

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA
V PRAZE
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA APLIKOVANÉ GEOINFORMATIKY A
ÚZEMNÍHO PLÁNOVÁNÍ



**Jak ovlivňuje metoda generování virtuálního
druhu modely druhové distribuce?**

BAKALÁŘSKÁ PRÁCE

Vedoucí práce: Ing. Lukáš Gábor

Autor práce: Tereza Hüttnerová

2018

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Tereza Hüttnerová

Územní technická a správní služba

Název práce

Jak ovlivňuje metoda generování virtuálního druhu modely druhové distribuce?

Název anglicky

How the Method of Generating Virtual Species Affects Species Distribution Models?

Cíle práce

Cílem práce je otestovat, jak ovlivňují odlišné metody generování virtuálního druhu modely druhové distribuce.

Formulace dílčích cílů je úkolem autora.

Metodika

Modely druhové distribuce (Species Distribution Models, dále jen SDMs) umožňující mapovat a monitorovat rozšíření živočichů či rostlin představují klíčový nástroj v biogeografii, makroekologii či ochraně přírody. Od počátku jejich využívání došlo k rozvoji mnoha statistických metod a zlepšila se dostupnost biologických a environmentálních dat. Právě data jsou v současnosti jedním ze zásadních limitujících faktorů modelování. Zatímco ověření vlivu kvality prostorových dat na SDMs pomocí reálných dat je obtížné, generováním tzv. virtuálního druhu lze snadno určit, jaký vliv mají například rozdílné metody sběru dat, jejich kvalita nebo použité měřítko environmentálních proměnných na výsledný model. V současnosti se k vytvoření virtuální reality používají dvě odlišné metody. Jedná se o (i) prahovou metodu a (ii) o pravděpodobnostní metodu. V nedávné době bylo navrženo také použití metody nazývané NicheLim. Úkolem autora je zpracovat literární rešerši na téma druhových distribučních modelů a generování virtuálního druhu. V praktické části pak pomocí tří výše uvedených metod vygenerovat virtuální druh a otestovat, jaký vliv má použití odlišných metod na modely druhové distribuce.

Doporučený rozsah práce

30 stran

Klíčová slova

Formulace klíčových slov je úkolem autora

Doporučené zdroje informací

- Huang M. et al. The Niche Limitation Method (NicheLim), a new algorithm for generating virtual species to study biogeography. *Ecological Modelling*, 2016, 320, 197-202.
- Leroy B. et al., Virtualspecies, an R package to generate virtual species distributions. *Ecography*. 2015, 001-009.
- Meynard CH. N., et al., Using virtual species to study species distributions and model performance. *Journal of Biogeography*. 2013, 40(1), 1-8.
- Meynard CH. N. & Kaplan M. D., The effect of a gradual response to the environment on species distribution modelling performance. *Ecography*. 2012, 35(6), 499-509.
- Miller J., Species Distribution Modelling. *Geography Compass*. 2010, 4(6), 490-509.
- Xiaoquan K. et al., Package 'sdmvspecies': Create Virtual Species for Species Distribution Modelling. 2015.
- Zurell D. et al., The virtual ecologist approach: simulating data and observers. *Oikos*. 2010, 119(4), 622-635.

Předběžný termín obhajoby

2017/18 LS – FŽP

Vedoucí práce

Ing. Lukáš Gábor

Garantující pracoviště

Katedra aplikované geoinformatiky a územního plánování

Elektronicky schváleno dne 16. 3. 2018

doc. Ing. Petra Šímová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 18. 3. 2018

prof. RNDr. Vladimír Bejček, CSc.

Děkan

V Praze dne 22. 03. 2018

Prohlášení

Prohlašuji, že jsem bakalářskou práci na téma *Jak ovlivňuje metoda generování virtuálního druhu modely druhové distribuce?* vypracovala samostatně, pod vedením Ing. Lukáše Gábora a že jsem uvedla všechny literární prameny, ze kterých jsem čerpala. Prohlašuji, že tištěná verze se shoduje s verzí odevzdanou přes Univerzitní informační systém.

Ve Vysoké Peci dne 24. 4. 2018

.....

Poděkování

Největší poděkování patří Ing. Lukášovi Gáborovi, vedoucímu bakalářské práce, za řadu podnětných konzultací a cenné rady při zpracování bakalářské práce. Dále bych ráda poděkovala své rodině za podporu při celé době studia.

Abstrakt

Druhové distribuční modely (SDMs), nazývané také jako habitatové modely, slouží k vymezení podmínek prostředí, které umožňují životaschopný výskyt druhu. SDMs jsou založeny na matematicko-statistickém modelování vztahů mezi environmentálními proměnnými a výskytovými daty. Výsledkem SDMs jsou prostorové predikce aktuální či potenciační distribuce druhu, případně indexy určující environmentální vhodnost prostředí. Modely se dají používat například k predikci rozšíření/zániku druhu, určení dopadů klimatického oteplování na výskyt druhů, k predikci šíření invazivních druhů, nalezení lokalit výskytu vzácných druhů anebo zjišťování ekologických nároků druhů. V současnosti představují největší limity SDMs vstupní data a způsob jejich pořízení. Zkoumat vliv dat na modely druhové distribuce s využitím reálných dat je nicméně obtížné. Řešením může být využívání virtuálního druhu díky kterému lze ověřovat modelovací algoritmy a testovat, jak odlišná kvalita vstupních dat ovlivňuje výsledné modely. Virtuální druh můžeme generovat třemi odlišnými metodami: (i) pravděpodobnostní metoda, (ii) prahová metoda a (iii) NicheLim. Výběr jedné z těchto metod může nicméně ovlivnit samotné modely a vést k mylným závěrům jednotlivých studií. V bakalářské práci byla proto virtuální realita vygenerována třemi uvedenými způsoby. Ty pak byly následně použity k vytvoření modelů druhové distribuce s využitím různého sample size, což umožnilo porovnat, zdali mohou ovlivnit jednotlivé metody generování virtuálního druhu výslednou predikci v závislosti na použitém sample size. Zájmovou oblastí bylo zvoleno území České republiky, vstupní environmentální data byla získána z databáze Worldclim v rozlišení 30 s (cca 1 km²). Velikost vzorků (sample size) byla stanoven na 50, 100, 500, 1 000, 2 000. Následně byla každá metoda vyhodnocena pomocí AUC a Schoener's D indexu, nejvyšší shoda byla zaznamenána při komparaci prahové a pravděpodobnostní metody. Nejvyšší výkon modelu byl dosažen při použití NicheLim metody, nejhorsích výsledků bylo dosaženo při použití pravděpodobnostní metody. Z analýzy dále vyplývá, že velikost sample size pozitivně ovlivňuje výkon modelu, lze obecně konstatovat, že čím vyšší sample size, tím vyšší schopnosti model dosahuje.

Klíčová slova: druhové distribuční modely, virtuální druh, prahová metoda, pravděpodobnostní metoda, NicheLim

Abstract

Species distribution models (SDMs), known also as habitat models, serve for the outlining of the environmental conditions enabling viable existence of a species. SDMs are based on mathematical-statistical modelling of relations between environmental variables and occurrence data. The result of SDMs consists in spatial predictions of an actual or potential species distribution, or indexes determining environmental suitability of the habitat. The models are used for prediction of distribution/extinction of a species, determination impacts of climatic warming on species occurrence, for prediction spread of invasive species, to locate rare species, identification of ecological requirements of species. The largest limits of SDMs are represented by input data and by the methods of its acquisition. It is difficult to investigate the impact of data on SDMs with the use of real data. A solution could be the use of a virtual species thanks to which it is possible to verify modelling algorithms and to test how the different quality of input data affects resulting models. A virtual species can be generated by using three different methods: (i) probability method, (ii) threshold method (iii) NicheLim. The selection of one of these methods can, however, affect the actual models and lead to erroneous conclusions of individual studies. For this reason, virtual reality was generated in three above mentioned ways in the Bachelor's Thesis. They were, in turn, subsequently used for creation of species distribution models with the use of different sample sizes, which made it possible to compare whether individual methods of generation of a virtual species can affect the resulting prediction, depending on the sample size used, or not. The study area was Czech Republic, the input environmental data was obtained from the Worldclim database in the resolution of 30 s (approx. 1 km²). The sample size was determined as 50, 100, 500, 1000, 2000. Each method was evaluated with AUC and Schoener's D index, the highest conformity was registered during comparison of the threshold and the probability methods. The highest performance of the model was achieved during the use of the NicheLim method, the worst results were achieved obtained during the use of the probability method. Larger sample size has positive effects on the performance of SDMs.

Key words

species distribution model, virtual species, probabilistic approach, threshold approach, NicheLim

Obsah

1. ÚVOD.....	11
2. LITERÁRNÍ REŠERŠE	13
2.1 Faktor M.....	14
2.2 Faktor B.....	14
2.3 Faktor A.....	14
2.2 Ekologická nika.....	15
2.3 Druhové distribuční modely.....	16
2.4 Vstupní data.....	17
2.4.1 Environmentální.....	18
2.4.2 Výskytová.....	20
2.5 Modelovací algoritmy.....	21
2.6 Validace modelů.....	24
2.7 Virtuální druh.....	27
2.7.2 Software.....	30
2.7.3 Metody vytváření virtuální reality.....	30
3. METODIKA	32
3.1 Charakteristika zájmového území.....	32
3.2 Vstupní data.....	32
3.3 Použitý software.....	33
3.4 Tvorba virtuálního druhu.....	34
3.4.1 Import dat.....	34
3.4.2 Odezvová funkce.....	35
3.4.3 Presence-absence mapy.....	36
3.4.4 Vzorky výskytu.....	37
3.5 Algoritmus Maxent.....	37
3.6 Validáční metriky.....	38

4. VÝSLEDKY	39
5. DISKUZE	43
6. ZÁVĚR	46
7. SEZNAM POUŽITÉ LITERATURY	47
8. SEZNAMY	55
8.1 Obrázky	55
8.2 Tabulky.....	56
8.3 Přílohy	57
9. PŘÍLOHY	58

SEZNAM POUŽITÝCH ZKRATEK

ANN	Artificial neural network (Umělé neuronové sítě)
AUC	Area under the ROC curve (Plocha pod ROC křivkou)
BAM	Biotic, Abiotic, Movement factors (Biotické, abiotické a hybné faktory)
DEM	Digital Elevation Models (Digitální model terénu)
ENFA	Ecological niche factor analysis (Analýza faktorů ekologické niky)
GA	Genetic algorithm (Genetický algoritmus)
GAM	Generalized additive model (Generalizovaný aditivní model)
GIS	Geographic Information System (Geografický informační systém)
GLM	Generalized linear model (Generalizovaný lineární model)
HSI	Habitat Suitability index (Index vhodnosti habitatu)
MARS	Multivariate adaptive regression model (Model mnohorozměrné adaptivní regrese)
Maxent	Maximum entropy model (Model maximální entropie)
PCA	Principal component analysis (Analýza hlavních komponent)
PCC	Percent correctly classified (Procento správně určených presencí a absencí)
ROC	Receiver Operating Characteristic Curve (Operační charakteristická křivka)
SDMs	Species distribution models (Druhové distribuční modely)
Se	Sensitivity (Senzitivita)
Sp	Specificity (Specifická)
TSS	True skill statistic (Statistika skutečných schopností modelu)
vSDMs	Virtual species distribution models (Virtuální druhové distribuční modely)

1. ÚVOD

Druhov \acute{e} distribuční modely (dále jen SDMs) jsou hojn \acute{e} využívány v ekologii a biogeografii s aplikací při ochran \acute{e} druhů, invazivní ekologii či určení dopadů klimatických změn na druhy. SDMs kvantifikují vztahy mezi environmentálními proměnnými a distribucí fauny a flóry (Franklin & Miller, 2009). V současné době existuje široká škála modelovacích algoritmů sloužících k vytvoření SDMs. Volba vhodného algoritmu je obtížná a objektivně nelze určit, jaká metoda přináší nejpřesnější výsledky. Výběr algoritmu závisí na použitých vstupních datech, cílech studie a požadovaných výstupech (Miller, 2010). Algoritmy lze rozdělit na modely pracující s presence-only a presence-absence daty. Absenční data jsou mnohdy nedostupná a jejich kvalita není zaručena. Do popředí zájmu se tedy dostávají algoritmy pracující s presence-only daty, například Maxent.

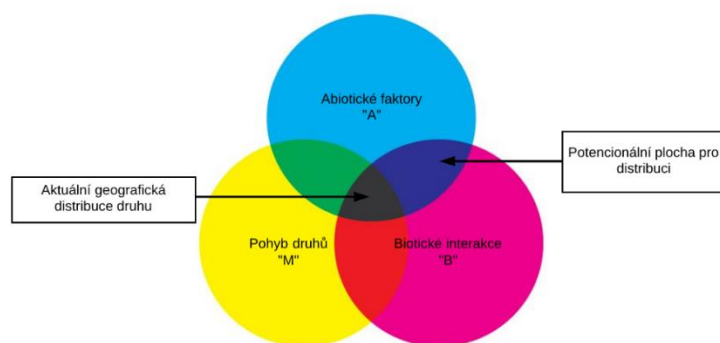
Hlavní limitací SDMs je špatná kvalita výskytových dat, jejich dostupnost a kompletnost. Testování chyb při použití reálných výskytových dat je obtížné. Řešením je použití virtuálního druhu. Virtuální druhy umožňují studovat působení nejrůznějších aspektů na SDMs (Huang et al., 2016). Modelaci virtuálního druhu pro biogeografické analýzy lze provést třemi metodami: 1. Pravděpodobnostní metodou založenou na náhodném výběru spojenou s pravděpodobností výskytu, 2. Prahovou metodou založenou na tzv. prahu, 3. NicheLim metodou odvíjející se od prahového zobrazení, ale lišící se v tvorbě environmentální vhodnosti prostředí, kdy je na každou environmentální proměnnou aplikován tzv. práh a až následně jsou environmentální vrstvy spojeny. Všechny metody kombinují vhodnost prostředí skrze environmentální proměnné.

I přes častou aplikaci prahové metody (například Hirzel et al., 2001, Bombi & D'Amen, 2012; Valera et al., 2014) dokazují nedávné studie výhodnější použití pravděpodobnostní metody (Meynard & Kaplan 2012; Moudrý 2015). Prahová metoda představuje extrémní případ, kdy je druh bez výjimky reprezentován pouze na jedné straně environmentálního gradientu (Meynard & Kaplan, 2012). NicheLim metoda představuje nový algoritmus pro generování virtuálního druhu (Huang et al., 2016).

Hlavním cílem práce je porovnat, jak může zvolená metoda generování virtuálního druhu ovlivnit predikční schopnost modelů. Pro možnost porovnání metod je nezbytné: 1) vytvoření virtuálního druhu pravděpodobnostní metodou, prahovou metodou, NicheLim metodou, 2) vytvoření modelů druhové distribuce pomocí algoritmu Maxent 3) porovnání vytvořených virtuálních druhů s modely druhové distribuce pomocí AUC a Schoener's D indexem.

2. LITERÁRNÍ REŠERŠE

Představa rovnováhy a stability v biogeografii je pomíjivá. Biotické interakce, abiotické faktory a pohyby druhů představují hlavní faktory ovlivňující potencionální distribuci druhů. Sobéron & Peterson (2005) zavedl nový pojem „BAM“, kterým se snaží názorně ilustrovat hlavní faktory ovlivňující druhové distribuční modely. Diagram „BAM“ (viz Obr.1) popisuje biotické (biotic=B), abiotické (abiotic= A) a hybné (movement= M) složky faktorů ovlivňující distribuci druhů. Tento diagram složí jako základní element k pochopení distribuce druhů.



Obr. 1 Diagram „BAM“ znázorňující aktuální druhovou distribuce a potencionální plochu pro distribuci druhu (= vhodné abiotické a biotické faktory)

(Zdroj: Sobéron & Peterson, 2005; autor)

Saupe et al. (2012) potvrzuje důležitost začleňování „BAM“ faktorů do SDMs a doporučuje přikládat význam těmto faktorům před samotnou tvorbou modelů. Začleňování biotických, abiotických a hybných faktorů a porozumění jejich příslušných rolí ve vztazích k predikci potencionální distribuce je pak důležité při SDMs invazivních druhů (Miller, 2014).

Navzdory důležitosti všech faktorů, je hlavní zkoumání zaměřené převážně na biotické faktory. A to především proto, že biotické faktory jsou nejdynamičtější a velice složité pro jejich zakomponování do modelování druhového rozšíření.

2.1 Faktor M

Pohyb druhů (Movement) představuje klíčovou roli při určování rozptylu druhů v geografické prostoru. Konkrétněji je faktor M určen kombinací schopností denního rozptylu druhu (kapacita pohybu druhu) a historickým vývojem přesunů, které otevíraly na jednu stranu možnost migrace do vzdálenějších lokalit, než byla obvyklá denní distribuce druhu, na druhou stranu byla migrace často znemožněna bariérami (Soberón & Peterson 2005). Mezi limitující bariéry řadíme pobřeží, koryta velkých řek, hory, propadliny a jiné topografické prvky. Tyto prvky zůstávají dlouhodobě fixovány na jednom místě a zapřičiňují izolovanost populací v mnoha částech světa. Základní snahou ekologů je nalezení areálu distribuce druhu a environmentálního vhodného prostředí pro výskyt druhů.

2.2 Faktor B

Biotické interakce jsou vztahy a vazby mezi organismy v ekosystému. Lze je rozdělit na vnitrodruhové (interakce jsou uskutečněny v rámci jednoho druhu) a na interakce mezidruhové (interakce se odehrávají mezi populacemi různých druhů). Většina SDMs pracuje s předpokladem, že biotické interakce nemají vliv na výskyt druhů (Huntley et al. 1995, Bakkenes et al. 2002), případně s domněním malého rozsahu ovlivnění (Pearson & Dawson 2003, Dormann et al. 2007, Heikkinen et al. 2007). Tato tvrzení však vyvrátil Araújo et al. (2014) podklady, které demonstrují, jak důležité je zahrnutí biotických interakcí do SDMs (Araújo & Luoto 2007).

Mezi základní biotické interakce řadíme: 1. Komensalismus, při kterém jedna populace využívá jinou populaci bez jejího poškození, 2. Mutualismus, který představuje kladné využití obou populací, například mykorhiza, 3. Predace, při které dochází ke spotřebování jednoho organismu (kořisti) jiným organismem (predátorem) za účelem získu energie (Begon et al., 1997).

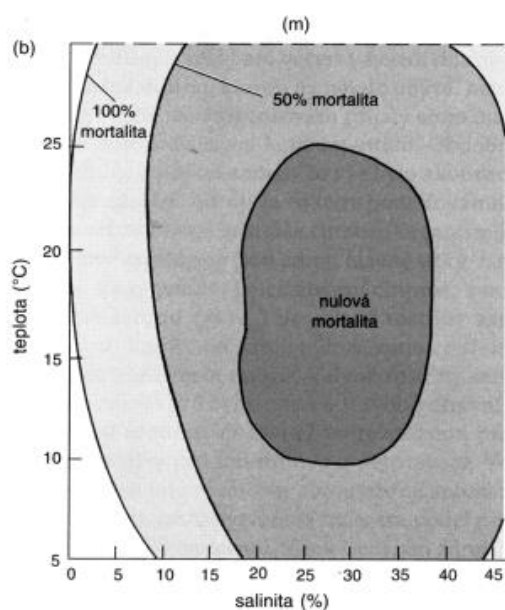
2.3 Faktor A

Abiotické podmínky představují vliv neživé přírody, která má určitou fyzikální a chemickou povahu. Abiotické podmínky zahrnují: aspekty klimatu, vztahy prostředí, půdní podmínky, jako příklad lze uvést světlo (dobu působení, intenzitu a směr dopadu světla), vodu, minerální látky, radiace, hluk, vibrace (Hutchinson, 1957).

Vhodné abiotické podmínky umožňují přetrvání druhu v areálu jeho rozšíření (Sobéron & Peterson, 2005). Pulliam (2000) zdůrazňuje důležitost abiotických podmínek pro výskyt druhu, a proto vyvinul simulaci založenou na kvantitativním popisu abiotických podmínek.

2.2 Ekologická nika

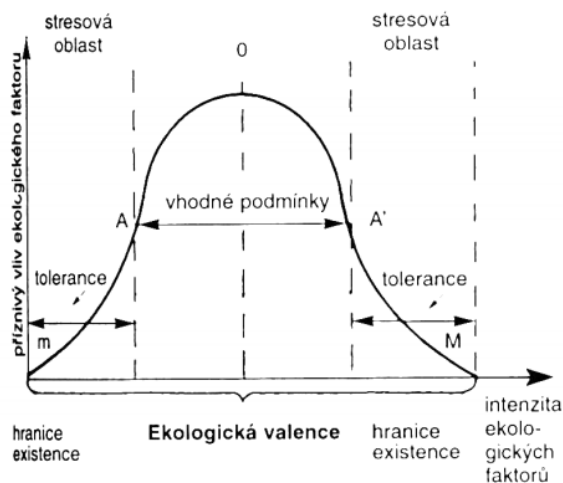
Ekologická nika (viz Obr. 2) vyjadřuje vztah mezi organismem a prostředím. Často je ekologická nika špatně interpretována a bývá označována jako místo výskytu druhu (stanoviště, habitat). Jiný pohled na ekologickou niku přinesl roku 1957 britsko – americký zoolog George E. Hutchinson, který ji definoval jako termín označující environmentální dimenzi, bez které organismy nedokáží přežít a nejsou schopné reprodukce. Hutchinson později definoval ekologickou niku jako reakci druhu na environmentální podmínky prostředí při absenci biotických interakcí. Jeho snahy vedly k vymezení potřeb organismů, lokaci vhodného místa k přežití a zdroje obživy (Heath & Smith, 1989). Jedním z hlavních faktorů ovlivňující vývin organismů je teplota. Pro každý druh existuje jiná optimální teplota, při které se dokáže vyvíjet a rozmnožovat. Mezi další faktory ovlivňující výskyt druhů řadíme: pH, vlhkost, srážky, rychlost větru. Ekologickou niku lze znázornit jako n-rozměrný prostor, kde veličina n představuje počet vstupující faktorů environmentálního prostředí (Heath & Smith, 1989; Austin, 2002; Chase & Leibold, 2003).



Obr. 2 Praktický příklad dvourozměrné ekologické niky, v závislosti na teplotě a salinitě vody je zde představena mortalita samic garnáta *Crangon septemspinosa*.

(Zdroj: Haefner, 1970)

Důležitá je adaptace druhů tj., přizpůsobení se druhu konkrétním podmínkám prostředí (viz Obr. 3). Druhy lze rozdělit na stenoekní, které jsou úzce vázané na vymezené podmínky existence (mají úzkou ekologickou valenci) a na euryekní druhy, které snášejí podmínky prostředí ve velkém rozsahu (mají širokou ekologickou valenci) (Peterson & Vieglais, 2001).



Obr. 3 Ukázka ekologické valence (vyjadřuje schopnost druhů snášet konkrétní faktory prostředí, například: teplota, srážky). Vrchol představuje optimální působení daného faktoru.

(Zdroj: Václavík, 2013)

2.3 Druhové distribuční modely

Založení SDMs se datuje do 70 let 19. století. V posledním desetiletí došlo k enormnímu nárůstu publikací zaměřujících se na metodiku a využití SDMs. Rozvoj geoinformačních systémů (GIS) umožnil modelovat s využitím přesnějších dat. Nové poznatky z výzkumů jsou používány v oblasti ochrany přírody, ekologie, a to především na lokální úrovni (Zimmermann et al., 2010).

SDMs představují statistické metody, které slouží k predikci rozšíření druhu v geografickém prostoru. Hlavní principem biologicko-geografického výzkumu je porozumění, jak a proč se organismy šíří v prostoru. SDMs kvantifikují korelaci mezi environmentálními faktory a distribucí flóry a fauny (Elith & Leathwick, 2009; Franklin, 1995, 2010; Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000; Miller & Rogan, 2007). Výsledkem SDMs jsou prostorové predikce znázorňující rozšíření druhů nebo určující vhodnost habitatu.

Modely mohou být použity například k ochraně ohrožených druhů (Araújo et al., 2002), testování biogeografických hypotéz (Leathwick, 1998), zkoumání evolučních procesů (Kozak et al. 2008), identifikaci vhodných míst ke znovu vysazení původních druhů a k nalezení výskytu vzácných druhů (Edwards et al. 2005). Dále je také možné, je využít k mapování změn druhové distribuci v důsledku globálního oteplování či antropogenní činnosti, která je lokálního nebo globálního měřítka a k predikci změn druhové distribuce v určitém časovém horizontu na určitém geografickém území, pokud víme, jak se budou environmentální faktory měnit. (DeAngelis & Mooij 2005, Araújo & Rahbek 2006, Thuiller et al., 2008).

Ekologové věnovali veliké úsilí za účelem: vytvoření nových statistických metod (Phillips et al., 2006), odstranění autokorelace (Dormann & Carsten, 2007), porovnání modelovacích algoritmů (Elith et al. 2006) a validaci modelů (Segurado et al., 2004). Araújo & Guisan (2006) uvádí problémy, na které by v budoucnu nemělo být při výzkumech opomíjeno: definování ekologické niky, zlepšení strategie při volbě vhodného modelovacího algoritmu, lepší validační metody, vhodný výběr environmentálních proměnných a jejich měřítka.

2.4 Vstupní data

Používání přesných dat je nezbytné pro vytvoření spolehlivé predikce. Správná volba typu dat, jejich dostupnost a volba měřítka patří mezi největší limitace v aplikaci SDMs. V roce 1994 kategorizoval Goodchild vstupní data na pole a objekty. Při pohledu na „pole“, geografické proměnné mohou být kategorického nebo kontinuálního charakteru, vždy mohou být změřeny (obsahují hodnoty). Příkladem je nadmořská výška (kontinentální proměnná) a typ vegetace (kategorická proměnná), které mohou být definovány v každém bodě po celém světě. Objekty jsou vzájemně nesouvisející geografické proměnné rozptýlené v prostoru. Příkladem je mapa zaznamenávající presenci druhu. Při modelaci SDM dochází k propojení výskytových dat (objektů) s environmentálními daty (polem) za účelem vytvoření nové mapy (např. distribuční). Při modelaci je nutné brát v potaz charakter vstupních dat, výběr modelovacího algoritmu, co má predikce reprezentovat a jakého výsledku chceme modelaci dosáhnout (Goodchild, 1994).

2.4.1 Environmentální

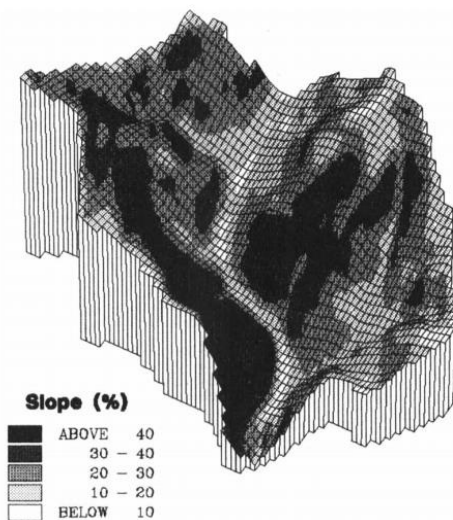
Environmentální data neboli prediktory představují faktory prostředí v zájmové oblasti, které ovlivňují distribuci organismů. Důležitá je při úpravě environmentálních dat volba správného měřítka. Není stanovené žádné univerzální měřítko, které by bylo vhodné pro všechny SDMs.

Kuemmerlen et al. (2014) uvádí základní kategorie environmentálních dat: topografické, bioklimatické, hydrologické, způsob využití půdy. Klimatická a topografická data jsou velmi často využívána při modelaci SDM, avšak kvalita klimatických dat nedosahuje takové kvality.

Vektorová a rastrová data

Vektorová data se používají pro popis jednotlivých geografických objektů: ve formátu bodů lze označit například lokaci památných stromů, ve formátu linií lze vyznačit řeky a potoky, ve formátu polygonů lze mapovat vodní nádrže a jezera. Rastrová data se používají pro reprezentaci spojitě se měnících jevů (teplota, digitální model terénu), základní jednotku tvoří buňka. Buňky jsou uspořádány do tzv. mozaiky a jejich velikost určuje pixel. Goodchild (2011) upozorňuje na fakt, že většina dostupných vektorových dat má nízkou kvalitu rozlišení, a proto doporučuje pracovat s rastrovými daty při vědeckých studiích.

Topografická data zahrnují nadmořskou výšku, akumulaci vody, svažitost, expozici svahů. Často používaným typem topografických dat je Digitální model terénu = DEM (viz Obr. 4), reprezentuje reliéf terénu, obsahuje informace o nadmořské výšce, nikoliv povrchu (v modelech DEM jsou odfiltrovány prvky, které nesouvisejí s georeliéfem: budovy, vegetace). Data se získávají z laserového skenování, fotogrammetrie, geodetického měření (Moore et al., 1991).



Obr. 4 Digitální model terénu představující svažitost terénu uváděnou v procentech, velikost pixelu je zde 30 metrů, celková rozloha projekce je 210 ha, zájmová plocha se nachází v Montaně. (Zdroj: Moore et al., 1991)

Bioklimatická data zahrnují roční průměrné teploty, sezónní průměrné teploty, srážky, rychlosti větru, maximální teplotu v nejteplejším měsíci, minimální teplotu v nejchladnějším měsíci. Bioklimatická data vznikají interpolací dat z meteorologických stanic. Teplota, světelný svit a vlhkost jsou základní faktory, které svým působením ovlivňují vývin druhů. V teoretických výzkumech nebylo specifikováno, jaké parametry stojí za primárními změnami při růstu a reprodukci organismů. Bioklimatická data jsou velmi důležitá při determinaci SDM, jejich použitelnost je často limitována nízkou kvalitou. Nejčastěji používaná bioklimatická databáze je WorldClim. Data jsou zde dostupná v několika rozlišeních (~ 1 km² až ~ 340 km²) a obsahem této databáze je 19 klimatických vrstev zobrazujících roční hodnoty teplot a srážek, dále sezónní a extrémní faktory (Miller, 2010).

Hydrologická data zahrnují průměrné roční odtoky, přítoky.

Způsob využití půdy znázorňuje ekologickou hodnotu, míru zastavění a devastaci půdy. Mezi základní typy způsobu využití půdy patří: orná půda, trvalý travní porost, zastavěná plocha, les, zahrady a sady (Kuemmerlen et al., 2014).

Tab. 1 Přehled environmentálních dat používaných při SDMs.

Organizace	Data
WorldClim	Světová bioklimatická databáze zahrnující 19 klimatických vrstev (rozlišení ~ 1 km ² až ~ 340 km ²)
PRISM	USA klimatická databáze zahrnující měsíční, roční a historické vrstvy (rozlišení 800 m až 4 km)
Global elevation and related data	Světová databáze zahrnující data o nadmořské výšce (rozlišení od 1 km)
UNEP	Globální databáze zahrnující data o typech půdy
NRCS	USA databáze zahrnující data o zmapované půdě

(Zdroj: Miller, 2010)

2.4.2 Výskytová

Data popisují distribuci druhu, používaná jsou dvojího typu: data presence-only a presence-absence. Data presence-only poskytují informace o přítomnosti druhu, data presence-absence ukazují jak prezenční data, tak i lokality na kterých se druh nevyskytuje – absenční data. Tento druh dat je hůře dostupný zvláště pro migrující druhy, sběr tedy vyžaduje vynaložení většího úsilí k zajištění spolehlivosti dat v porovnání s presence-only daty (Miller 2010, Pearson et al., 2007). Proto jsou data presence-only častěji používaná při predikci druhů, lze je snadno získat čerpáním z již existujících dat, které jsou dostupné v muzeích, databázích a přírodovědných sbírkách. Problém těchto dat spočívá v jejich nekompletnosti. Sběr mohl probíhat jen na určitých lokalitách při jednotlivých výzkumech, a nikoliv v rámci systematického mapování. (Phillips et al. 2004, Elith & Leathwick 2009).

Tab. 2 Zdroje výskytových dat.

Organizace	Data
VertNet	Globální muzejní databáze historie obratlovců, zahrnující 4 sítě databází – MaNIS, HerpNET, ORNIS a FishNet
North American Breeding Bird Survey	Databáze čítající informace o více jak 400 druzích ptáků, sběr dat probíhal od roku 1966
NatureServe	Databáze v Severní a Jižní Americe zahrnují informace o rostlinách, zvířatech a ekosystému
National Invasive Species Information Center	Americká a regionální databáze zahrnují informace o invazivních druzích
Global Biodiversity Information Facility	Informace o rostlinách, zvířatech, houbách a mikroorganismech

(Zdroj: Miller, 2010)

2.5 Modelovací algoritmy

Volba vhodného algoritmu je velice důležitá, protože každá modelovací metoda přináší jiné výsledky. Objektivně nelze určit, jaký algoritmus je nejvhodnější a nejpřesnější. Klíčové kroky (viz Obr. 5) pro kvalitní model zahrnují: shromáždění relevantních dat, vyhodnocení komplexnosti a přesnosti výskytových dat, vhodná selekce modelovacího algoritmu (závisí na tom, zda je predikována pravděpodobnost výskytu druhu tzv. probability nebo spíše vhodnost prostředí pro druh tzv. suitability) a validace modelu (Miller, 2010).

Presence-only metody

Pokud jsou k dispozici presence-only data je možné využít metody: Gower Metric, ENFA, Maxent (Carpenter et al., 1993).

Algoritmus ENFA (Analýza faktorů ekologické niky) byl vyvinut k predikci potencionálního rozšíření pro použití presence-only dat. Toto zobrazení, popsal Hutchinson jako ekologickou teorii niky, vytvořenou z vhodné mapy habitatu, která nepřímou ukazuje potencionální distribuci druhu (Hirzel et al, 2002). Původně byla tato metoda vytvořena k predikci výskytu fauny. Zvířecí monitoring byl často

náchylný k nesprávným záznamům výskytu, kvůli časté migraci zvěrstva a skrývání se během monitoringu zájmového pole (Hirzel et al., 2002).

Maxent (Model maximální entropie)

Maxent je automatický algoritmus využívající prezenční data a tzv. background data (environmentální data o náhodném vzorku) pro určení pravděpodobné distribuce. Maxent odhaduje distribuci na základě maximální entropie (Stockwell, 1999). Distribuce druhu je reprezentována jako pravděpodobnost distribuce π skrze soubor X z míst provedeného výzkumu. Z toho vyplývá, že π značí nezáporné hodnoty každého místa X , následně jsou tyto hodnoty sečteny do jedné $\pi(x)$ (Phillips & Dudík, 2008). Prvky v Maxentu jsou odvozeny z environmentálních proměnných dvojího typu: kontinuální a kategorické. Kontinuální proměnné značí libovolné reálné prvky, které korespondují s kvantitativním měřením jako je nadmořská výška, roční srážky, maximální teploty. Kategorické proměnné představují limitovaný počet vzájemně nesouvisících prvků (půdní a vegetační typ).

Maxent je považován za velice schopný nástroj při práci s řídkými a nepravidelně rozloženými daty nebo při menší zájmové ploše. Funguje za předpokladu systematického sběru dat z celé zájmové plochy. Je využíván pro potencionální distribuci vzácných a ohrožených druhů, předpověď budoucí distribuce druhů při působení environmentálních proměnných a při rozdělení ekologické niky (Phillips & Dudík, 2008).

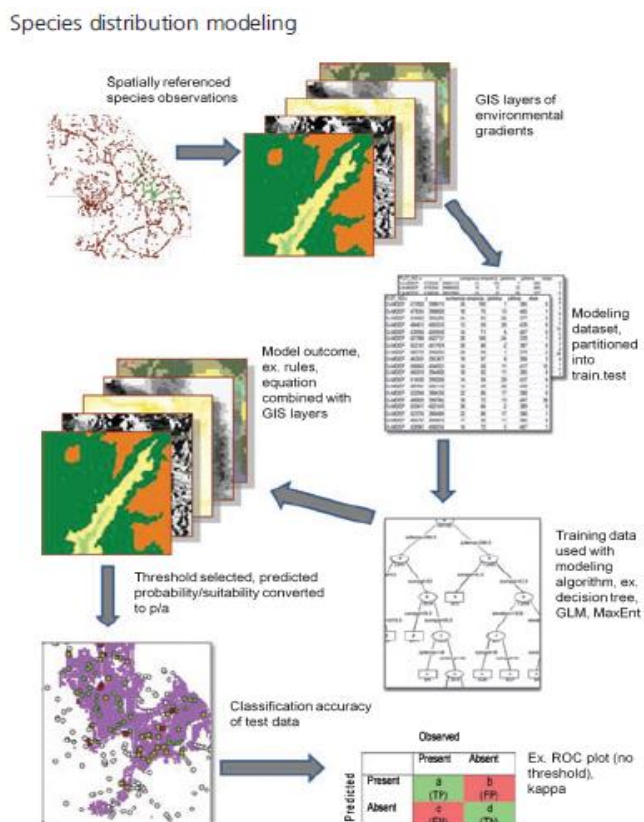
Presence-absence metody

Pro data presence-absence lze využít metody: GA (genetický algoritmus) (Stockwell et al., 2006), ANN (umělé neuronové sítě) (Pearson & Dawson, 2003), MARS (model mnohorozměrné adaptivní regrese), GLM (generalizovaný aditivní model) a jeho neparametrickou nadstavbu GAM=generalizovaný aditivní model (Araújo & Luoto, 2007).

GAM a MARS jsou přizpůsobivé algoritmy, vhodné pro identifikaci typu druhové reakce. Algoritmus MARS je v porovnání s GAM rychlejší a výsledky lépe převádí do predikční mapy, ale jeho použití je nevhodné pro presence-absence data (Franklin & Miller 2009; Leathwick et al. 2005). GAM produkuje konzistentní

výsledky bez ohledu na nastavení odezvové funkce (Santika & Hutchinson, 2009) a při komparaci s GLM přináší kvalitnější a přesnější výsledky (Araújo et al. 2005; Meynard & Quinn 2007). Modely GLM našli hojně využití pro SDMs díky své tolerantnosti při práci s různými proměnnými. Guisan et al. (2002) a Elith et al. (2006) označuje GLM za jeden z nejvhodnějších algoritmů pro SDMs. Přesnost GLM ovlivňuje negativně extrémně vysoká či nízká druhová prevalence, při použití prahové metody namísto pravděpodobnostního zobrazení, přináší příliš optimistické výsledky při zkoumání schopnosti rozlišování mezi presencemi a absencemi (Meynard & Kaplan, 2013). Nevhodné pro validaci GLM modelu je použití AUC (Jiménez-Valverde, 2012).

ANN představuje regresní algoritmus, jehož výsledkem je numerický výstup. Používá se pro klasifikaci druhů (Franklin & Miller, 2009). Výhodou je schopnost algoritmu zpracovávat rozlišná vstupní data (Pearson & Dawson, 2003).



Obr. 5 Proces SDM – vstupující výskytová data a environmentální proměnné, selekce modelovacího algoritmu, mapa vhodnosti prostředí/pravděpodobnosti výskytu transformována do presence-absence mapy, validace modelu.

(Zdroj: Miller, 2010)

2.6 Validace modelů

Validace představuje zpětnou kontrolu modelu, ověřuje správnost volby modelovací metody a její přesnost. Výběr validační metody by měl probíhat s ohledem na typ vstupních dat (Moudrý, 2015). Při validaci modelu dochází k rozdělení dat na část kalibrační a validační. Tento aktse nazývá Data – splitting. Kalibrační část je určena pro zkoušku modelu, je zastoupena nejčastěji 75 % původních dat. Zbývající data jsou použita pro validační část, která slouží k vyhodnocení výkonu modelu. Rozdělení dat je uskutečněno náhodným principem a žádná data nejsou duplicitní. Nevýhodou rozdělení je možná ztráta informací o datech. Kalibrační a validační data by měla být vzájemně nezávislá a pocházet např.: z jiného časového období (Fleishman et al., 2002), z odlišného areálu výskytu (Peterson, 2003), z odlišných vstupních proměnných – jiné měřítko (Areújo et al., 2005).

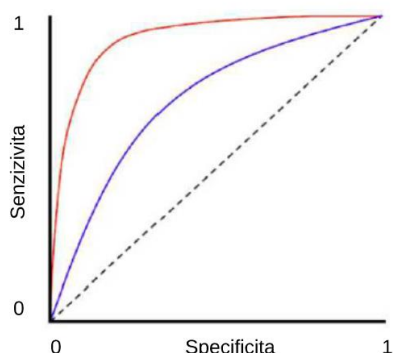
Validace modelů se dělí podle přístupů, kterými lze hodnotit modely:

- a) threshold-independent indexy (hodnotí distribuce vytvořené pomocí spojitých dat): AUC (Domínguez-Domínguez et al., 2006; Pearson et al., 2007; Moudrý & Šimová, 2012, 2013).
- b) threshold-dependent indexy (hodnotí binárních predikcí): senzitivita, specificita (Moudrý & Šimová, 2012 b).

ROC křivka (Operační charakteristická křivka) měří predikční schopnost modelu, na osách křivky je vynesena senzitivita a specificita. Senzitivita (Se) vyjadřuje podíl správně předpověděných prezencí, specificita (Sp) představuje podíl správně předpověděných absencí (viz Obr. 6). Z této křivky je odvozena AUC hodnota (Area Under the ROC curve), která měří schopnost modelu rozlišovat mezi presencemi a absencemi a určuje, zda byl každý záznam vybrán náhodně. Metoda byla vyvinuta za účelem odhalení radarových signálů a adaptována pro lékařské výzkumy (Pepe, 2000). AUC je široce využívána statistická metoda k vyhodnocení diskriminační kapacity SDMs (Jiménez-Valverde, 2012).

Nabývá hodnot od 0-1, kdy 1 značí 100 % spolehlivé odlišení míst, kde je druh přítomen od lokalit, kde se nenachází. Pokud je hodnota AUC nižší než 0.5, model nelze brát v potaz (Krzanowski & Hand, 2009). Výhodou je hodnocení modelu

z pohledu prezencí a absencí současně (senzitivita a specificita ukazuje pouze, jak si model vedl v jedné kategorii). Nelze spoléhat na validace jen skrze AUC, doporučuje se používat i jiné kalibrační metody (Pearson et al., 2007)



Obr. 6 Příklad ROC křivky.

(Zdroj: Pearson et al., 2007)

Tab. 3 Klasifikace prediktivní přesnosti modelu pomocí AUC

Přesnost modelu	Hodnota AUC
výborná	0.9- 1
dobrá	0.8 – 0.9
průměrná	0.7 – 0.8
podprůměrná	0.6 – 0.7
nedostačující	0.5 – 0.6

(Zdroj: Kienast et al., 2012)

Niche overlap (překryv niky)

Rozmanitost v přístupu a měření byl použit k měření niche overlap (Schoener, 1970). Metodu lze aplikovat pro porovnání taxonomických, geografických a časových skupin, přesněji lze porovnávat jeden druh při odlišném nastavení (přirozeně či invazivně se vyskytující), případně pro stejné druhy v rozličném časovém horizontu (například před a po klimatické změně). Dá se aplikovat pouze na metodické studie založené na použití virtuálního druhu, protože při práci s reálnými daty by nebylo možné výsledný model s čím porovnat.

Postup vyžaduje dodržení 3 kroků:

- 1) výpočet hustoty výskytů a určení environmentálních proměnných, výsledkem je multivariabilní analýza
- 2) rozměry překryvu niky vycházející z multivariabilní analýzy
- 3) statistický test (Warren et al., 2008).

Schoener's D index, sloužící pro určení překryvu niky, nabývá hodnot od 0 (žádný překryv) do 1 (identická potencionální distribuce). Rödder & Engler (2011) doporučuje používat následující klasifikaci pro snadnější interpretaci výsledků.

Tab. 4 Klasifikace přesnosti modelu při použití Schoener's D indexu

Přesnost modelu	Hodnota Schoener's D indexu
velmi vysoký překryv	0.8–1.0
vysoký překryv	0.6–0.8
nevýrazný překryv	0.4–0.6
nízký překryv	0.2–0.4
žádný, případně velmi limitující překryv	0.0–0.2

(Zdroj: Rödder & Engler, 2011)

Chybová matice tvořená presencemi a absencemi (viz Obr. 7): a= správně určené presence, b= špatně určené presence, c= špatně určené absence, d= správně určené absence ($N = a + b + c + d$). Slouží k vyhodnocování schopnosti presence-absence modelů, při SDMs dochází ke dvěma nejčastějším chybám: b a c (Fielding & Bell, 1997).

		Realizovaná distribuce	
		Presence	Absence
Predikční distribuce	Presence	a	b
	Absence	c	d

Obr. 7 Chybová matice složená z presence a absence dat

(Zdroj: Fielding & Bell, 1997)

Tab. 5 Přehled základních validačních metod.

Validační metoda		Vzorec
PCC	Procento správně určených presencí a absencí	$(a + d) / N$
KAPPA	Procentně vyjádřená shoda mezi přesností predikční a realizované distribuce	$[(a+d) - (((a+c) (a+b) + (b+d) (c+d)) / N)] / [N - (((a+c) (a+b) + (b+d) (c+d)) / N)]$
Sensitivita (Se)	Poměr správně určené realizované presence	$a/(a+c)$
Specificita (Sp)	Poměr správně určené realizované absence	$d/(b+d)$
PPP	Procentně znázorněné skutečně predikované presence	$a/(a+b)$
NPP	Procentně znázorněné skutečně predikované absence	$d/(c+d)$
TSS	Statistika skutečných dovedností	$Se + Sp - 1$

(Zdroj: Miller,2010)

2.7 Virtuální druh

Hirzel et al (2001) poprvé použil pojem „virtuální druh“. Testování SDMs při použití reálných dat je omezené, jednak kvůli jejich neúplnosti a nepřesnosti a také kvůli vznikajícím chybám modelu: sampling bias, positional error, spatial scale. Momentálně je kvalita výskytových dat označována za hlavní problém limitující SDM (Duputié et al., 2014). Testování chyb výskytových dat při použití reálných dat je komplikované. Možné řešení představuje použití virtuálního druhu, který umožňuje zkoumat vlivy rozličných kvalit dat na schopnosti SDMs (Moudrý, 2015).

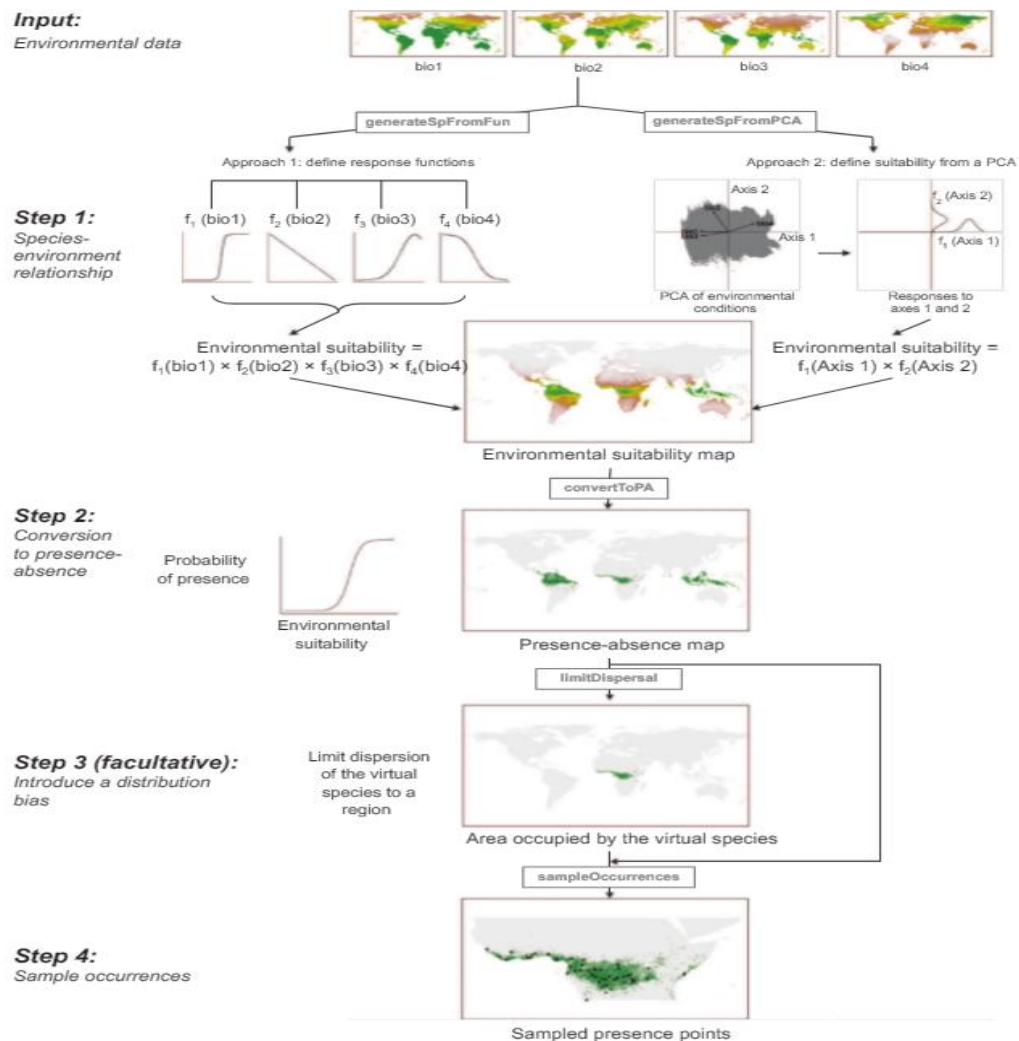
Virtuální druhy jsou zjednodušené modely reálných druhů, které představují jedinou možnost, jak studovat nejrůznější aspekty působící na SDMs. Modelace virtuálního druhu představuje environmentální niku ideálního druhu (Valera & Gomez-Rodriguez, 2015). Virtuální druhy jsou využívány pro tvorbu metodických studií, ve kterých je zkoumán vliv různých metod a odlišná kvalita dat na SDMs. Dále jsou používány ke kvantifikaci reakcí druhů na klimatické změny, k predikci potencionálního přesunu druhů v geografické sféře a k testování metod, které jsou využívány k predikci druhů v geografickém prostoru. Momentálně existují tři metody používané ke konstrukci virtuálního druhu: pravděpodobnostní, prahová, NicheLim. Účelem je zamezení omylů při modelaci reálného výskytu druhu (Austin et al., 2006; Hirzel et al., 2001; Zurell et al., 2010; Naimi et al., 2011; Miller, 2014; Valela et al., 2014; Fourcade et al., 2014). Václavík & Meentemeyer (2012) použil virtuální druh k testování druhové distribuce invazivních druhů v různých fázích jejich invaze (krátkodobý výskyt, adaptace invazivního druhu na místní podmínky– mutace, období růstu a intenzivního šíření, dovršení obsazenosti přírodních stanovišť, eliminace původního druhu) a doporučuje brát zřetel na určení fáze invaze při výběru algoritmu a validace modelu.

V posledních letech byl virtuální druh používán k vylepšení predikčních metod. Slouží k lepší kontrole nad kvalitou prezenčních a / nebo absenčních dat, pomáhají k lepšímu vyhodnocování správnosti predikce a k lepšímu porozumění ekologických problémů se kterými se vědci setkávají při studiích reálných druhů (Naimi et al., 2011; Jiménez-Valverde et al., 2009; Barbet-Massin et al., 2012; Santika & Hutchinson, 2009). Pokud je virtuální zobrazení druhu úspěšné, dalším krokem, by měla být aplikace stejného modelovacího algoritmu na reálná data.

Virtuální druh je vygenerován na základě environmentálních vrstev, Miller (2014) nedoporučuje používání virtuálních environmentálních vrstev z důvodu až moc zjednodušené virtuální distribuci. Jednotlivým environmentálním vrstvám může být přiřazena konkrétní váha, která upravuje vztahy při generování virtuální distribuce.

Většina virtuálních druhů je tvořena na základě HSI (index vhodnosti habitatu), (Austin et al., 2006; Hirzel et al., 2001; Zurell et al., 2010; Naimi et al., 2011; Miller, 2014; Valera et al., 2014). HSI je generován ve třech krocích: 1) selekce

environmentálních proměnných, 2) selekce vhodné metody pro tvorbu HSI, metody pro generování HSI jsou založeny na základních funkcích, např.: Gaussova křivka, lineární křivka (rostoucí nebo klesající průběh), 3) kombinace jednotlivých částí HSI vedoucí k vytvoření multi-variabilního habitatu (viz Obr. 8). Výsledky modelů by měly analyzovány i s ohledem na typ a kvalitu vstupních dat.



Obr. 8 Tvorba virtuálního druhu: 1. vstupní environmentální proměnné, 2. simulace reakcí virtuálního druhu na environmentální gradienty lze vytvořit a) nadefinováním odezvové funkce – Gaussova, lineární, kvadratická křivka b) generováním PCA (Analýza hlavních komponent)-analýza, která je založena na dekoleraci dat, snižuje rozměrná data a zároveň se snaží zachovávat co nejvíce informací o datech, 3. vytvořená environmentální vhodného prostředí, 4. transformace do presence-absence mapy, 5. volba vzorku dat s možností nastavení limitací, v jaké části území se druh bude vyskytovat.

(Zdroj: Leroy et al., 2015)

2.7.2 Software

Virtuální druh lze vygenerovat pomocí VIRTUALSPECIES a SDMVSPECIES package (balíčku). Duan et al (2014) popsal použití SDMVSPECIES package pro prahovou metodu zobrazení. Současná verze SDMVSPECIES (ver. 0.2.1) obsahuje 4 metody k vytvoření virtuálního druhu: Niche syntheses method, Pick mean method, Pick median method, Artificial bell-shaped response method (Jiménez-Valverde & Lobo, 2007). VIRTUALSPECIES package byl sestaven pro pravděpodobnostní metodu zobrazení (Leroy et al., 2015).

2.7.3 Metody vytváření virtuální reality

Pravděpodobnostní metoda

Pravděpodobnostní metoda představuje náhodný proces spojený s pravděpodobností výskytu (viz Obr. 9). Výhodou této metody je pozvolné reagování na environmentální proměnné a možnost opakované simulace se stejnými či diferenčními podmínkami, což přináší lepší vyhodnocování výsledků a chyb modelů. Ke kontrole druhové reakce na environmentální proměnné používá Meynard & Kaplan (2012) logistickou funkci, která je vhodná při použití pravděpodobnostní metody.

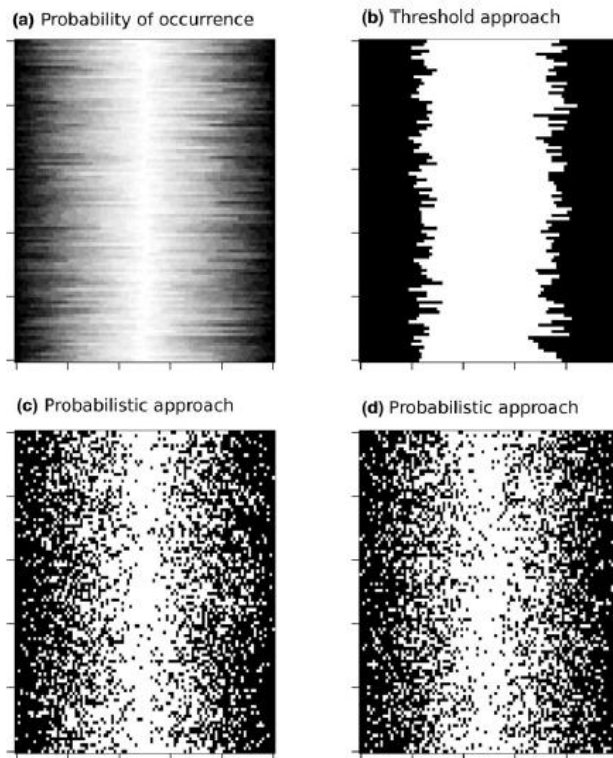
Prahová metoda

Bombi & D' Amen (2012) použil k vytvoření virtuálního druhu uměle stanovenou specifickou hodnotu (tzv. práh), druh je poté zobrazen, pokud se vyskytuje nad stanovenou (nikdy ne pod) hranicí (viz Obr. 9). Mnoho studií donedávna využívalo prahovou metodu zobrazení, která výsledky rázně rozčleňuje bez ohledu na ekologické faktory.

Meynard & Kaplan (2012) uvádí důvody, proč by prahová metoda neměla být používána:

1. Podle ekologické teorie je podporována myšlenka dynamického užívání prostoru, zahrnující proces kolonizace a zániku druhů, použití prahové metody je nevhodné pro tvorbu databází,
2. při modelaci přináší nesprávné a nekompletní výsledky,
3. schopnost modelu rozlišit presence a absence je nízká, nedokáže reagovat na environmentální změny,

4. mnoho SDMs je založeno na použití logistické křivky (generovaný lineární model GLM a generovaný aditivní model GAM), model vytvořeným tímto algoritmem je nevhodný pro použití prahové metody,
5. opakování simulace je omezené.



Obr. 9 Porovnání prahové a pravděpodobnostní metody A) pravděpodobnost výskytu druhu při použití presence-absence dat (černá barva značí vysokou pravděpodobnost výskytu, bílá barva značí nízkou pravděpodobnost výskytu). B) znázorňuje použití prahové metody, C a D znázorňuje pravděpodobnostní metodu s použitím presence-only dat.

(Zdroj: Meynard & Kaplan, 2012)

NicheLim

Slovní spojení NicheLim pochází z výrazu niche limitation method (metoda limitující nikou). NicheLim metoda se odvíjí od prahové metody. Tato metoda je založena na bioklimatické ekologické modelaci niky (Busby, 1991). V prvním kroku modelace je využita prahová metoda pro každou proměnnou, která slouží k transformaci hodnot do binominálních vrstev (popisující četnost výskytu). Vynásobením těchto hodnot získá environmentální vhodnost prostředí. Hlavním diferenciací při použití NicheLim metody je, že k analýzám dochází až po aplikaci prahové metody (Huang et al., 2016).

3. METODIKA

3.1 Charakteristika zájmového území

Výzkum byl proveden na území České republiky, rozloha zájmové plochy činí přibližně 78 866 km². Česká republika je vnitrozemský stát, zaujímající středoevropskou polohu. Reliéfem krajiny jsou převážně pahorkatiny a vrchoviny. Podnebí je mírné a přechodné mezi kontinentálním a oceánským. Průměrná roční teplota se pohybuje v rozmezí 5,5 až 9 °C, nejteplejší oblastí jsou nížiny, nejchladnější jsou horské oblasti. Množství srážek převážně stoupá s nadmořskou výškou.

3.2 Vstupní data

WorldClim

Vstupní environmentální data byla použita z databáze WorldClim, která je souborem globálních klimatických vrstev. Bioklimatické proměnné jsou získávány měřením z meteorologických stanic a následně interpolovány z odvozených měsíčních teplotních a srážkových hodnot v souřadnicovém systému WGS84. Databáze obsahuje roční trendy, například: průměrné roční teploty, roční srážky. Dále sezónní, extrémní a limitující environmentální faktory. Hodnoty jsou nejčastěji uváděné za časové období rok, kvartál (perioda 3 měsíců). Data jsou poskytnutá bezplatně pro studijní účely na webové stránce <http://www.worldclim.org/> (Hijmans et al, 2005).

Tab. 6 Vstupní data reprezentují následující environmentální proměnné.

Environmentální proměnná	Označení	Průměr	Standartní odchylka
1.	BIO5	20 °C	1.5 °C
2.	BIO6	-7 °C	0.5 °C
3.	BIO12	651 mm	135 mm

1. BIO5 = Max Temperature of Warmest Month, zobrazující maximální teploty nejteplejšího měsíce, 2. BIO6 = Min Temperature of Coldest Month, zobrazující minimální teploty nejchladnějšího měsíce, 3. BIO12 = Annual Precipitation, zobrazující roční srážky

(Zdroj: autor)

3.3 Použitý software

ArcGis Desktop

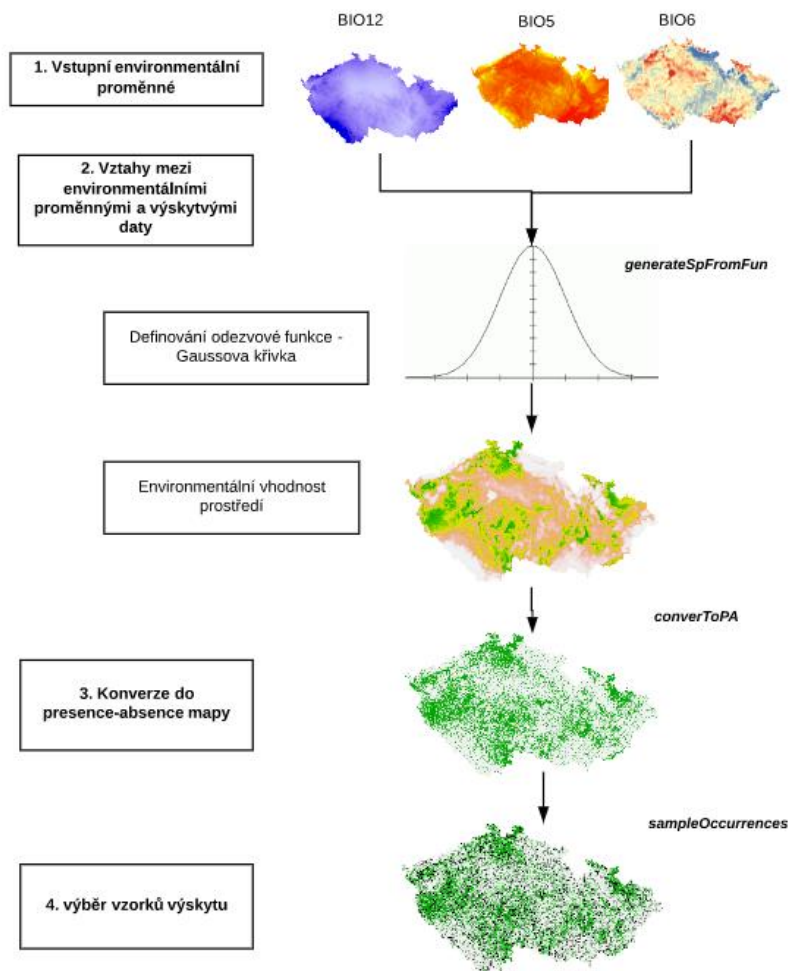
Aplikace používaná pro mapově orientované práce, včetně prostorové analýzy dat. Licence je poskytována v několika úrovních. Skládá se z několika aplikací: ArcMap, ArcCatalog, ArcToolbox.

V první kroku byly nahrány do softwaru ArcGis klimatické vrstvy: BIO5, BIO6, BIO12 v rozlišení 30 s (odpovídá přibližně 1 km²). Souřadnicový systém byl určen projec_WGS84. Klimatické vrstvy byly oříznuty na území České republiky pomocí funkce Extract by Mask.

R

Volně stažitelný programový software R (použitá verze softwaru: R.3.4.3) dostupný na webových stránkách <http://cran.r-project.org/>. Slouží pro statistické výpočty a grafické úlohy. R Studio představuje jednotné prostředí zahrnující konsolové okno sloužící pro vstupní příkazy, dále grafické okno sloužící pro grafické výstupy, editor sloužící pro záznam příkazů, editaci a odesílání do konsolového okna, záznamy příkazů lze uložit ve formě skriptu. Pro vytvoření virtuálního druhu je potřeba několika rozšíření, a to dismo, maps, raster, rdgal, rJava, sp, sqldf, virtualspecies).

3.4 Tvorba virtuálního druhu



Obr. 10 Schéma tvorby virtuálního druhu: vstupní environmentální data – BIO5; BIO6; BIO12, vztahy mezi environmentálními proměnnými a výskytovými daty definovány pomocí Gaussovy křivky, převedení environmentální vhodnosti prostředí do presence-absence mapy a volba vzorků výskytu druhu.

(Zdroj: autor)

3.4.1 Import dat

Vytvořené environmentální proměnné: BIO5, BIO6, BIO12 byli importovány do R Studia pomocí příkazu „*raster*“. Dalším krokem je vytvoření tzv. odezvové funkce – nadefinování Gaussovy křivky, která umožní reakci každé environmentální proměnné (viz Obr. 10)

3.4.2 Odezvová funkce

Gaussova křivka vychází z průměrných hodnot (mean) a standardních odchylek (sd).
-tvar obecné normalizované křivky:

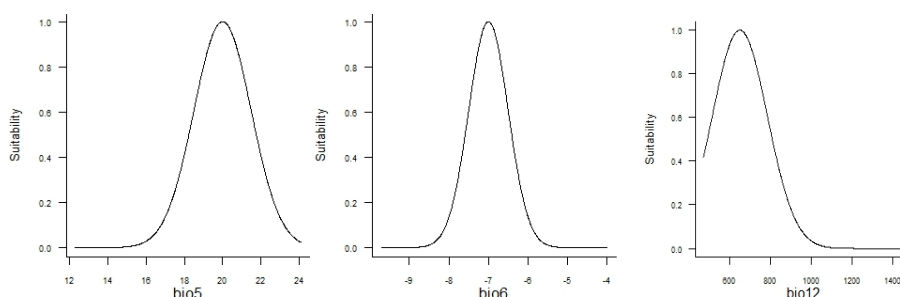
$$f_n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Parametr μ představuje průměrnou hodnotu (mean), parametr σ značí standardní odchylku.

Po každou environmentální proměnnou byl nastaven formát Gaussovy křivky pomocí hodnot mean a sd.

```
valera.Parameters <- formatFunctions(bio5= c(fun = 'dnorm', mean = 20, sd = 1.5) ,  
                                     bio6 = c (fun = 'dnorm' , mean = -7, sd = 0.5),  
                                     bio12 = c (fun = 'dnorm' , mean = 651, sd = 135))
```

Funkcí plotResponse (extenze virtualspecies) lze vykreslit grafy vyjadřující vztahy mezi environmentálními proměnnými a virtuálním druhem. Znáznorňují výstupní hodnoty z generateSpFromFun (viz Obr. 11).



Obr. 11 Vztahy mezi environmentálními proměnnými a virtuálním druhem. Na ose x jsou vyneseny hodnoty ve °C pro BIO5 a BIO6 a v mm pro BIO12, na ose y je vynesena vhodnost prostředí, kde 1 značí nejvhodnější lokalitu a 0 nejméně vhodná místa pro výskyt druhu.

(Zdroj: autor)

Environmentální vhodnost (pravděpodobnost výskytu) druhu je v každém pixelu vypočítána na základě vynásobení environmentálních proměnných – BIO5, BIO6, BIO12 (viz Obr. 10). Hodnoty nabývají od 0 do 1 (0: nejméně vhodné podmínky pro výskyt druhu, 0 z 10 druhů se zde bude vyskytovat, 1: nejvhodnější podmínky pro výskyt druhů, 10 z 10 druhů se zde bude vyskytovat).

3.4.3 Presence-absence mapy

Následným krokem je konverze do presenčně – absenční mapy (viz Obr. 10). Tento krok lze provést třemi způsoby modelace virtuálního druhu: pravděpodobnostní, prahovou a NicheLim metodou. (convertToPA).

Pravděpodobnostní metoda

Modelace je založena na náhodném výběru s pravděpodobností výskytu.

```
gentle.PA <- convertToPA(Valera.species,  
                          PA.method = "probability",  
                          beta = 0.5, alpha = -0.2,  
                          plot = F)
```

Beta vyjadřuje bod změny logistické funkce a alpha určuje sklon logistické funkce. Pravděpodobnostní metoda umožňuje opakování simulace i stejnými podmínkami, proto byla modelace provedena 50 x.

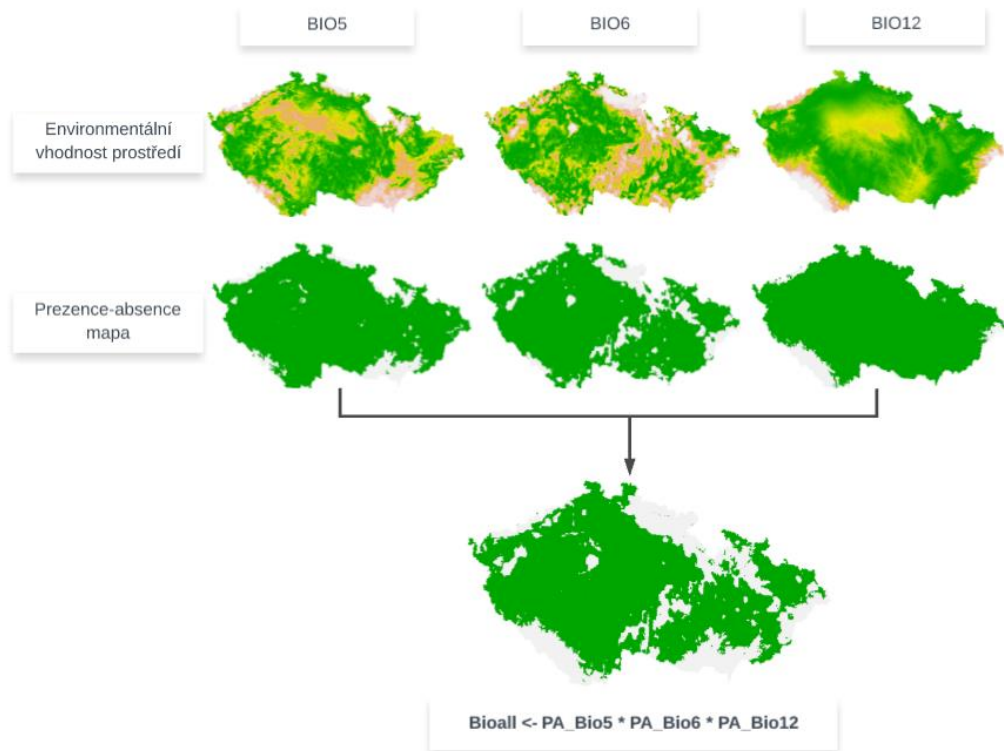
Prahová metoda

Při prahovém zobrazení byla aplikována tzv. prahová hodnoty 0.2, druh byl vykreslen, pokud se nacházel nad stanovenou hranicí prahu. Možnost opakování této modelace je omezená, proto byla simulace provedena 1x.

NicheLim metoda

NicheLim metoda je založena na prahovém zobrazení, ale na každou environmentální proměnnou je nahlíženo zvlášť. Nejprve byla aplikována odezвовá funkce (Gaussova křivka) na každou environmentální proměnnou – BIO5, BIO6, BIO12. Následně byla vygenerována environmentální vhodnost prostředí, na kterou byl aplikován tzv. práh (prahová hodnota zvolena 0.2).

Vytvořené presence-absence mapy environmentálních proměnných BIO5, BIO6, BIO12 byly vynásobeny za účelem vytvoření jedné presenčně-absenční mapy na kterou lze následně vybrat vzorky výskytu (viz Obr. 12).



Obr. 12 Schéma tvorby virtuálního druhu NicheLim metodou, na každou environmentální proměnnou je nejprve aplikována tzv. prahová hodnota a až poté jsou vrstvy spojeny.

(Zdroj: autor)

3.4.4 Vzorky výskytu

Výběr výskytových bodů byl vygenerován náhodným způsobem. Počet prezenčních záznamů byl stanoven na: 50, 100, 500, 1 000, 2 000. Absenční záznamy činily dvojnásobný počet prezenčních výskytů, tedy 100, 200, 1 000, 2 000, 4 000.

3.5 Algoritmus Maxent

Pro vytvoření SDM byl vybrán Maxent verze 3.4.1, který je volně stažitelný na webové stránce www.biodiversityinformatics.amnh.org. Elith et al. 2006 doporučuje jeho použití z důvodu přinášení obdobných či lepších výsledků v porovnání s jinými modelovacími algoritmy. Maxent pracuje s presence-only daty, které jsou častěji používaná při SDMs a lze je snadněji získat. Maxent je vhodný pro modelaci vzácných druhů, které lze špatně zmapovat či pro menší zájmové území. Maxent byl použit s packagem dismo a bylo ponecháno defaultní nastavení, se kterým Maxent poskytuje

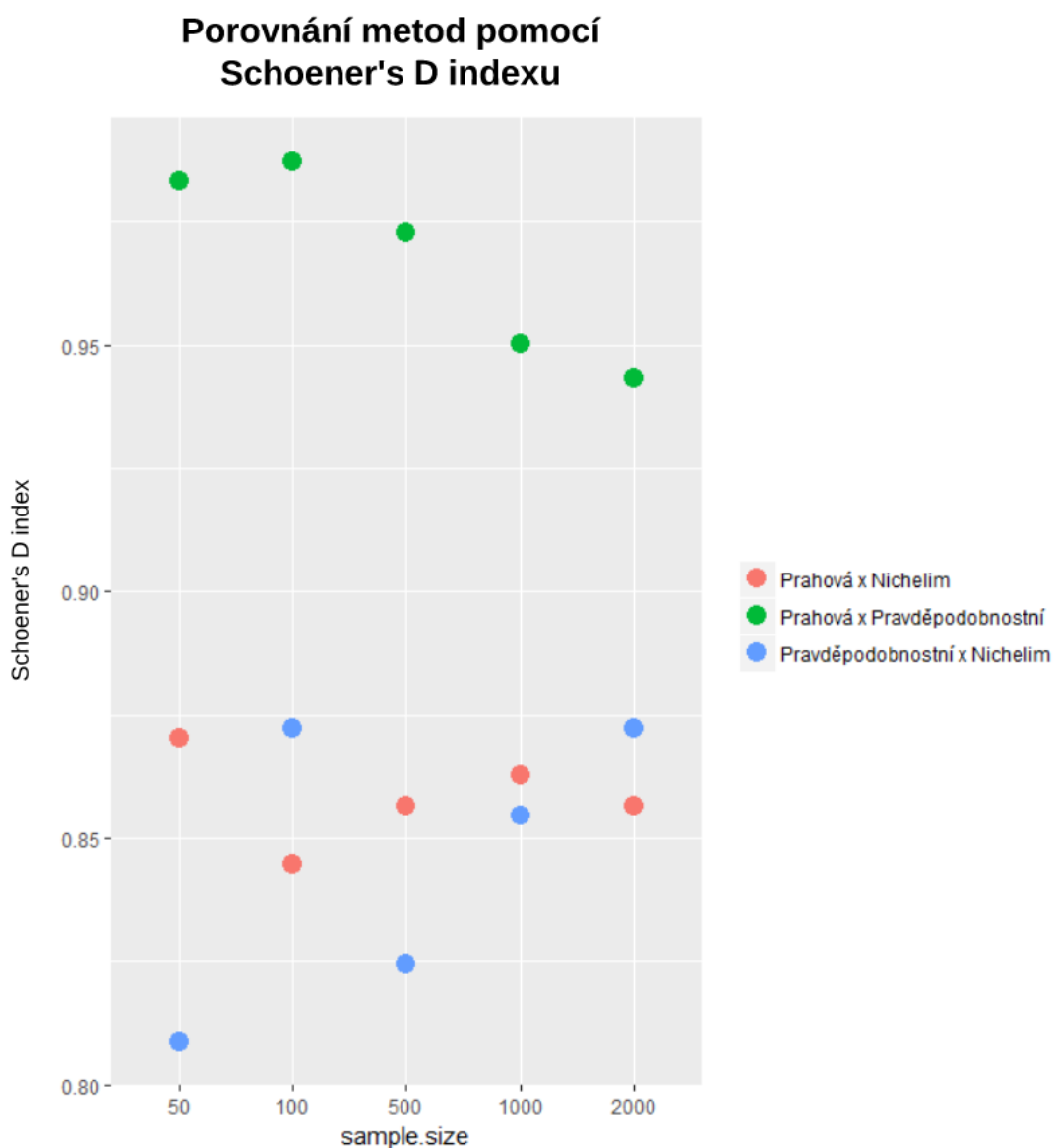
kvalitní predikci – „default prevalence“ 0,5 a „betamultiplier“ 10 (Phillips & Dudík, 2008).

3.6 Validační metriky

Data byla rozdělena náhodným způsobem v poměru 80 % kalibračních dat a 20 % validačních dat. Pro vyhodnocení schopnosti modelů byla nejprve vybrána metrika AUC. Pearson et al., (2007) doporučuje používat při validaci pomocí AUC i jiné metriky. Proto byla následně provedena metoda překryvu niky hodnocena Schoener's D indexem. Pro porovnání jednotlivých metod tvorby virtuálního druhu byl použit Schoener's D index.

4. VÝSLEDKY

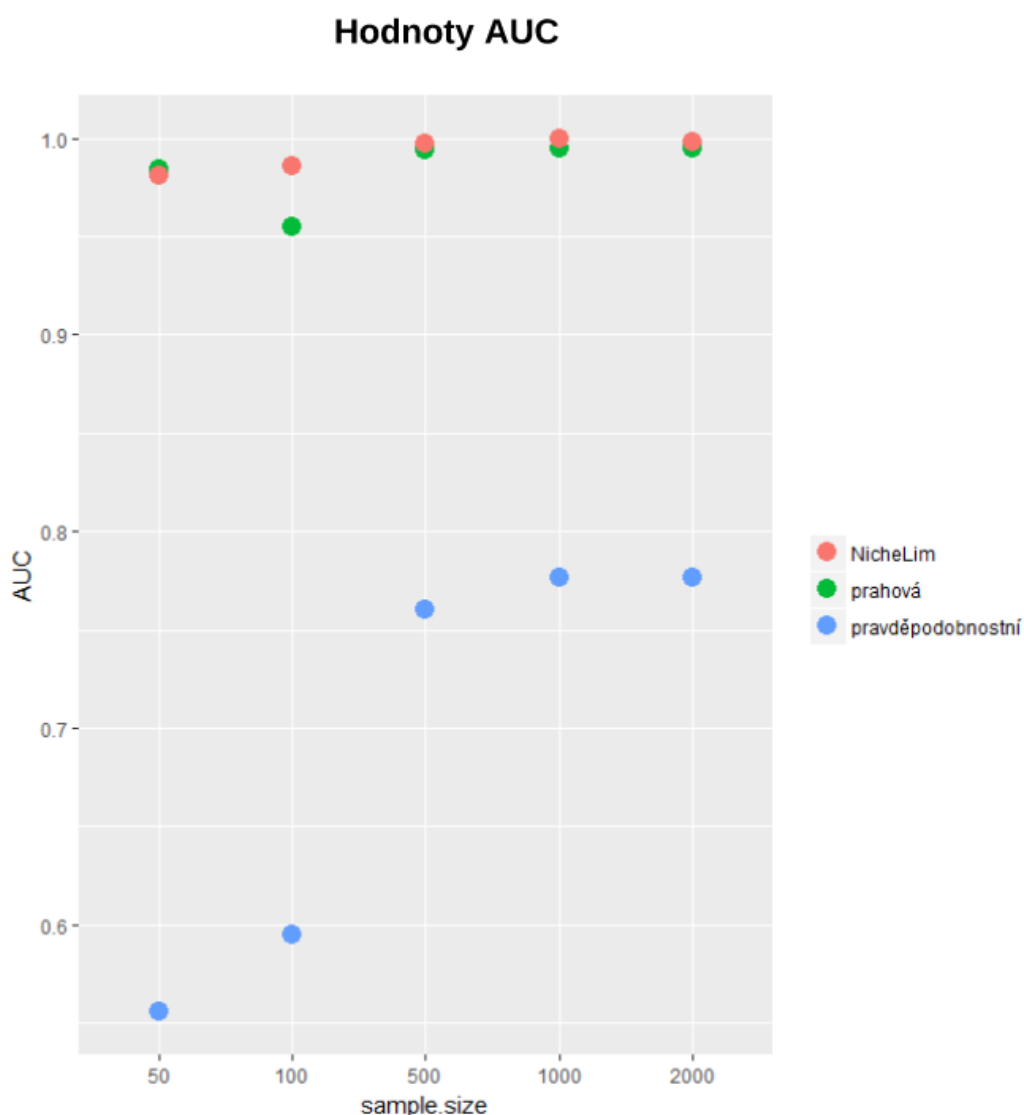
Nejprve byly porovnány metody generování virtuálního druhu mezi sebou pomocí Schoener's D indexu. Dále byla vyhodnocena schopnost modelu rozlišovat mezi presencemi a absencemi (validační metrika AUC) pro prahovou, pravděpodobnostní a NicheLim metodu zvlášť. Pomocí Schoener's D indexu byl vyhodnocen překryv virtuální vhodnosti prostředí s modelem druhové distribuce.



Obr. 13 Hodnoty Schoener's D indexu při porovnání prahové a NicheLim metody (červená barva), prahové a pravděpodobnostní metody (zelená barva), pravděpodobnostní a NicheLim metody (modrá barva)

(Zdroj: autor)

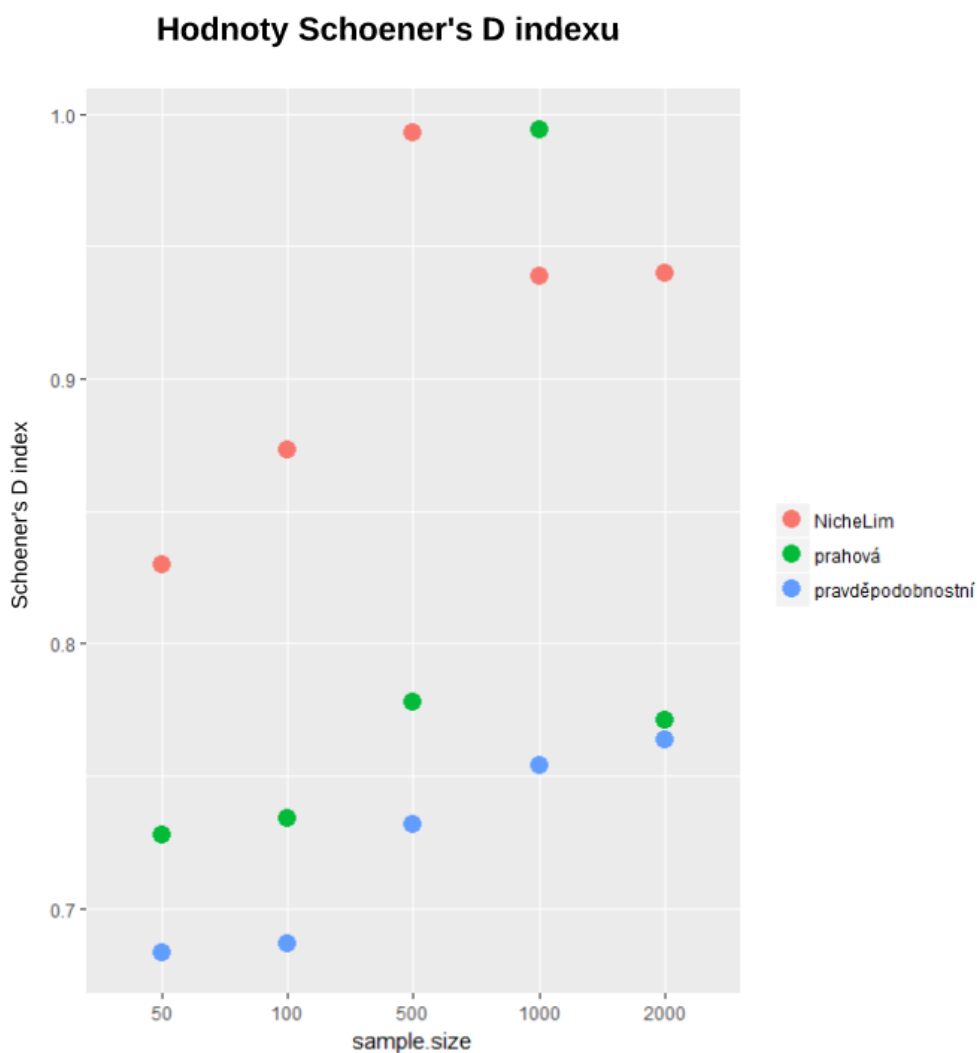
Pro možnost porovnání metod virtuálního druhu je potřeba provést analýzu skrze překryv niky. Komparace byla provedena nejprve pro prahovou a NicheLim metodu, dále pro prahovou a pravděpodobnostní metodu a poslední překryv byl proveden pro pravděpodobnostní a NicheLim metodu. Z výsledků zobrazených na obrázku 13 je patrné, že nejvyššímu překryvu je dosaženo při komparaci prahové a pravděpodobnostní metody. Pro snazší interpretaci výsledků lze použít klasifikaci překryvu niky (Tab. 4) od Rödder & Engler (2011). Při použití této klasifikace, lze vyhodnotit překryv, jako velmi vysoký při komparaci všech tří uvedených metod používaných pro generování virtuálního druhu.



Obr. 14 Hodnoty AUC pro pravděpodobnostní, prahovou a NicheLim metodu.

(Zdroj: autor)

Z obrázku 14 lze vyčíst, že nejvyšších hodnot AUC dosahuje NicheLim metoda, k výjimce dochází při sample size 50, kdy nejvyšší hodnotu- 0,9842 dosáhla prahová metoda. NicheLim metoda dosahuje konstantních hodnot od sample size 500 do 2 000, kdy výsledné hodnoty dosahují 0.9974, 0.9997, 0.9983. Zvyšující se velikost sample size má pozitivní vliv na pravděpodobnostní metodu, kde dochází k zřetelnému nárůstu hodnot AUC (sample size 50- hodnota AUC 0.5557 a při sample size 2 000 je hodnota AUC 0.7768). Pro snazší interpretaci výsledků lze použít klasifikaci prediktivní schopnosti AUC (Tab. 5) od Kienast et al. (2012). Podle této klasifikace dosahuje model při použití NicheLim metody výborné přesnosti při všech sample size, prahová metoda zobrazení dosahuje taktéž výborné schopnosti modelu, pravděpodobnostní metoda dosahuje při sample size 50, 100 nedostačující přesnosti, při sample size 500, 1 000, 2 000 dosahuje průměrné přesnosti modelu.



Obr. 15 Hodnoty Schoener's D indexu pro pravděpodobnostní, prahovou a NicheLim metodu
(Zdroj: autor)

Kromě AUC, které je často kritizováno (Pearson et al., 2007) byly modely vyhodnoceny druhou validační metrikou, a to překryvem niky hodnocenou Schoener's D indexem (viz Obr. 15). Nejvyšších hodnot překryvu dosahuje NicheLim metoda, k výjimce dochází při sample size 1 000, kdy má vyšší přesnost překryvu prahová metoda zobrazení (0.9940). Zvyšující se sample size má pozitivní vliv na hodnoty překryvu niky při použití pravděpodobnostní metody zobrazení (při sample size 50 dosahuje překryv niky 0.6833, při sample size 2 000 překryv niky činí 0.7637). Pro snažší interpretaci výsledků lze použít klasifikaci překryvu niky (Tab. 4) od Rödder & Engler (2011). Při použití této klasifikace, dosahuje NicheLim metoda velmi vysokému překryvu niky. Prahová metoda zobrazení dosahuje vysokého překryvu niky při sample size 50, 100, 500 a 2 000, při sample size 1 000 činí hodnota velmi vysokého překryvu. Pravděpodobnostní metoda dosahuje při všech sample size vysokého překryvu niky.

Všechny tabulky výsledných hodnot, AUC a Schoener's D indexu, jsou uvedené v příloze D.

5. DISKUZE

V posledních letech byl virtuální druh používán především ke studiu hodnotícím různé aspekty modelování: vliv kvality prostorových dat nebo vliv zvolené metody modelování na predikční výkon modelu. Virtuální druh umožňuje lépe pochopit různorodé aspekty ovlivňující výkony modelů (Naimi et al., 2011). Při tvorbě virtuálního druhu lze nasimulovat jakékoliv vztahy mezi druhy a environmentálními proměnnými (Leroy et al., 2015).

Cílem této práce bylo porovnat, jak výběr metody generování virtuálního druhu ovlivňuje výsledné modely druhové distribuce. Pro možnost porovnání bylo nezbytné vytvořit virtuální druh pravděpodobnostní, prahovou a NicheLim metodou. Do nedávna byla hojně využívána prahová metoda (Hilzer et al., 2001). Momentálně je ale stále více využívána pravděpodobnostní metoda z důvodu pozvolné reakce na environmentální proměnné a také díky možnosti opakování simulace (Meynard & Kaplan, 2012; Moudrý, 2015). Prahová metoda nepodporuje dynamické šíření druhů v prostoru. Meynard & Kaplan (2012) nedoporučuje používat prahovou metodu z důvodu nadhodnocování výkonu modelu. Mnoho modelovacích algoritmů je založeno na principech logistické křivky (například GAM a GLM), tyto algoritmy nelze používat při prahové metodě. Možným řešením, je použití algoritmů založených na jiných principech (Meynard & Kaplan, 2012). Dalším problémem prahové metody je nemožnost opakování simulaci, při použití pravděpodobnostní metody lze simulaci opakovat se stejnými či odlišnými podmínkami.

Pro možnost komparace pravděpodobnostní, prahové a NicheLim metody byla použita metrika Schoener's D indexu. Metody byly nejprve porovnány mezi sebou a z výsledných hodnot je patrné, že nejvyššímu překryvu bylo dosaženo při komparaci prahové a pravděpodobnostní metodě při sample size 100, s narůstajícím sample size se hodnota Schoener's D indexu snižovala. Při komparaci prahové a NicheLim metody bylo dosaženo největšího překryvu při sample size 100. Při komparaci pravděpodobnostní a NicheLim metody bylo dosaženo nejnižšímu překryvu, hodnoty se pohybovaly v rozmezí od 0.8086 do 0.8724.

Dále byly samotné modely validovány pomocí AUC a Schoener's D indexu. Metoda AUC je hojně využívána při výzkumech (Domínguez-Domínguez et al. (2006); Araújo & Luoto, 2007; Moudrý & Šimová, 2012, 2013; Kuemmerlen et al., 2014). AUC je ale kritizováno z důvodu nadhodnocování schopnosti modelů realizované distribuce. Jiménez-Valverde (2012) označuje AUC dokonce za nevhodnou validační metriku. Lobo et al. (2008) uvádí za nejzávažnější problém AUC, chybovost při určení podílu správně předpokládaných absencí a výsledných hodnot AUC. Pearson et al., (2007) doporučuje používat při validaci pomocí AUC i jiné kalibrační metody, proto byla použita i druhá validační metrika – překryv niky, respektive Schoener's D index. Validace modelu skrze překryv niky je hojně používaná již od svého založení (Pyron & Burbrink, 2009; Rödder & Lötters, 2009; McCormack et al., 2010; Medley, 2010; Nakazato et al., 2010). Rödder & Engler, (2011) doporučují používat pro hodnocení překryvu niky Schoener's D index, označuje ho za jeden z nejvhodnějších.

Pravděpodobnostní metoda dosahovala při validačních metrikách (AUC, Schoener's D index) nejnižších hodnot, zatímco NicheLim metoda dosahovala nejvyšších hodnot. Momentálně většina metodických studií doporučuje používat pravděpodobnostní metodu zobrazení (Meynard & Kaplan, 2012; Moudrý, 2015), z mé studie dosáhla tato metoda nejhorších výsledků, což vede k zamyšlení, jestli je to správný směr, kterým by se vědci měli ubírat.

Počet sample size (záznamy výskytu) má efekt na výsledný model. Nelze objektivně určit, jaká velikost sample size přináší nejpřesnější výsledky modelů, záleží na zvoleném modelovacím algoritmu. Stockwella & Peterson (2002) uvádí 50 záznamů, Wisz et al. (2008) se přiklání ke 100 záznamů, při kterých dochází k maximální přesnosti modelu. Hanberry et al. (2012) ve své studii porovnával modely s velikostí sample size od 30 do 2 500, z jeho výsledků je patrné, že modely s nižším sample size (<200) dosahují nepřesných hodnot. Mé výsledky se s touto studií shodují: při použití pravděpodobnostní metody hodnoty AUC činily 0.5945 (sample size 100), při vyšším sample size 500, hodnota AUC dosahovala 0.7599. Obecně lze z metodických studií usoudit, že čím větší sample size, tím bude přesnost modelu větší (Pearce & Ferrier 2000, Stockwell & Peterson 2002). Výsledky práce toto tvrzení dokazují, k nejvyšší přesnosti při použití pravděpodobnostní metody dochází při sample size 2 000 (hodnota AUC: 0.7768, Schoener's D index: 0.7637), při použití prahové metody

dosahuje model nejvyšší přesnosti při sample size 2 000 pro hodnotu AUC (0.9950) a při sample size 1 000 pro hodnotu Schoener's D indexu (0.9940), při použití NicheLim metody dosahuje model nejvyšší přesnosti při sample size 1 000 pro hodnotu AUC (0.9997) a pro hodnotu Schoener's D indexu při sample size 500 (0.9930). Lineárně narůstající přesnost modelu se zvyšujícím se sample size je zaznamenána při pravděpodobnostní metodě.

6. ZÁVĚR

Výsledky této studie jsou zaměřené na zkoumání vlivu použité metody generování virtuálního druhu na modely druhové distribuce. Virtuální druh byl vytvořen pravděpodobnostní, prahovou a NicheLim metodou. Pro vytvoření modelů druhové distribuce byl vybrán algoritmus Maxent. Validace modelů proběhla pomocí metriky AUC a překryvu niky klasifikovanou Schoener's D indexem. Metody generování virtuálního druhu byly porovnány Schoener's D indexem. Velmi vysokého překryvu bylo dosaženo při komparaci prahového a pravděpodobnostního zobrazení (0.9432–0.9872). Nejvyšší schopnosti modelu rozlišit mezi presencemi a absencemi (validace pomocí AUC) dosáhla NicheLim metoda. Při hodnocení schopnosti modelu skrze Schoener's D index byl nejvyšší překryv zaznamenán při použití NicheLim metody. Dále lze podotknout, že modely s vyšším počtem výskytů (sample size) přinášejí přesnější výsledky, u pravděpodobnostní metody byl zaznamenán lineární nárůst hodnot AUC a Schoener's D indexu (vyšší schopnost modelu) s navyšujícím se sample size. Virtuální druh je používán k testování a vylepšování predikčních schopností SDMs, proto výběr adekvátní metody virtuálního druhu představuje klíčový krok pro nové posuny v oblasti biogeografie. Závěrem lze říci, že můj výzkum poukazuje na vhodné použití NicheLim metody, která je založena na teorii ekologické niky a podporuje základy ekologických teorií (Huang et al., 2016). Nezbytný je další výzkum, při kterém by byla provedena komparace metod virtuálního druhu při použití jiného modelovacího algoritmu.

7. SEZNAM POUŽITÉ LITERATURY

Araújo, M., Luoto, M., 2007: "The importance of biotic interactions for modelling species distributions under climate change." *Global Ecology and Biogeography* 16.6, 743-753.

Araújo, M., Pearson, R., Thuiller, W., Erhard, M., 2005: Validation of species–climate impact models under climate change. *Global Change Biology* 11(9), 1504-1513.

Araújo, M., Guisan A., 2006: Five (or so) challenges for species distribution modelling. *Journal of biogeography* 33.10, 1677-1688.

Araújo, M., Rahbek, C., 2006: How does climate change affect biodiversity? *Science* 313.5792, 1396-1397.

Araújo, M., Marcon, J., Maccheroni, W., van Elsas, J., van Vuurde, J., Azevedo, J., 2002: Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. *Applied and environmental mikrobiology* 68(10), 4906-4914.

Austin, M., 2002: Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological modelling* 157.2-3, 101-118.

Austin, M., Belbin, L., Meyers, J., Doherty, M., Luoto, M., 2006: Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. *ecological modelling* 199(2), 197-216.

Bakkenes, M., Alkemade, J., Ihle, F., Leemans, R., Latour, J., 2002: Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global change biology* 8(4), 390-407.

Barbet-Massin, M., Jiguet, F., Albert, C., Thuiller, W., 2012: Selecting pseudo-absences for species distribution models: how, where and how many?. *Methods in ecology and evolution* 3(2), 327-338.

Begon M., Harper J., Townsend C., 1992: *Ecology: individuals, populations and communities*, Univerzita Palackého, Olomouc.

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(4), 640-651.

Busby, J., 1991: BIOCLIM – a bioclimate analysis and prediction system. *Plant protection quarterly*, 56-87.

Carpenter, G., Gillison, A., Winter, J., 1993: DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity & Conservation* 2(6), 667-680.

DeAngelis, D., Wolf, M., 2005: "Individual-based modeling of ecological and evolutionary processes." *Annu. Rev. Ecol. Evol. Sys.* 36, 147-168.

Domínguez-Domínguez, O., Martínez-Meyer, E., Zambrano, L., De, L., Pérez-Ponce, G., 2006: Using ecological-niche modeling as a conservation tool for freshwater species: live-bearing fishes in central Mexico. *Conservation Biology* 20(6), 1730-1739.

Dormann, F., Carsten, P., 2007: Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30.5, 609-628.

Duan, R., Kong, X., Huang, M., Fan, W., Wang, Z., 2014: The predictive performance and stability of six species distribution models. *PLoS One* 9(11), 112-764.

Duputié, A., Zimmermann, N., Chuine, I., 2014: Where are the wild things? Why we need better data on species distribution. *Global Ecology and Biogeography* 23(4), 457-467.

Edwards, A., Ritter, R., Abel, K., Manning, A., Panhuysen, C., Farrer, L., 2005: Complement factor H polymorphism and age-related macular degeneration. *Science* 308, 421-424.

Elith, J., Graham, C., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., Li, J., 2006: Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 129-151.

Elith, J., Leathwick, J., 2009: "Species distribution models: ecological explanation and prediction across space and time." *Annual review of ecology, evolution, and systematics* 40, 677-697.

Fielding, A., Bell, J., 1997: A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation* 24(1), 38-49.

Fleishman, E., Ray, C., Sjögren-Gulve, P., Boggs, C., Murphy, D., 2002: Assessing the roles of patch quality, area, and isolation in predicting metapopulation dynamics. *Conservation Biology* 16(3), 706-716.

Fourcade, Y., Engler, J., Rödder, D., Secondi, J., 2014: Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS one* 9(5), 97-122.

Franklin J., Miller J., 2009: Mapping species distributions: spatial inference and prediction. Cambridge: Cambridge University Press.

Franklin, J., 1995: Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in physical geography* 19.4, 474-499.

- Franklin, J., 2010:** Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions* 16(3), 321–330.
- Goodchild, M., 1994:** Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues. *Journal of Vegetation Science* 5(5), 615-626.
- Goodchild, Michael F., 2011:** Scale in GIS: An overview. *Geomorphology* 130, 5-9.
- Guisan, A., Zimmermann, N., 2000:** Predictive habitat distribution models in ecology. *Ecological modelling* 135.2-3, 147-186.
- Guisan, A., Edwards Jr, T., Hastie, T., 2002:** Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling* 157(2-3), 89-100.
- Guisan, A., Thuiller, W., 2005:** "Predicting species distribution: offering more than simple habitat models." *Ecology letters* 8.9, 993-1009.
- Haefner Jr, P., 1970:** The effect of low dissolved oxygen concentrations on temperature-salinity tolerance of the sand shrimp, *Crangon septemspinosa* Say. *Physiological Zoology* 43.1, 30-37.
- Hanberry, B., He, H., Dey, D., 2012:** Sample sizes and model comparison metrics for species distribution models. *Ecological Modelling* 227, 29–33.
- Heath, J., Smith, A., 1989:** Growth factors in embryogenesis. *British medical bulletin* 45.2, 319-336.
- Heikkinen, R., Luoto, M., Virkkala, R., 2007:** Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography* 16(6), 754–763.
- Hijmans, R., Cameron, S., Parra, J., Jones, P., Jarvis, A., 2005:** Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology* 25(15), 1965-1978.
- Hirzel, A., Hausser, J., Chessel, D., Perrin, N., 2002:** Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83(7), 2027-2036.
- Hirzel, A., Helfer, V., Metral, F., 2001:** Assessing habitat-suitability models with a virtual species. *Ecological modelling* 145(2-3), 111-121.
- Huang, M., Kong, X., Valera, S., Duan, R., 2016:** The Niche Limitation Method (NicheLim), a new algorithm for generating virtual species to study biogeography. *Ecological modelling* 320, 197-202.

- Huntley, B., Berry, P., Cramer, W., McDonald, A., 1995:** Special paper: modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography*, 967-1001.
- Hutchinson G., 1957:** A Treatise on Limnology. JOHN WILEY & SONS, New yourk.
- Chase J., Leibold M., 2003:** Ecological niches: linking classical and contemporary approaches. University of Chicago Press, London.
- Jiménez-Valverde, A., 2012:** Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21(4), 498-507.
- Jiménez-Valverde, A., Lobo, J., 2007:** Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta oecologica* 31(3), 361-369.
- Kienast, F., Bolliger, J., Zimmermann, N., 2012:** Advanced Landscape Ecology: Lecture Notes, 1-12.
- Kozak, K., Graham, C., Wiens, J., 2008:** "Integrating GIS-based environmental data into evolutionary biology." *Trends in ecology & evolution* 23.3, 141-148.
- Krzanowski W., Hand D., Wojtek J., 2009:** ROC curves for continuous data. CRC Press, Boca Raton.
- Kuemmerlen, M., Schmalz, B., Guse, B., Cai, Q., Fohrer, N., Jähnig, S., 2014:** Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. *Ecological Modelling* 277, 77-86.
- Leathwick, J., 1998:** "Are New Zealand's *Nothofagus* species in equilibrium with their environment?." *Journal of Vegetation Science* 9.5, 719-732.
- Leathwick, J., Rowe, D., Richardson, 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(12), 2034-2052.
- Leroy, B., Meynard, C., Bellard, C., Courchamp, F., 2015:** virtualspecies, an R package to generate virtual species distributions. *Ecography* 39(6), 599-607.
- McCormack, J., Zellmer, A., Knowles, L., 2010:** Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution* 64(5), 1231-1244.
- Medley, K., 2010:** Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. *Global ecology and biogeography* 19(1), 122-133.

- Meynard, C., Kaplan, D., 2012:** The effect of a gradual response to the environment on species distribution modelling performance. *Ecography* 35(6), 499-509.
- Meynard, C., Kaplan, D., 2013:** Using virtual species to study species distributions and model performance. *Journal of Biogeography* 40(1), 1-8.
- Meynard, C., Quinn, J., 2007:** Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography* 34(8), 1455-1469.
- Miller J., 2010:** Species Distribution Modelling. *Geography Compass* 4(6), 490-509.
- Miller, J., 2014:** Virtual species distribution models: using simulated data to evaluate aspects of model performance. *Progress in Physical Geography* 38(1), 117–128.
- Miller, J., Rogan, J., 2007:** "Using GIS and remote sensing for ecological mapping and monitoring." *Integration of GIS and remote sensing*, 233-268.
- Moore, I., Grayson, R., Ladson, A., 1991:** Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes* 5(1), 3-30.
- Moudrý V., 2015:** Modelling species distributions with simulated virtual species. *Journal of Biogeography* 42(8), 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(11), 2083-2095.
- Moudrý, V., Šímová, P., 2013:** Relative importance of climate, topography, and habitats for breeding wetland birds with different latitudinal distributions in the Czech Republic. *Applied Geography* 44, 165-171.
- Naimi, B., Skidmore, A., Groen, T., Hamm, N., 2011:** Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of Biogeography* 38(8), 1497-1509.
- Nakazato, T., Warren, D., Moyle, L., 2010:** Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany* 97(4), 680-693.
- Pearce, J., Ferrier, S., 2000:** Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133, 225–245.
- Pearson, R., Dawson, T., 2003:** Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?. *Global ecology and biogeography* 12(5), 361-371. .
- Pearson, R., Raxworthy, C., Nakamura, M., Peterson, A., 2007:** Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of biogeography* 34(1), 102-117.

- Pepe, M., 2000:** An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* 56(2), 352-359.
- Peterson, A., 2003:** Predicting the geography of species' invasions via ecological niche modeling. *The quarterly review of biology* 78(4), 419-433.
- Peterson, A., Vieglais, D., 2001:** Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem: A new approach to ecological niche modeling, based on new tools drawn from biodiversity informatics, is applied to the challenge of predicting potential species' invasions. *BioScience* 51(5), 363-371.
- Phillips, S., Anderson, R., Schapire, R., 2006:** Maximum entropy modeling of species geographic distributions. *Ecological modelling* 190.3-4, 231-259.
- Phillips, S., Dudík, M., 2008:** Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31(2), 161-175.
- Phillips, S., Dudík, M., Schapire, R., 2004:** A maximum entropy approach to species distribution modeling. In *Proceedings of the twenty-first international conference on Machine learning*, 83.
- Pulliam, H., 2000:** On the relationship between niche and distribution. *Ecology letters* 3.4, 349-361.
- Pyron, A., Burbrink, F., 2009:** Lineage diversification in a widespread species: roles for niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Molecular ecology* 18(16), 3443-3457.
- Robin, E., Guisan, A., Rechsteiner, L., 2004:** An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of applied ecology* 41.2, 263-274.
- Rödger, D., Engler, J., 2011:** Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography* 20(6), 915-927.
- Rödger, D., Lötters, S., 2009:** Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Global Ecology and Biogeography* 18(6), 674-687.
- Santika, T., Hutchinson, M., 2009:** The effect of species response form on species distribution model prediction and inference. *Ecological Modelling* 220(19), 2365-2379.
- Saupe, E., Barve, V., Myers, C., Soberón, J., Barve, N., Hensz, C., Lira-Noriega, A., 2012:** Variation in niche and distribution model performance: the need for a priori assessment of key causal factors. *Ecological Modelling* 237, 11-22.

Segurado, P., Araújo M., 2004: An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31.10, 1555-1568.

Schoener, T., 1970: Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology* 51(3), 408-418.

Sobéron, J., Peterson, A., 2005: Interpretation of models of fundamental ecological niches and species' distributional areas. *Ecology letters* 8.6, 533-550.

Stickel, L., 1954: A comparison of certain methods of measuring ranges of small mammals. *Journal of Mammalogy* 35.1 (1954): 1-15.

Stockwell, D., 1999: The GARP modelling system: problems and solutions to automated spatial prediction. *International journal of geographical information science* 13(2), 143-158.

Stockwell, D., Beach, J., Stewart, A., Vorontsov, G., Vieglais, D., Pereira, R., 2006: The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecological modelling* 195(1-2), 139-145.

Stockwell, D., Peterson, A., 2002: Effects of sample size on accuracy of species distribution models. *Ecological modelling* 148(1), 1-13.

Thuiller, W., Albert, C., Araújo, M., Berry, P., Cabeza, M., Guisan, A., Sykes, M., 2008: Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in plant ecology, evolution and systematics* 9(3-4), 137-152.

Townsend, C., Begon M., Harper J., 2008: *Essentials of ecology*. 3rd ed. Malden, MA: Blackwell Pub.

Václavík T., 2013: Modelování habitatu a distribuce druhů [online]. [cit. 2018-04-04]. Dostupné z: http://tova.euweb.cz/downloads/08_KEKO_sdm.pdf

Václavík, T., Meentemeyer, R., 2012: Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions* 18(1), 73-83.

Valera, S., Gomez-Rodriguez, M., 2015: Modeling adoption and usage of competing products. In *Data Mining (ICDM)*, IEEE International Conference, 409-418.

Valera, S., Anderson, R., García-Valdés, R., Fernández-González, F., 2014: Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37(11), 1084-1091.

Warren, D., Glor, R., Turelli, M., 2008: Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62(11), 2868-2883.

Wisz, M., Hijmans, R., Li, J., Peterson, A., Graham, C., Guisan, A., 2008: Effects of sample size on the performance of species distribution models. *Diversity and distributions* 14(5), 763-773.

Zurell, D., Berger, U., Cabral, J., Jeltsch, F., Meynard, C., Münkemüller, T., Grimm, V., 2010: The virtual ecologist approach: simulating data and observers. *Oikos* 119(4), 622-635.

Zimmermann, N., Edwards, T., Graham, C., Pearman, P., Svenning, J., 2010: New trends in species distribution modelling. *Ecography* 33(6), 985-989.

8. SEZNAMY

8.1 Obrázky

Obr. 1 Diagram „BAM	13
Obr. 2 Praktický příklad dvourozměrné ekologické niky	15
Obr. 3 Ukázka ekologické valence	16
Obr. 4 Digitální model terénu.....	19
Obr. 5 Proces SDM	23
Obr. 6 Příklad ROC křivky.....	25
Obr. 7 Chybová matice	26
Obr. 8 Tvorba virtuálního druhu.....	29
Obr. 9 Porovnání prahové a pravděpodobnostní	31
Obr. 10 Schéma tvorby virtuálního druhu.....	34
Obr. 11 Vztahy mezi environmentálními proměnnými a virtuálním druhem.....	35
Obr. 12 Schéma tvorby virtuálního druhu NicheLim metodou.....	37
Obr. 13 Hodnoty Schoener's D indexu – srovnání metod.....	39
Obr. 14 Hodnoty AUC	40
Obr. 15 Hodnoty Schoener's D indexu	41

8.2 Tabulky

Tab. 1 Přehled environmentálních dat používaných při SDMs	20
Tab. 2 Zdroje výskytových dat	21
Tab. 3 Klasifikace hodnot AUC.....	25
Tab. 4 Klasifikace hodnot Schoener's D indexu.....	26
Tab. 5 Přehled základních validačních metod	27
Tab. 6 Vstupní data reprezentují následující environmentální proměnné.....	32

8.3 Přílohy

Příloha A: Vstupní environmentální data

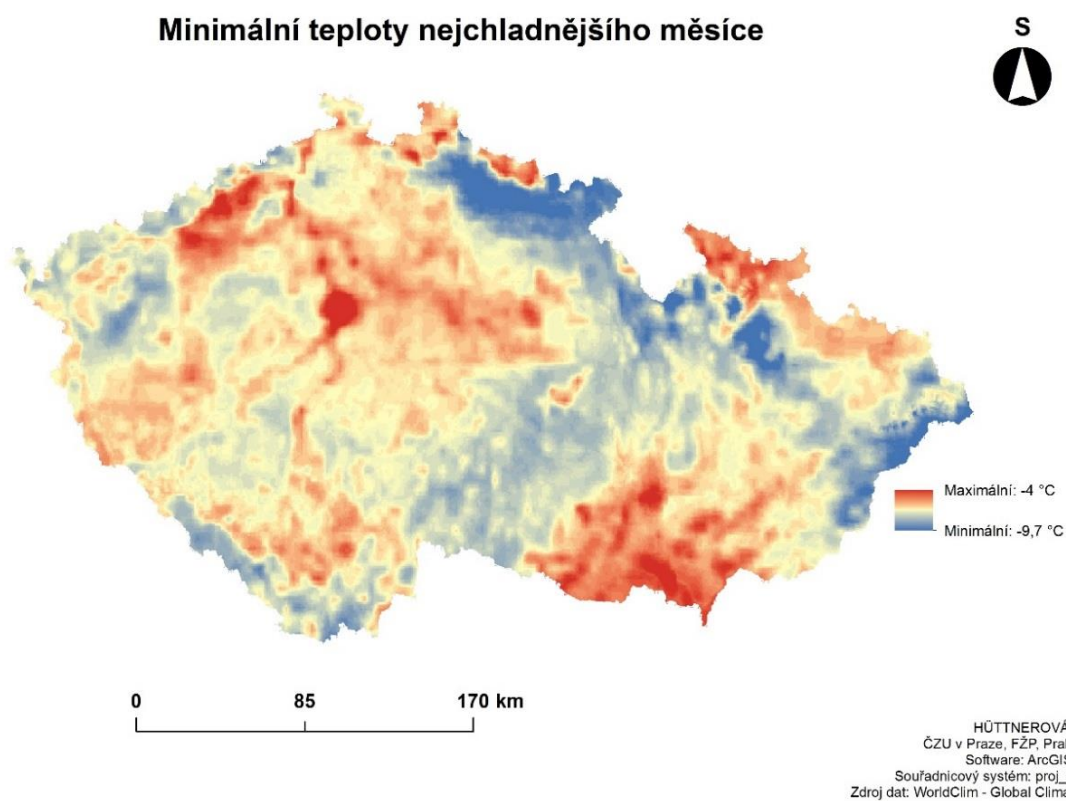
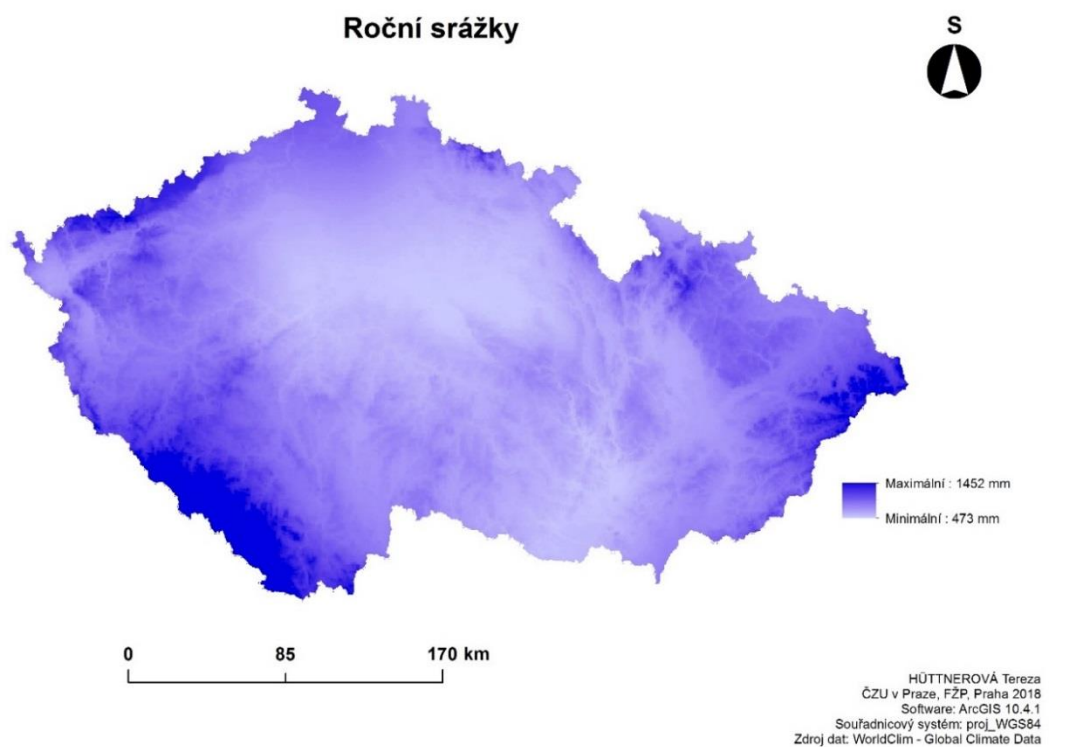
Příloha B: Environmentální vhodnost prostředí

Příloha C: Presence – absence mapy

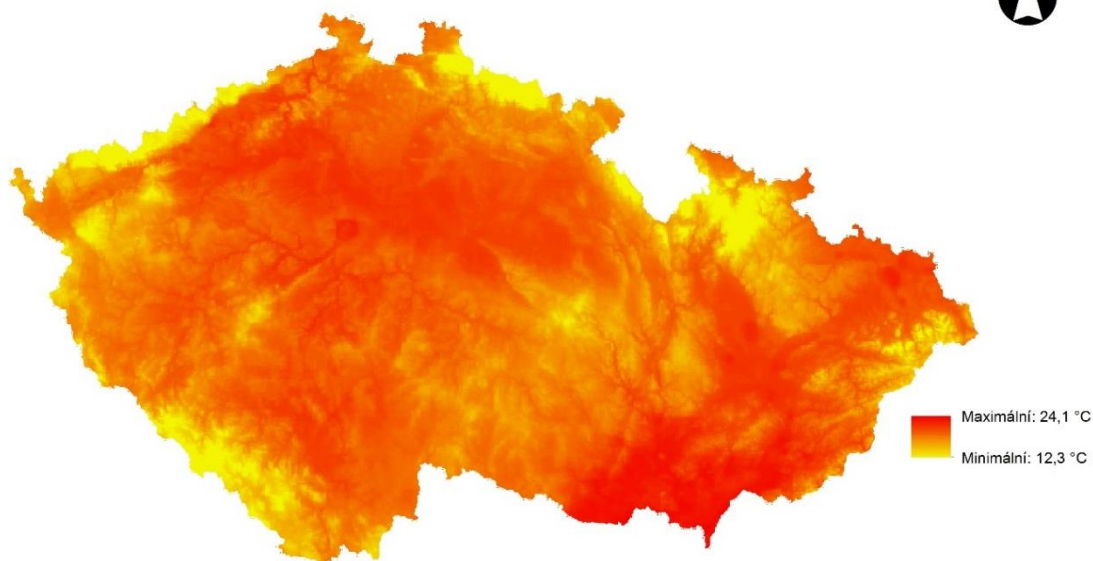
Příloha D: Tabulky s hodnoty AUC, Schoener's D indexu

9. PŘÍLOHY

Příloha A: Vstupní environmentální data, zdroj dat: databáze Worldclim



Maximální teploty nejteplejšího měsíce

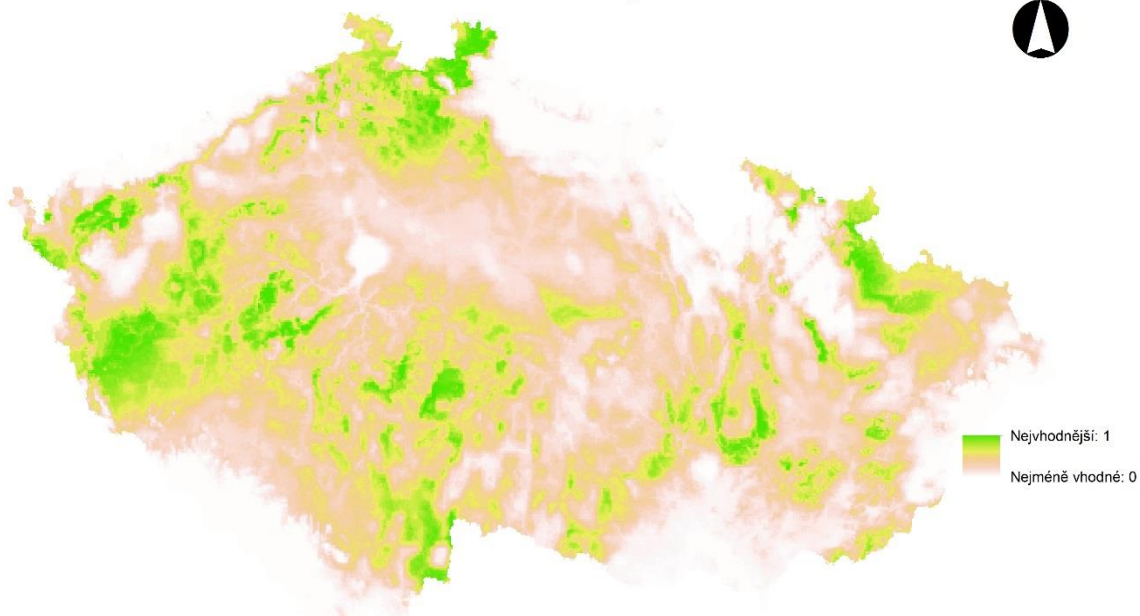


0 85 170 km

HÜTTNEROVÁ Tereza
ČZU v Praze, FŽP, Praha 2018
Software: ArcGIS 10.4.1
Souřadnicový systém: proj_WGS84
Zdroj dat: WorldClim - Global Climate Data

Příloha B: Environmentální vhodnost prostředí

Environmentální vhodnost prostředí

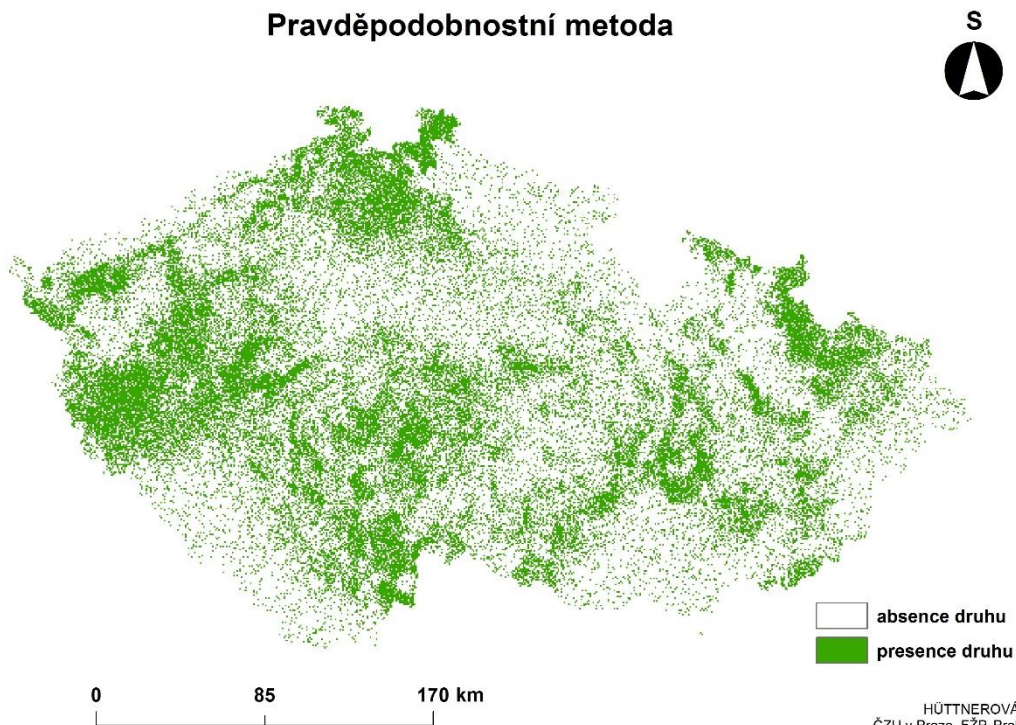


0 75 150 km

HÜTTNEROVÁ Tereza
ČZU v Praze, FŽP, Praha 2018
Software: ArcGIS 10.4.1
Souřadnicový systém: proj_WGS84
Zdroj dat: WorldClim - Global Climate Data

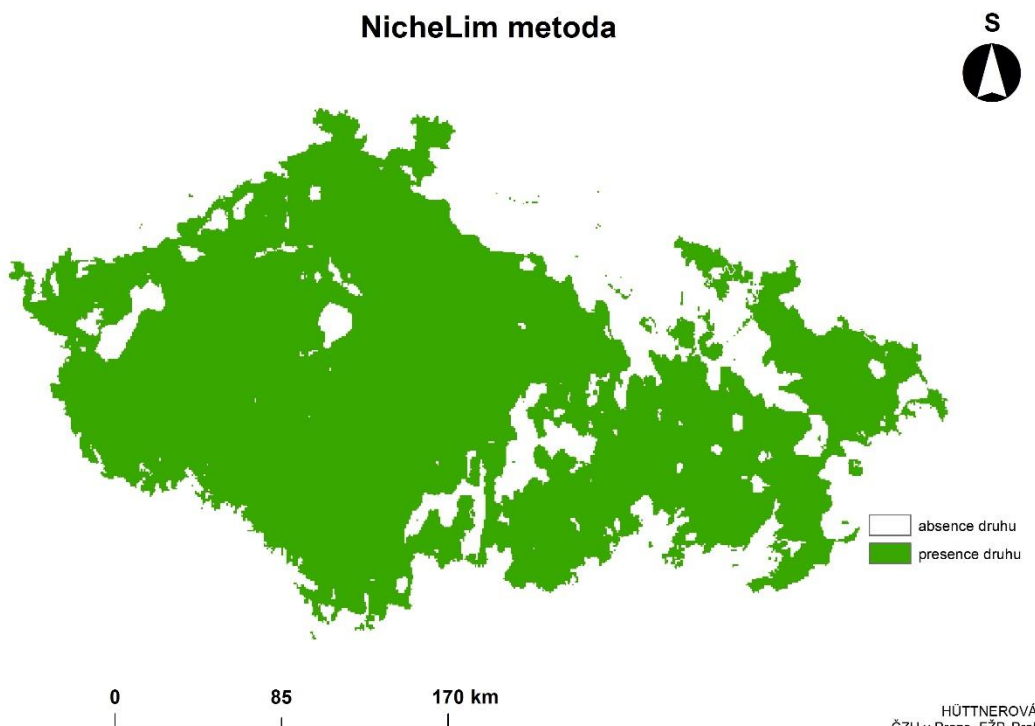
Příloha C: Presence-absence mapy

Pravděpodobnostní metoda



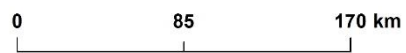
HÜTTNEROVÁ Tereza
ČZU v Praze, FŽP, Praha 2018
Software: ArcGIS 10.4.1
Souřadnicový systém: proj_WGS84
Zdroj dat: WorldClim - Global Climate Data

NicheLim metoda



HÜTTNEROVÁ Tereza
ČZU v Praze, FŽP, Praha 2018
Software: ArcGIS 10.4.1
Souřadnicový systém: proj_WGS84
Zdroj dat: WorldClim - Global Climate Data

Prahová metoda



HÜTTNEROVÁ Tereza
ČZU v Praze, FŽP, Praha 2018
Software: ArcGIS 10.4.1
Souřadnicový systém: proj_WGS84
Zdroj dat: WorldClim - Global Climate Data

Příloha D: Tabulky s hodnoty AUC a Schoener's D indexem

Tab. D1 Srovnání jednotlivých metod pomocí Schoener's D indexu

	Prahová x Pravděpodobnostní	Pravděpodobnostní x NicheLim	Prahová x NicheLim
Sample.size	Schoener's D	Schoener's D	Schoener's D
50	0.9834	0.8086	0.8701
100	0.9872	0.8724	0.8446
500	0.9727	0.8243	0.8566
1000	0.9502	0.8545	0.8626
2000	0.9432	0.8723	0.8563

Tab. D2 Hodnoty AUC a Schoener's D indexu pro pravděpodobnostní metodu

Sample	ø AUC	ø Schoener's D
50	0,5557	0,6833
100	0,5945	0,6868
500	0,7599	0,7316
1000	0,7766	0,7539
2000	0,7768	0,7637

Tab. D3 Hodnoty AUC a Schoener's D indexu pro NicheLim metodu

Sample	AUC	Schoener's D
50	0,9810	0,8300
100	0,9863	0,8730
500	0,9974	0,9930
1000	0,9997	0,9390
2000	0,9983	0,9400

Tab. D4 Hodnoty AUC a Schoener's D indexu pro prahovou metodu

Sample	AUC	Schoener's D
50	0,9842	0,7280
100	0,9553	0,7340
500	0,9937	0,7780
1000	0,9948	0,9940
2000	0,9950	0,7710