0463	0.6655	0.0054
n = 27182				

(autor, 2017)

Tabulka přílohy 2: Výsledky GLM s kvadratickým členem

		β	SE	z value	hladina významnosti
varianta 0	Intercept	1.13E+02	2.12E+00	-53.34	<0.001
	dtm	1.45E-01	3.36E-03	43.02	<0.001
	dtm ²	-7.60E-05	1.80E-06	-42.31	<0.001
	sklon	3.35E+00	8.34E-02	40.18	<0.001
	sklon ²	-1.11E-01	2.76E-03	-40.35	<0.001
	orientace	2.96E-01	6.64E-03	44.54	<0.001
	orientace ²	-8.46E-04	1.89E-05	-44.83	<0.001
	vegetace	-1.52E-01	4.75E-03	-31.95	<0.001
varianta A	Intercept	-5.23E+01	8.48E-01	-61.6	<0.001
	dtm	1.02E-01	1.80E-03	56.56	<0.001
	dtm ²	-5.31E-05	9.58E-07	-55.36	<0.001
	sklon	7.16E-01	1.69E-02	42.47	<0.001
	sklon ²	-2.67E-02	5.93E-04	-44.98	<0.001
	orientace	2.86E-02	7.72E-04	37.08	<0.001
	orientace ²	-8.81E-05	2.16E-06	-40.88	<0.001
	vegetace	-7.79E-02	2.38E-03	-32.73	<0.001
varianta B	Intercept	-4.25E+01	6.75E-01	-62.986	<0.001
	dtm	9.03E-02	1.46E-03	61.748	<0.001
	dtm ²	-4.72E-05	7.81E-07	-60.425	<0.001
	sklon	1.77E-01	8.18E-03	21.634	<0.001
	sklon ²	-5.75E-03	2.24E-04	-25.712	<0.001
	orientace	7.36E-04	5.80E-04	1.269	0.205
	orientace ²	-3.71E-06	1.56E-06	-2.378	0.017
	vegetace	-4.77E-02	1.81E-03	-26.353	<0.001
varianta C	Intercept	-4.06E+01	6.43E-01	-63.065	<0.001
	dtm	8.72E-02	1.40E-03	62.293	<0.001
	dtm ²	-4.56E-05	7.49E-07	-60.885	<0.001
	sklon	9.27E-02	7.14E-03	12.975	<0.001
	sklon ²	-3.14E-03	1.85E-04	-17.011	<0.001
	orientace	-6.05E-04	5.68E-04	-1.066	0.286
	orientace ²	1.55E-06	1.53E-06	1.014	0.311
	vegetace	-3.63E-02	1.73E-03	-21.001	<0.001
n = 27182					

(autor, 2017)