Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

SROVNÁNÍ VÝPOČETNÍCH ALGORITMŮ PRO PROSTOROVOU DISTRIBUCI DRUHŮ

Bakalářská práce

Eliška VLČKOVÁ

Vedoucí práce Mgr. Jitka Doležalová

Olomouc 2016
Geoinformatika a geografie

ANOTACE

Tato bakalářská práce se zabývá tématem prostorové distribuce druhů. Je představen základní úvod do problematiky včetně vymezení hlavních pojmů. Hlavní prostor je věnován algoritmům dostupným skrz programy Open Modeller a Biomod včetně zaměření na jejich vstupní požadavky, princip výpočtu a oblast nasazení. Na základě rozboru algoritmů a s ohledem na aspekty důležité při modelovacím procesu jsou sestaveny dva rozhodovací stromy určené k výběru vhodného algoritmu uživatelem. V praktické části práce jsou vybrané algoritmy simulovány na čtyřech druzích motýlů v CHKO Bílé Karpaty.

KLÍČOVÁ SLOVA

Prostorová distribuce druhů; OpenModeller; Biomod; algoritmus, rozhodovací strom

Počet stran práce: 56

Počet příloh: 24 (z toho 5 volných a 3 elektronické)

ANOTATION

This bachelor thesis deals with the topic of species distribution modelling (SDM). Firstly, there are desribed the main terms and basic introduction to the issue is done. The aim of this thesis is to focuse on different approaches and different algorithms which are available in software OpenModeller and package Biomod. Based on that, there are constructed two decision trees for potentional users to simplify the process of selection the appropriate method. In decision trees was considered the aim of experiment, quality of data and primary type of input data. Furthermore, there was a few algorithms that was selected to demonstrate their functions on. Species distribution models were applied to predict the distribution of four butterfly species in CHKO Bílé Karpaty area.

KEYWORDS

Species distribution modeling (SDM); OpenModeller; Biomod; algorithm; decision tree

Number of pages: 56

Number of appendixes: 24

Prohlašuji, že

- bakalářskou/diplomovou práci včetně příloh, jsem vypracovala samostatně a uvedla jsem všechny použité podklady a literaturu.

- jsem si vědoma, že na moji bakalářskou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,

- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, bakalářskou práci užívat (§ 35 odst. 3),

- souhlasím, aby jeden výtisk bakalářské práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

- souhlasím, že údaje o mé bakalářské práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé bakalářské práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé bakalářské práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne 12.8.2016 Eliška Vlčková

Poděkování

Děkuji vedoucímu práce Mgr. Jitce Doležalové za podněty a připomínky při vypracování práce. Dále děkuji doc. RNDR. Vilému Pechancovi, Ph.D za konzultace.

OBSAH

SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Význam
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AUC	Area Under the ROC Curve
BIC	Bayesian Information Criterion
BIOMOD	Biodiversity Modeling
BRT	Boosted Regression Trees
CTA	Classification Tree Analysis
CSV	Comma Separated Value
DMR4G	digitální model reliéfu 4. generace
ENFA	Ecological Niche Factor Analysis
FDA	Flexible Discriminant Analysis
GAM	Generalized Additive Models
GARP	Genetic Algorithm for Rule-Set Production
GBM	Generalized Boosted Networks
GIS	Geografický informační systém
GLM	Generalized Linear Models
JAR	Jihoafrická republika
LIDAR	Light Detection and Ranging
LDA	Linear Discriminant Analysis
MARS	Multivariate Adaptive Regression Trees
MAXENT	Maximum Entropy
MDA	Mixture Discriminant Analysis
PCA	Principal Component Analysis
PD-Map	Potential Distribution Map
PD-Model	Potential Distribution Model
PSU	Practical Salinity Unit
RF	Random Forests
ROC	Relative Operating Characteristic
SDM	Species Distribution Modeling
SHP	Esri Shapefile
SNN	Simulated Neural Network
SRE	Surface Range Envelope
SVM	Support Vector Machines
SW	Software
TSS	True Skill Statistic

ÚVOD

Jaký bude potenciální dopad klimatické změny na živočišné druhy? Jaké bude riziko expanze zavlečeného invazivního druhu? Jaká bude potenciální expanze přenašečů smrtelných onemocnění? Jak moc velký podíl má člověk na masovém vymírání druhů? Všechny tyto otázky mají jedno společné – k jejich odpovědi je využito metod pro modelování potenciální distribuce druhů (species distribution modeling – SDM). Se stále větší komplexností a složitostí dnešního světa jsou tyto data miningové metody velice mocným nástrojem v hledání závislostí tam, kde to člověk nedokáže. Rychle se vyvíjející informační technologie spolu se zlepšujícími se metodami dálkového průzkumu Země značně usnadňují procesy modelování potenciální distribuce druhů stejně jako stále lepší dostupnost kvality a sběru dat.

Práci lze považovat za jakýsi úvod do prostorové distribuce druhů včetně přehledu algoritmů dostupných skrz dva největší programy věnující se této problematice – program OpenModeller a balík funkcí implementovaný do programu Rstudio s názvem Biomod. Jelikož se v české literatuře této problematice co do komplexnosti nikdo nevěnuje, je tato práce vhodným dokumentem, po které by měl uživatel-začátečník sáhnout. Uvádí ucelený přehled pojmů, teoreticky vymezuje algoritmy dostupné skrze dva výše uvedené programy a v neposlední řadě díky dvěma rozhodovacím stromům napomáhá uživateli v rozhodnutí, jaký algoritmus použít s ohledem na cíl experimentu a typ a kvalitu vstupních dat. Použití vybraných algoritmů v závěru demonstruje na čtyřech druzích motýlů, jejichž výsledky jsou také vhodně zhodnoceny.

1 CÍLE PRÁCE

Hlavním cílem bakalářské práce je tvorba rozhodovacího stromu pro výběr adekvátního algoritmu dostupného v programech OpenModellech a Biomod v závislosti na cíli a kvalitě vstupních dat. Mezi teoretické cíle patří podrobá rešerše literatury a zpracování teoretických východisek algoritmů. Mezi cíle v praktické části je zařazena simulace na vybraných ecosystem providers dostupných skrz portál florabase.cz. Oproti tomuto cíli byla změněna testovací data a algoritmy byly simulovány na nálezových datech čtyř druhů motýlů převzatých z diplomové práce Sylvie Hartmannové. K dosažení hlavního cíle bakalářské práce byly zohledněny aspekty důležité při modelování prostorové distribuce druhů, na jejichž základě byly vytvořeny dva rozhodovací stromy.

Praktické dílčí cílé bakalářské práce jsou:

- ✓ vypracování rešerše,
- ✓ teoretické vymezení dostupných algoritmů,
- ✓ simulace vybraných algoritmů a zhodnocení výsledků,
- ✓ zohlednění aspektů důležitých při modelování prostorové distribuce druhů,
- ✓ vytvoření rozhodovacího stromu.

2 METODY A POSTUPY ZPRACOVÁNÍ

Použité metody

Protože práce je primárně o teoretickém vymezení algoritmů, použité metody lze zmínit pouze v procesu simulací. Pro zpracování dat bylo využito základních geoprocessingových nástrojů s cílem získat jednotné rozlišení jednotlivých environmentálních vrstev včetně jejich převzorkování na požadovanou velikost buňky. K nálezových datům motýlů byly ke každému bodu přiřazeny souřadnice a následně byla atributová tabulka vyexportována do formátu xlsx. Tabulky byly jednotlivě vhodně upraveny pro vstup do programu OpenModeller a následně i do programu Biomod, kde bylo hlavně nutné nahradit desetinné čárky za tečky.

Pro zhodnocení výsledků z programu OpenModeller byly výsledné predikce pravěpodobnosti výskytu ve formátu ASCII nahrány do prostředí GIS. Z nich byly v první řadě vytvořeny mapové výstupy pro každý algoritmus. V dalším bodě byly použitím reklasifikace vybrány ty buňky, kde byla pravděpodobnost výskytu druhu větší než 80 %. Pomocí raster calculatoru byl ve finále vytvořen jeden z hlavních cílů práce – statistický report.

Použitá data

V práci byly použity celkem dva druhy dat – oboje převzaty z diplomové práce Sylvie Hartmannové (Hartmannová, 2016). Jako nálezová data byly použity 4 bodové vrstvy ve formátu Esri Shapefile zobrazující výskyt 4 druhů motýlů: babočky jilmové (*Nymphalis polychloros*), bělopáska topolového (*Limenitis populi*), modráska hnědoskvrnného (*Polyommatus Dafnis*) a perleťovce většího (*Argynnis aglaja*). Tato dato vznikla v rámci projektu „Analýza biodiverzity v CHKO Bílé Karpaty jako podklad pro stanovení nové zonace vhodného managementu cenných území“ (2003–2006). Jedním z výstupů projektu je i Atlas rozšíření vybraných druhů živočichů CHKO Bílé Karpaty vyobrazující mapy rozšíření vybraných druhů motýlů, střevlíkovitých brouků a hnízdících ptáků. Druhým typem dat, jež byly v práci použity bylo 6 rastrových vrstev obsahující environmentální údaje na území CHKO Bílé Karpaty.

Vrstvy byly následující:

- digitální model reliéfu 4. generace (DMR4G),
- sklon,
- orientace,
- teplota (průměrná teplota v červnu 2000),
- geologické podloží,
- využití krajiny (landuse).

Použité programy

Hlavní část práce byla zpracována v gisových programech ArcMap 10.2 a ArcGIS Pro od společnosti ESRI. V programu ArcMap 10.2 byly převáděny a upravovány vrstvy do potřebných formátů. Následně v něm byly analyzovány výsledky jednotlivých algoritmů, a to primárně pomocí funkce Raster Calculator a Reclassify. Program ArcGIS Pro byl využit pro vygenerování hexagonální sítě, pro jehož potřebu bylo nutné importovat toolbox Create Hexagon Tessellation.

V tabulkových procesorech Microsoft Excel a LibreOffice Calc byla upravována a následně konvertována nálezová data čtyř druhů motýlů.

Pro SDM byly využity dva programy: OpenModeller Desktop v1.1.0, který nad sebou dovoluje vícero rozhraní, jako například příkazové řádky, desktopové rozhraní, webové rozhraní a rozhraní webové služby. Pro největší jednoduchost bylo k práci vybráno desktopové rozhraní programu. Druhým programem použitým pro demonstraci algoritmů byl balík funkcí BIOMOD (Biodiversity Modeling) napsaný v jazyce R a implementovaný v programu Rstudio.

Pro tvorbu veškerých diagramů a ukázkových postupu byla použita webová aplikace LucidChart.

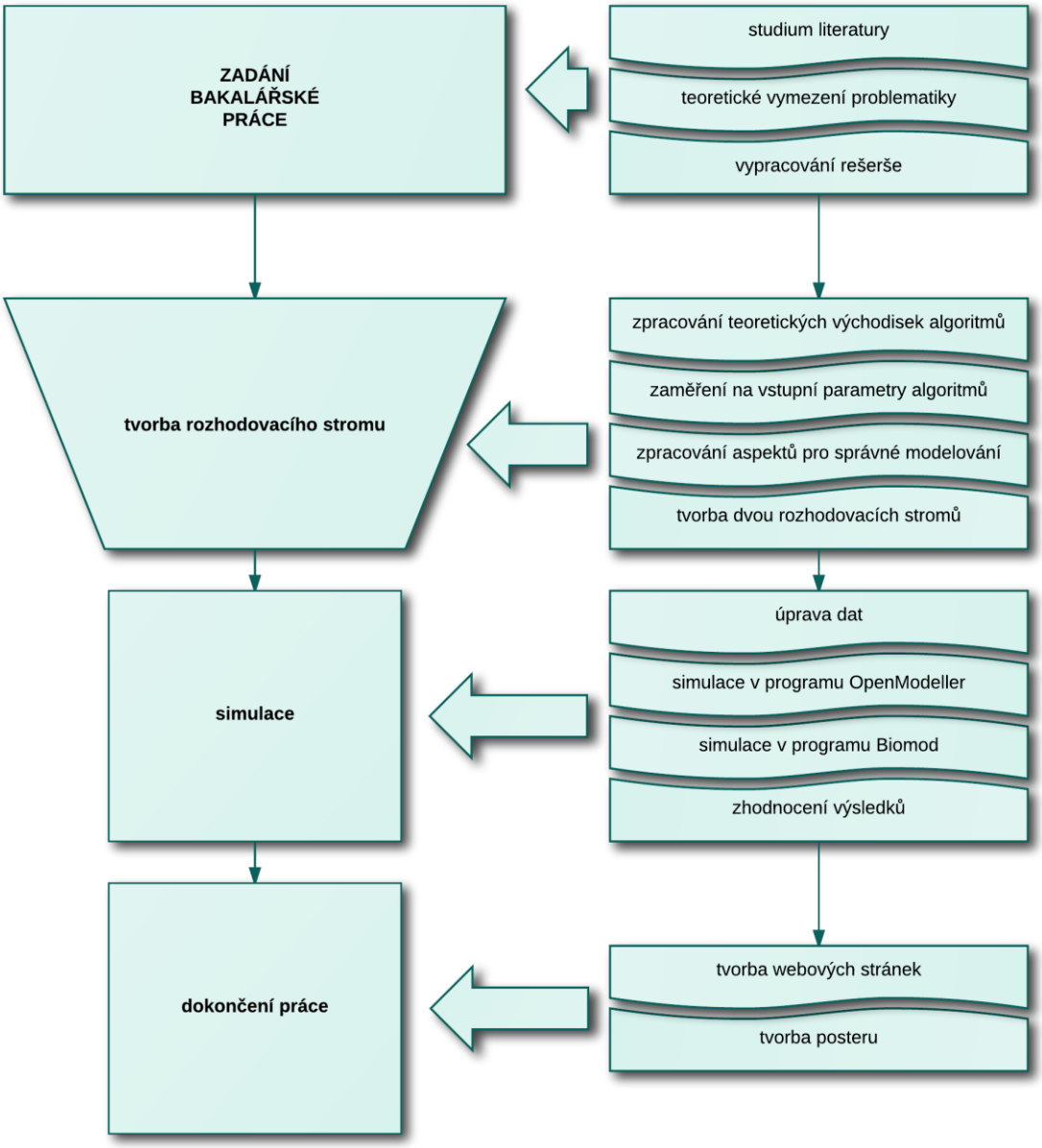
Program PSPad Editor byl použit pro tvorbu webových stránek a pro tvorbu posteru byly využity programy Inkscape 0.91 a GIMP 2.

Postup zpracování

Práce byla sestavena v krocích jako na obrázku níže (Obr. 2.1). Ve stanoveném postupu jsou zohledněny všechny dílčí části práce. V teoretické části práce byla po studiu literatury vypracována rešerše s teoretickým vymezením problematiky. Tvorba rozhodovacího stromu vycházela ze zpracování teoretických východisek algoritmů se zaměřením na jejich vstupní parametry. Dále byly zpracovány aspekty modelování prostorovoé distribuce druhů, je jejichž základu byly vypracovány dva rozhodovací stromy.

V praktické části byly provedeny simulace čtyř druhů motýlů s cílem demonstrovat jednotlivé algoritmy v praxi. Prvním krokem simulací byla úprava dat tak, aby byl možný jejich import do programů OpenModeller a Biomod. Po provedení simulací v obou výše uvedených programech byly výsledky zhodnoceny.

V rámci dokončení práce byl vytvořen poster a webové stránky bakalářské práce.



Obr. 2.1 Postup zpracování bakalářské práce

3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Prostorová distribuce druhů (species distribution modeling – SDM) je ústředním principem řešení mnoha problémů v biogeografii, evoluční ekologii, rozšiřování populací a biologie ochraně životního prostředí (Drake, 2014). V literatuře je často použito také její synonymum, a to modelování ekologických nik (ecological niche modeling), přičemž ekologická nika je soubor faktorů životního prostředí, v němž jsou druhy schopny přežít a ideálně se i rozmnožovat. Tyto faktory jsou zpravidla abiotické (neživé), např. teplota, nadmořská výška, salinita, vlhkost, srážky, vegetační pokrytí, apod., a určují tzv. fundamentální, neboli základní niku. Spolu s biotickými faktory, jako je například mezidruhová konkurence a kompetenční vyloučení, určují realizovanou niku. Podle Petersona (2011) je fundamentální nika rozsah teoretických možností a realizovaná nika je ta část, která je opravdu druhem obsazena, včetně konkurencí jiného druhu který obývá sousední oblast.

Jestliže bude vzat v potaz n maximální možný počet podmínek a zdrojů ekologické niky, lze o ní uvažovat jako o n-rozměrném nadprostoru (Peterson a Vargas, 1993 v Peterson a Vieglais, 2001), což je také základním stavebním kamenem celého ekologického myšlení.

Potenciální distribuce může být vyjádřena geograficky jako realizovaná nika v určitém čase, kdy musí dojít ke splnění biotických a abiotických podmínek organismu. Je nutné si uvědomit, že reálná distribuce druhu často nekoresponduje s modelem potenciální distribuce druhu (Muñoz, 2009). Stabilní populace mohou být nalezeny pouze v takových regionech, které byly k dispozici pro daný druh již od jejich vzniku (pomocí přírodních, antropogenetických či jiných důvodů šíření).

Podle Dormana (2012) mohout být modely prostorové distribuce druhů klasifikovány do dvou skupin, a to na: (1) korelativní modely a (2) mechanistické modely (Obr. 3.2). Korelativní modely spojují nálezová data druhu s prostorovými environmentálními vrstvami ze studované oblasti a produkují mapu pravděpodobnosti výskytu či relativní vhodnosti prostředí pro výskyt druhu (Kumar a kol., 2014). Mechanistické modely používají funkční vlastnosti druhu a psychologickou toleranci pro vhodné nastavení modelu (Kearney a kol., 2010). Do korelativních modelů mohou být dosazena existující nálezová data z muzeí či herbářů (Kearney a kol., 2010; Elith a kol., 2006), zatímco mechanistické modely potřebují detailní experimentální data, která pro zkoumaný druh nemusí být vždy dostupná (Dormann a kol., 2012). Protože mechanistické modely vyžadují expertní posudky ohledně fyziologie druhu a znalosti nároků druhů na podmínky prostředí, jejich použití je stále limitováno, stejně tak jako jejich interpretace.

V případě korelativních modelů jsou požadovány tři typy vstupních dat, a to: nálezová data (*occurrence points*), environmentální data (*environmental layers*) a specifické parametry algoritmu. Nálezová, či také výskytová data, jsou záznamy o tom, kde byli sbíráni či pozorováni jedinci určitého druhu. Právě výběr vhodného algoritmu pro modelování je určován počtem a druhem nálezových dat, dostupností absenčních dat, typem a počtem environmentálních proměnných a také účelem experimentu.

Nálezová data lze rozdělit do následujících čtyř kategorií:

- Prezenční data, což jsou záznamy o potvrzeném výskytem druhu (*presence only data*), jež jsou požadovány klusterovými algoritmy (např. algoritmus Bioclim). Jsou reprezentována číslem 1.

- Absenční data (*absence data*) jsou pravým opakem prezenčních dat, neboť potvrzují negativní výskyt druhu; jsou tedy reprezentována číslem 0. Spolu s prezenčními daty jsou požadována jako vstupy u klasifikačních algoritmů, nicméně pokud nejsou absenční data k dispozici (v ČR jsou v celorepublikovém měřítku těžko získatelná (Brych, 2009)), klasifikační algoritmy si mohou samy vytvořit tzv. pseudoabsenční body (Muñoz a kol., 2009).
- Pseudoabsenční body (*pseudoabsence data*) jsou pseudonegativní nálezy, náhodně vygenerované napříč celou studovanou oblastí. Toto často vede k celkovému snížení přesnosti modelu, obzvláště když jsou body generovány bez jakékoliv znalosti rozmístění druhu (Muñoz a kol. 2009).
- Pozaďová data (*background data*) je integrovaná vrstva environmentálních faktorů, jež byla dopředu vytvořena uživatelem sjednocením environmentálních vrstev.

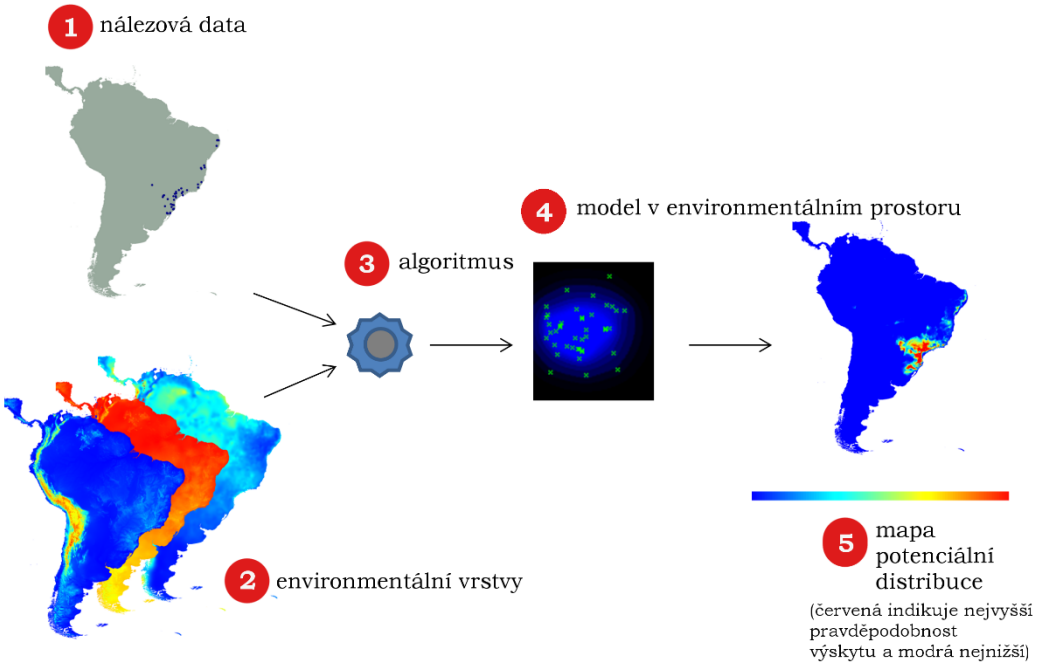
Nálezová data se typicky skládají z unikátního identifikačního čísla, taxonomické identifikace, lokace (zeměpisná šířka, zeměpisná délka a ideálně také datum a přidružený kód), hojnosti výskytu a příslušného data. Příkladem sbírky výskytových dat je projekt „speciesLink“ (2001–2005), jehož cílem bylo integrovat data o druzích a exemplářích z národních historických muzeí, herbářů a národních sbírek a udělat tato data otevřená a volně dostupná na Internetu. V rámci projektu byl vyvinut také první prototyp multiplatformního software Open Modeller (Sutton a kol., 2007). Podle Muñoze (2009) obsahují v současnosti veškeré biologické sbírky cca 2,5 miliardy záznamů sbíraných v průběhu posledních tří set let.

Environmentální vrstvy se nejčastěji vyskytují ve formě georeferencovaných rastrů reprezentujících abiotické faktory na zkoumaném území. Geoprostorová rastrová data jsou vymezena rozsahem pokrytí (souřadnice v rozích rastru) a přidružena referenčním systémem a maticí buněk obsahující aktuální data pro danou oblast.

Předpovědní modely jsou vyvíjeny skrz tříkroký proces:

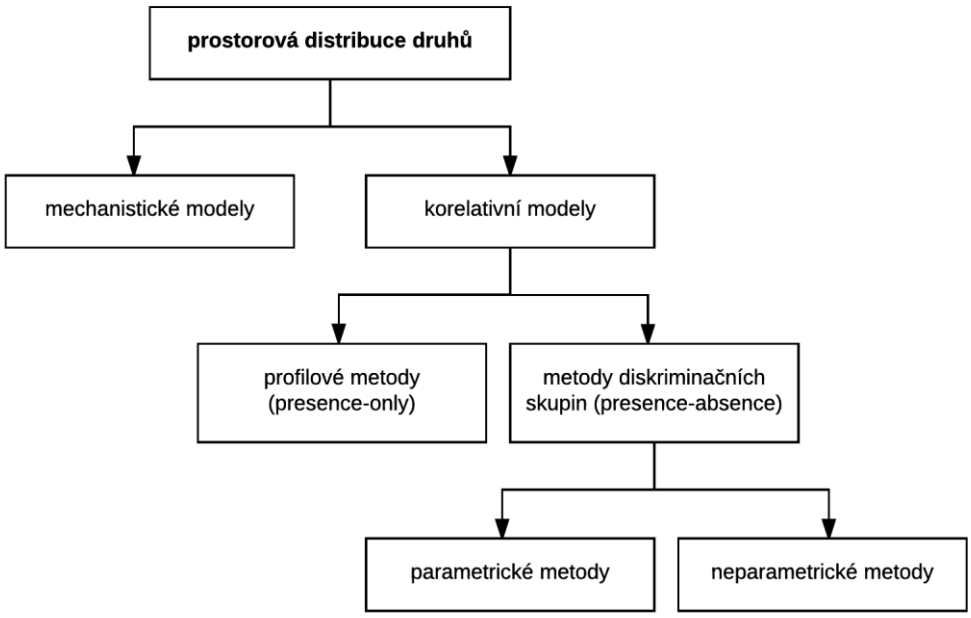
- modelování nik v ekologickém prostoru,
- vyhodnocení těchto modelů nik založené na nativní distribuci,
- projekce modelu zpět do cílené geografické oblasti (Peterson a Vieglais, 2001).

Algoritmy založené na korelativním přístupu generují modely potenciální distribuce (Potential Distribution Model – PD-Model). K vytvoření PD-Modelu nejprve načtou pro každý výskytový bod příslušné environmentální hodnoty ze vstupních rastrů a následně tyto body transformují do struktur zvaných vzorky (samples) jejichž elementy reprezentují environmentální podmínky v každém čtverci. Poté je tento soubor vzorků použit k nalezení reprezentace druhové niky vybraným algoritmem. Výsledný PD-Model může být reprezentován buď datovým modelem pravděpodobnosti, nebo matematickou funkcí udávající vztah mezi environmentálními podmínkami a vhodností prostředí pro existenci druhu. Při projekci modelu zpět do cílené geografické oblasti jsou environmentální podmínky iterativně načítány ze sady rastrů pro každou pozici buňky asociované s cílovou oblastí. Tato sada rastrů může být stejná jako sada použitá ke generování PD-Modelu (nativní projekce), či může být použit soubor rastrů z jiného časového období nebo jiné geografické oblasti. Algoritmus vrací hodnoty predikce korespondující s vhodností prostředí pro výskyt daného druhu, přičemž každá z těchto hodnot je poté zapsána do adekvátní buňky v georeferencované výstupní mapě. Finálním výsledkem je mapa potenciální distribuce (Potential Distribution Map – PD-Map), která reprezentuje potenciální distribuci druhu v určité oblasti a v určitém čase (Muñoz, 2009). Proces modelování je znázorněn na obrázku níže (Obr. 3.1).



Obr. 3.1 Korelativní přístup modelování prostorové distribuce druhů (převzato z: http://openmodeller.sourceforge.net/overview.html)

Korelativní modely lze dále dělit na profilové modely (*profile models*) a modely diskriminačních skupin (*group discriminant models*) (Obr. 3.2) (Esfahani, 2008). Modely využívající pouze prezenční data se řadí do profilových metod; modely využívající prezenční a absenční data jsou řazena do kategorie druhé (a mohou být dále rozdělena na parametrické a neparametrické metody).



Obr. 3.2 Klasifikace modelů pro SDM (upraveno dle Esfahani, 2008)

Parametrické (či také globální) metody se snaží vysvětlit vztah mezi závislou proměnnou a nezávislými proměnnými v rámci celého rozsahu dat jednotnou formou – přímkou nebo rovinou. V literatuře je závislá proměnná také často nazývána jako odpovědná proměnná, environmentální proměnná nebo prediktor. Oproti nim, neparametrické metody (lokální) se přizpůsobují konkrétníím oblastem v prostoru a v globálním měřítku se mohou lišit (např. metoda MARS).

V rámci České republiky lze uvést příklad využití prostorového modelování v projektu „Vyhodnocení migrační propustnosti krajiny pro velké savce a návrh ochranných a optimalizačních opatření." a publikaci *Ochrana průchodnosti krajiny pro velké savce* (Anděl, 2010). Jako demonstraci reálného použití SDM lze uvést práci Tomáše Václavíka a kol. (2010).

Během modelovacího procesu mohou potenciálně nastat dva typy chyb: (1) systémové a (2) náhodné. Systémové chyby (commission errors), což jsou chyby způsobené modelem, a představují vynechání ve skutečnosti obydleného území. Oproti tomu náhodné chyby (ommission errors) zahrnují území, na kterém se ve skutečnosti daný druh nevyskytuje. Tento druh chyby zahrnuje celkem dvě komponenty: reálnou náhodnou chybu, kdy jsou zahrnuty kombinace environmentálních podmínek druhu i když ve skutečnosti do jeho ekologické niky nespadají, a zjevnou náhodnou chybu, neboť druh je zde nepřítomen z důvodu prostorové interakce či historických událostí. V tomto smyslu zjevná náhodná chyba představuje skutečné rysy v druhové distribuční ekologii: „ne všechna obyvatelná území jsou skutečně obývána" (Peterson a kol., 1999).

Existuje celá řada metod, jež mohou být použity k modelování potenciální distribuce druhů. Některé z těchto metod byly vyvinuty speciálně za tímto účelem, jako například Bioclim či GARP. Ovšem také mnoho metod (umělá neuronová síť, klasifikační regresní strom, …) bylo primárně vyvíjeno v jiné oblasti výzkumu, nicméně jejich využití se dostatečně osvědčilo i při SDM. Stejně tak opačně mohou být statistické modely (většina je dostupná v software Biomod (Thuiller, 2003)) využity nejen pro predikci potenciální distribuce, ale pro modelování jakýkoliv dvoučlenných dat (gen, molekula, ekosystém) v závislosti na případných vysvětlujících proměnných.

Mnoho algoritmů má svou desktopovou platformu, ovšem taktéž většina z nich je implementována v software OpenModeller a Biomod (Tab. 3.1).

Protože proces modelování je obvykle poměrně složitý a časově náročný a vyžaduje velkou odbornost v řadě nástrojů a software (různé formáty, jiné software), za účelem zjednodušení a vytvoření uživatelsky přívětivějšího prostředí byl vyvinut software OpenModeller (Sutton a kol., 2007). Framework je schopen se vypořádat s různými projekcemi, souřadnicovými systémy a formáty, takže použitím tohoto rámce se budou uživatelé schopni více soustředit na analýzu než na samotnou přípravu dat. Software Biomod (Thuiller, 2003) je zřejmě tím nejaktuálnějším dostupným software pro SDM a je obsažen ve formě balíku funkcí napsaných v jazyce R.

Tab. 3.1 Algoritmy dostupné skrz programy OpenModeller a Biomod

metoda	implementace	počet vstupních parametrů
Artificial Neural Network	Biomod, OpenModeller	B:5; OM:6*
Aqua Maps	OpenModeller	7
Bioclim	OpenModeller	1
Classification Tree Analysis	Biomod	4
Climate Space Model	OpenModeller	
Consensus	OpenModeller	7
ENFA	OpenModeller	7
Envelope Score	OpenModeller	0
Environmental Distance	OpenModeller	3
Flexible Discriminant Analysis	Biomod	1
GARP (single run)	OpenModeller	4
GARP (with best subsets)	OpenModeller	11
Generalized Additive Model	Biomod	6
Generalized Boosted Model	Biomod	11
Generalized Linear Model	Biomod	4
Maximum Entropy	Biomod, OpenModeller	B:18; OM:14*
Multivariate Adaptive Regression Splines	Biomod	6
Niche Mosaic	OpenModeller	1
Random Forest	OpenModeller	3
Support Vector Machines	OpenModeller	9
Surface Range Envelope	Biomod	4
Virtual Niche Generator	OpenModeller	4

* B = Biomod, OM = Open Modeller

4 TEORETICKÁ VÝCHODISKA ALGORITMŮ

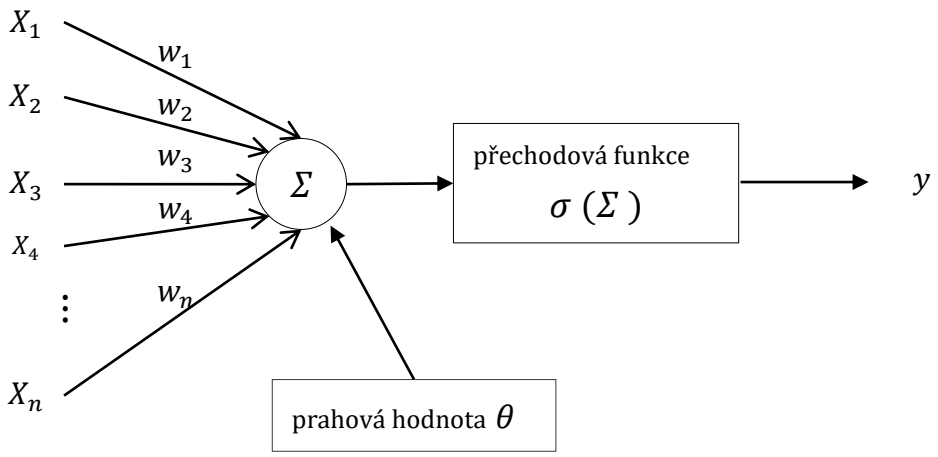
4.1 Artificial Neural Network (ANN)

Tato metoda je implementována jak v software Biomod (Venables a Ripley, 2002), tak i v software OpenModeller, kde je dostupná v aktuální verzi 0.2 (openModeller: Documentation, 2015).

Umělá neuronová síť (Arctifical Neural Network – ANN), také známá jako simulovaná neuronová síť (Simulated Neural Network – SNN) či neuronová síť, je skupina umělých neuronů, jenž jsou vzájemně propojeny; čili prakticky jde o snahu napodobit činnosti lidského mozku ve smyslu získání poznatků ze sítě pomocí procesu učení. Skupina neuronových sítí využívá matematický nebo výpočetní model pro zpracování informací na základě spojitého přístupu k výpočtu. Ve většině případů se jedná o adaptivní systém, který mění svou strukturu založenou na vnitřních nebo vnějších informacích které jí proudí. Používá se k modelování vztahu mezi vícerozměrnou vstupní proměnnou x a vícerozměrnou výstupní proměnnou y.

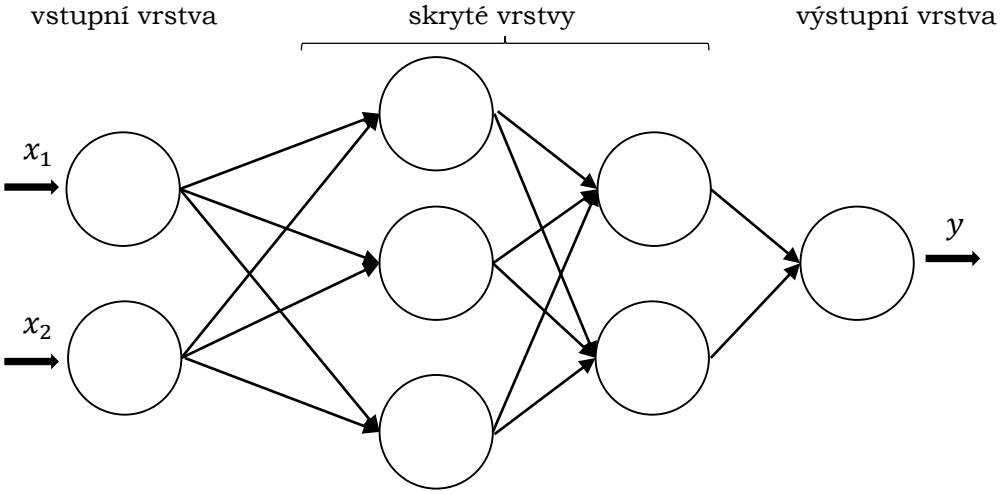
Může být použita pro modelování složitých vztahů mezi vstupy a výstupy či při hledání určitého vzoru v datech bez nutnosti znalosti vztahů mezi proměnnými.

Základní stavební jednotkou je neuron, který má libovolný počet vstupů x_i a jeden výstup y (Obr. 4.1). Citlivost vstupů je určena vahami w_i přiřazenými jednotlivým vstupním proměnným. Každý prvek x_i je vynásoben příslušnou vahou w_i a od výsledné hodnoty je odečten práh θ, který aktivuje výstup neuronu. Následně je přechodovou funkcí σ transformován vnitřní potenciál a dochází k vygenerování výstupu. Mezi nejčastěji používané funkce patří např.: sigmoida se střední hodnotou v bodě $(0;0)$, logistická funkce, prahová funkce, hyperbolický tangens, apod.



Obr. 4.1 Model neuronu (upraveno dle Muller, 1995)

V případě vícevrstvé neuronové sítě dochází k rozlišení tří typů vrstev: vstupní, výstupní a skryté (jsou skryty vnějšímu pozorovateli). Na obrázku (Obr. 4.2) lze vidět příklad dopředné čtyřvrstvé neuronové sítě s topologií 2-3-2-1.



Obr. 4.2 Vícevrstvá neuronová síť s topologií 2-3-2-1

Oproti neuronům sousedních vrstev s úplným propojením, mezi neurony jedné vrstvy propojení neexistuje. Hloubka sítě je obecně udávána jako počet transformací (je rovno počtu skrytých vrstev) včetně vrstvy výstupní.

Jedním ze dvou případů učení je učení s učitelem, kdy je znám pro jednotlivé tréninkové vstupy požadovaný výstup a na základě rozdílu (chyby) mezi požadovaným a skutečným výstupem algoritmus provádí korekci vah, čímž dochází k adaptaci sítě. Druhý případ – učení bez učitele – nemá žádné vnější kritérium správnosti a dochází pouze k hledání vzorků se společnými vlastnostmi.

Množina všech dostupných dat bývá často rozdělena na trénovací a testovací množinu. Trénovací množina ovlivňuje rychlost a kvalitu učení a testovací množina ověřuje výkonnost neuronové sítě po ukončení adaptace. Pro výsledné posouzení kvality predikčních schopností neuronové sítě slouží křížová validace, pro kterou je nutno vymezit validační množinu.

Tématu prostorové distribuce pomocí neuronových sítí je věnována kniha *Neural Networks for Hydrological Modelling* (Abrahart a kol., 2004) a konkrétní použití neuronové sítě pro prostorovou distribuci druhů demonstroval např. Friedrich Recknagel (Recknagel, 1997), který predikoval abundanci a sukcesi modrozelených řas v jezeře Kasumigaura v Japonsku. Srovnáním neuronových sítí s klasifikačními stromy, logistickou regresní analýzou a lineární diskriminační analýzou se ve své práci zabývali Olden a Jackson (Olden a Jackson, 2002), kde byla modelována přítomnost a nepřítomnost celkem 27 druhů ryb v závislosti na stanovištních podmínkách ve 286 jezerech v jižním Ontariu a v Kanadě. Zatímco všechny testované přístupy vykazovaly velmi podobné výsledky při testování na simulovaných lineárních datech, v případě nelineárních dat klasifikační stromy a neuronové sítě výrazně překonaly ostatní tradiční přístupy. Srovnání podobných metod bylo již dříve uvedeno v práci Stephanie Manel z názvem *Alternative methods for predicting species distribution: an illustration with Himalayan river birds* (Manel a kol., 1999).

Parametry (OpenModeller)

- počet neuronů ve skryté vrstvě (number of neurons in the hidden layer)
  - rozsah hodnot: $\{1, \cdots, \infty\}$
  - typ: přirozená čísla
  - výchozí hodnota: 14
  - Je přídavnou vrstvou ke vstupním a výstupním vrstvám a určuje, kolik neuronů mají obsahovat jednotlivé vrstvy.
- rychlost učení (learning rate)
  - rozsah hodnot: $< 0; 1 >$
  - typ: reálná čísla
  - výchozí hodnota: 0,3
  - Trénovací parametr kontrolující velikost váhy a ovlivňující samotnou adaptaci algoritmu.
- hybnost (momentum)
  - rozsah hodnot: $< 0; 1 >$
  - typ: reálná čísla
  - výchozí hodnota: 0,05
  - Hybnost je koeficient, který zabraňuje systému konvergovat k lokálnímu minimu či sedlovému bodu tím, že k aktuálnímu gradientu přičte navíc zlomek gradientu minulého kroku (Daubner, 2015). Vysoký parametr hybnosti může pomoci ke zvýšení rychlosti konvergence systému. Nicméně nastavení příliš vysoké hodnoty vytváří riziko přestřelení minima, což může způsobit nestabilitu systému. Oproti tomu příliš nízká hodnota parametru se nemůže dostatečně vyvarovat lokálním minimům a také může systém zpomalit.
- trénovací typ (training type)
  - rozsah hodnot: {0,1}
  - typ: přirozená čísla
  - výchozí hodnota: 1
  - 0 = trénováno epochou, 1 = trénováno minimální chybou
- epocha učení (epoch)
  - rozsah hodnot: $\{1, \cdots, \infty\}$
  - typ: přirozená čísla
  - výchozí hodnota: 5 000 000
  - Interval, ve kterém dojde k předložení všech vzorů trénovací množiny alespoň jednou. Určuje, kdy bude zastaveno trénování, jakmile počet iterací přesáhne epochu. Při trénování minimální chybou reprezentuje maximální počet iterací.
- minimální chyba (minimum error)
  - rozsah hodnot: $< 0; 0{,}5 >$
  - typ: reálná čísla
  - výchozí hodnota: 0,01
  - Udává minimální střední kvadratickou chybu epochy. Druhá odmocnina součtu kvadrátů rozdílů mezi požadovanými a skutečnými výstupy je vydělena počtem vzorů (pouze pro trénování minimální chybou).

Parametry (Biomod)

- data pro učení („*NbCV*“)
  - výchozí hodnota: 5
  - Počet křížových validací k nalezení optimální velikosti.
- počet neuronů ve skryté vrstvě („*size*“)
- decay („*decay*“)
  - výchozí hodnota: NULL
  - Penalizace příliš velkých vah pomocí jednoduché metody „weight decay“. Metoda se snaží postupně normalizovat váhy hran do uzších intervalů kolem počátku souřadnic (Civín, 2006).
- rozsah váhových hodnot („*rang*“)
  - výchozí hodnota: 0,1
  - Inicializace vah neuronů ze zadaného intervalu.
- maximální počet iterací („*maxit*“)
  - výchozí hodnota: 200

4.2 Aqua Maps

Aqua maps je modelovací nástroj přizpůsobený k modelování distribuce mořských organismů, jenž byl vyvinut jako součást projektu Incofish (Kesner-Reyes a kol., 2012). Základním principem je modelový přístup založený na zóně životních nároků, kde má každá proměnná asociovaný preferovaný rozsah a širší možný rozsah. Bez preferovaného rozsahu je pravděpodobnost výskytu 1. Mezi preferovaným a možným rozsahem pravděpodobnosti výskytu proměnná kolísá mezi 0 a 1 (lineární rozklad) a mimo možný rozsah je pravděpodobnost výskytu rovna 0.

Celková pravděpodobnost výskytu je vypočítána vynásobením všech jednotlivých pravděpodobností. Tento algoritmus se od ostatních tradičních algoritmů liší tím, že k práci vyžaduje specifickou sadu vrstev v následujícím pořadí: maximální hloubka v metrech, průměrná roční koncentrace ledu, průměrná roční vzdálenost k pevnině v kilometrech, průměrná roční produkce (chlorofyl A, měřený v $mgC/m^2/den$), průměrná roční slanost u dna v jednotkách PSU (Practical Salinity Unit), průměrná roční povrchová slanost v jednotkách PSU, průměrná roční teplota u dna ve stupních Celsia, průměrná roční povrchová teplota ve stupních Celsia (openModeller: Documentation, 2015).

Preferované rozsahy jsou obvykle vypočítány na základě 10. a 90. percentilů a jsou dále upravovány pomocí mezikvartilových hodnot a také zajištěním minimální velikosti zóny životních nároků založené na předem definovaných hodnotách.

Expertní informace, ke kterým se přistupuje skrz algoritmus, pochází z FishBase a jsou uloženy v místní SQLite databázi[1]. K nalezení informace v databázi musí být veškeré výskytové body označeny vědeckým názvem (rod a druh). Výskyty jsou jednoznačné a jsou citlivé na velikost písmen i při operacích. V této verzi (2016) obsahuje interní databáze informace o více než 7 000 mořských druzích.

Predikcí distribuce mořských druhů v globálním měřítku se ve své práci zabýval Jonatan Ready a kol. (2002) či Gianpaolo Coro (2016), který ve své studii uveřejnil potenciální dopad klimatické změny na mořské druhy. Jednalo se o vyhodnocení a porovnání současných a budoucích možných scénářů (rok 2050) distribuce ze 406 zkoumaných mořských druhů s použitím algoritmu Aqua Maps.

- [1] http://www.aquamaps.db

Miranda C. Jones (2012) ve své práci *Modelling commercial fish distributions: Prediction and assessment using different approaches* modeluje výskyt mořských a bezobratlých živočichů použitím metod Aqua Maps, Maxent a Sea Around Us Project.

Parametry:

- použití vrstev z povrchu hladiny (use surface layers)
    - rozsah hodnot: {−1,0,1}
    - typ: přirozená čísla
    - výchozí hodnota: -1
    - pouze pro teplotu a slanost
    - Použití vrstev s povrchovou teplotou vody a salinitou (1 = ano, 0 = ne, -1 = o použití rozhodne algoritmus). Při výchozím nastavení se bude algoritmus snažit najít rozsah hloubky pro druhy v jeho interní databázi. Pokud bude minimální hloubka menší nebo rovna 200 m, pak algoritmus vrstvy s povrchovou teplotou vody a salinitou použije. V opačném případě použije vrstvy s informacemi ze dna. Tento parametr může být využit k vynucení použití těchto dvou typů vrstev algoritmem.
- použít rozsah hloubky (use depth range)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - Je-li k dispozici, při výpočtu pravděpodobnosti výskytu použije rozsah hloubky.
- použít koncentraci ledu (use ice concentration)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - Při výpočtu pravděpodobnosti výskytu využít koncentraci ledu.
- použít vzdálenost k pevnině (use distance to land)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 0
    - Při výpočtu pravděpodobnosti použít zónu tolerance ke vzdálenosti k pevnině.
- primární produkce (use primary production)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - Využít zónu tolerance pro primární produkci.
- použít salinitu (use salinity)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - Využít slanost při výpočtu pravděpodobnosti.
- použít teplotu (use temperature)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - Využít teplotu při výpočtu pravděpodobnosti.

4.3 Bioclim

Bioclim je korelativní modelovací nástroj schopen interpolovat až 35 environmentálních proměnných. Může být použit ke třem účelům:

1. popis životního prostředí, ve kterém byl druh nalezen,
2. identifikace dalších lokací, které by mohl druh potenciálně obývat,
3. určení lokací pro potenciální výskyt druhu vzhledem ke změně klimatu.

Implementuje algoritmus bioklimatické zóny tolerance. Pro každou zadanou proměnnou algoritmus najde průměr a směrodatnou odchylku (za předpokladu normálního rozdělení). Každá proměnná má svou vlastní zónu tolerance reprezentovanou intervalem $[m - c \cdot s, m + c \cdot s]$, kde m je průměr, **c** je vstupní mezní parametr a s je směrodatná odchylka. Každá proměnná má také navíc horní a dolní limit převzatý z minimálních a maximálních hodnot ze vstupní sady. Následně algoritmus porovná hodnoty environmentálních proměnných v jakékoliv lokalitě s percentilovým rozložením hodnot známých míst výskytu („tréninkové body“). Čím blíže je lokace k mediánu, tím je vhodnější. Výsledky percentilu (*percentile scores*) jsou mezi 0 a 1 (pokud jsou predikované hodnoty větší než 0,5, jsou odečteny od čísla 1). Poté je spočítáno minimální procentuální skóre napříč všemi environmentálními proměnnými, konečná hodnota se odečte od 1 a vynásobí 2, čímž se dosáhne hodnot výsledků v intervalu $\langle 0; 1 \rangle$. Hodnota 1 se bude vyskytovat velmi zřídka, neboť k jejímu dosažení by musela mít lokace hodnotu mediánu trénovacích dat pro všechny uvažované proměnné. Oproti tomu hodnota 0 bude velmi častá, protože je přiřazena ke všem buňkám s hodnotou environmentální proměnné mimo percentilový rozsah pro alespoň jednu veličinu (openModeller: Documentation, 2015)

V predikovací funkci lze ignorovat jeden z okrajů distribuce (např. limitujícím faktorem bude málo srážek, ale mnoho srážek ne).

V tomto modelu může být každý bod klasifikován jako:

1. Vhodný: pokud všechny přidružené hodnoty spadají do vypočítaných zón tolerance. Výstupem je hodnota 1.
2. Okrajový: pokud jedna nebo více přidružených hodnot nespadá do vypočítaných zón tolerance, přesto stále spadá do intervalu mezi horním a dolním limitem. Výstupem je hodnota 0,5.
3. Nevhodný: pokud jedna nebo více hodnot spadá mimo interval mezi horním a dolním limitem. Výstupem je hodnota 0.

Příkladů využití lze uvést značné množství. Zajímavou prací je studie Morgane Barbet-Massin a kol. (2014). Autoři na celkem 243 druzích ptáků a 6 relativně nekorelovaných bioklimatických proměnných spustili 6 algoritmů (GAM, GLM, FDA, ANN, BRT a RF) s cílem poukázat na důležitost výběru relevantních environmentálních proměnných. Na důsledky klimatické změny se zaměřuje práce Carvalha a kol. (Carvalho a kol., 2015), kde je zkoumána potenciální expanze ničivek (*Leishmania*), jež způsobují smrtelné onemocnění zvané Leishmanióza a jsou činitely častých epidemií. Pro modelování byly použity algoritmy Bioclim, Maximum Entropy, GARP, Random Forest a DOMAIN.

V důsledku mimořádné vzácnosti aplikací prediktivních modelovacích přístupů pro italskou faunu byla v roce 2009 publikována studie s názvem *Modelling Bedriaga's rock lizard distribution in Sardinia* (Bombi a kol., 2009). Práce se zaměřuje na použití algoritmů Bioclim, GAM, GLM, MAXENT a ENFA s cílem detekovat potřebné klimatické podmínky, předpovědět potenciální distribuci a identifikovat nejzranitelnější populace ještěrky Bedriagovy (*Archaeolacerta bedriagae*). Srovnáním *presence-only* metod (GARP,

Bioclim, ENFA, apod.) se zabývá práce Asafa Tsoara (2007), kde uvádí, že po testování se metody Bioclim a ENFA ukazovaly jako ty s nejnižší prediktivní přesností.

Parametry:

- limit směrodatné odchylky (standard deviation cutoff)
    - rozsah hodnot: $(0; \infty)$
    - typ: reálná čísla
    - výchozí hodnota: 0,674
    - Procentuální limit pro začlenění do bioklimatické zóny tolerance.
    - Příklady („procento začlenění“, „hodnota parametru“): (50.0%, 0.674); (68.3%, 1.000); (90.0%, 1.645); (95.0%, 1.960); (99.7%, 3.000)

4.4 Classification Tree Analysis (CTA)

Metoda založena na identifikaci specifického prahu pro každou environmentální proměnnou poskytuje dobrou alternativu k regresním technikám. Stejně jako GAM, nespoléhají na apriorní hypotézy o vztahu závislých a nezávislých proměnných. Data jsou opakovaně rozdělena do homogenních skupin, které nejlépe vysvětlují výskyt či absenci druhu, až dojde k vytvoření stromu klasifikačních pravidel. Heterogenita uzlu může být interpretována jako odchylka Gaussova modelu (pro regresní strom) nebo multinomiálního modelu (pro klasifikační strom).

Použití klasifikačních a regresních stromů bylo publikováno v článku *Classification and regression trees: a powerfull yet simple technique for ecological data analysis* (De'ath, 2000). Potenciální výhody používání klasifikačních a regresních stromů lze nalézt v práci Marca P. Vayssierese a kol. (2000).

Parametry:

- metoda („*method*“)
    - rozsah hodnot: ‚anova', ‚poisson', ‚class', ‚exp'
    - výchozí hodnota: ‚class'
- parms („*parms*“)
    - výchozí hodnota: 1
    - Nepovinný parametr pro rozdělení funkce. Pokud byla zvolena metoda ANOVA, zůstává tento parametr prázdný. Poissonovo rozdělení má jediný parametr, a to variační koeficient podle dosavadní distribuce.
- cost („*cost*“)
    - výchozí hodnota: NULL
- rozsah váhových hodnot („*control*“)
    - výchozí hodnota: 0,1
- maximální počet iterací („*maxit*“)
    - výchozí hodnota: 200

4.5 Climate Space Model

Jedná se o metodu založenou na principu faktorové analýzy a spolu s metodou ENFA, která je založena na principu analýzy hlavních komponent, patří mezi metody redukce počtu původních proměnných. Faktorová analýza objasňuje kovarianci a korelaci původních proměnných pomocí několika společných komponent (latentních proměnných). K nevýhodám patří nutnost zadat počet společných faktorů ještě před provedením samotné analýzy.

Metody založené na PCA byly například testovány v práci Robertsona a kol. (2001), kde byla na bioklimatických datech předpovídána distribuce tří invazivních rostlin v JAR (Jihoafrická republika), Lesothu a Svazijsku.

Parametry:

- počet náhodných vlastních čísel (number of random eigenvalues)
  - rozsah hodnot: $\{1, \cdots, 1000\}$
  - typ: přirozená čísla
  - výchozí hodnota: 8
  - Metoda výběru hlavních komponent založena na vlastních hodnotách (eigenvalues) získaných z náhodných dat.
- počet standardních odchylek (number of standard deviations)
  - rozsah hodnot: $\langle -10; 10 \rangle$
  - typ: reálná čísla
  - výchozí hodnota: 2
  - Ve chvíli, kdy jsou sečteny proměnné (suma stadardizovaných rozptylů), je toto číslo přidáno k průměru těchto proměnných. Jsou zachovány ty komponenty, které jsou nad touto hranicí.
- minimální počet komponent (minimum number of components in the model)
  - rozsah hodnot: $\{1, \cdots, 20\}$
  - typ: přirozená čísla
  - výchozí hodnota: 1
  - Obvykle je přijatelných 3 a více komponent. Je-li vybrán nedostatečný počet komponent, model selže či vrátí chybný výsledek.
- zobrazit detailní průběh procesu (show very detailed debugging info)
  - rozsah hodnot: {0,1}
  - typ: přirozená čísla
  - výchozí hodnota: 0

4.6 Consensus

Jedná se o takový druh algoritmu, který ve svých parametrech přijímá další algoritmy. Může tedy generovat jak jednotlivé modely, tak výsledky agregovat do modelu jednoho.

Maximální počet algoritmů je omezen na 5.

Ke specifikaci algoritmu je třeba znát jeho ID (Identification) a také názvy jeho parametrů v programu openModeller.

Před sloučením modelů jsou všechny jednotlivě transformovány do binárního modelu na základě nejnižší prahové hodnoty. Každému algoritmu mohou být jednak přiřazeny různé váhy a jednak minimální úroveň shody mezi algoritmy. Pokud je například použito 5 algoritmů s minimální úrovní 3, bude výsledek nulový ve chvíli, kdy

se na predikci shodnou méně než 3 algoritmy. Konečný model bude tedy zobrazovat pouze ty oblasti, kdy se zadaný počet algoritmů shodne na predikci výskytu (openModeller: Documentation, 2015)

Parametry:

Musí být uvedeno ID algoritmu a následně jméno parametru a jeho hodnota, odděleno čárkou a uzavřeno v závorkách.

- algoritmus 1
  - typ: řetězec hodnot
  - např.: RF(NumTrees=10,VarsPerTree=0,ForceUnsupervisedLearning=0)
  - Stejný postup následuje i u dalších čtyř algoritmů. Pokud stačí použít méně než 5 algoritmů, stačí nechat pole prázdná.
- algoritmus 2
- algoritmus 3
- algoritmus 4
- algoritmus 5
- váhy (weights)
  - typ: řetězec hodnot
  - výchozí hodnota: 1.0 0.0 0.0 0.0 0.0
  - Sekvence vah pro jednotlivé algoritmy. Může být použito k většímu důrazu na určité algoritmy. Váhy jsou odděleny mezerou a jako oddělovač desetinných míst je použita tečka.
- shoda (agreement)
  - rozsah hodnot: {1,2,3,4,5}
  - typ: přirozená čísla
  - výchozí hodnota: 1
  - Udává minimální úroveň shody mezi algoritmy. Jako pozitivní hodnoty s predikcí výskytu budou vráceny pouze ty hodnoty, které se shodují v zadaném počtu algoritmů.

4.7 Ecological Niche Factor Analysis (ENFA)

Tento algoritmus používá upravenou analýzu hlavních komponent (Principal Component Analysis – PCA) k modelu založenému pouze na prezenčních datech. Pozorované prostředí je porovnáno s pozaďovými daty studovaného území (absenční body jsou v souboru s výskytem klasifikována jako background data).

PCA je multivariační technika vytvářející soubor latentních proměnných (tzv. hlavních komponent), které jsou váženou lineární kombinací původních proměnných (James a McCulloh, 1990 v Roberston a kol., 2001). Cílem je objasnit rozptyl původních proměnných a snížit dimenzi prostoru. PCA je provedena na tréninkovém datasetu, kde zkonstruuje matematický hyperprostor, ve kterém každá osa definuje jednu hlavní komponentu. Pokud jsou všechny hodnoty prediktorů zobrazeny v hyperprostoru definovaného trénovací množinou, pak lze vypočítat vzdálenost od každého nevzorkovaného místa do původního hyperprostoru (Roberston a kol., 2001). Tato vzdálenost je následně použita pro výpočet bioklimatické zóny tolerance.

První faktor se nazývá “marginalita” druhů (tedy okrajovost) a je definována jako ekologická vzdálenost mezi optimem a středním stanovistěm bez pozaďových dat. Marginalita vysvětluje jak moc se ekologické optimum druhu odchyluje od nejfrekventovanějších podmínek v území. Ostatní faktory jsou nazvány “specializace” a jsou definovány jako poměr ekologického rozptylu ve středním stanovišti k

pozorovanému cíli druhů (Hirzel a Arlettaz, 2003) a udávají toleranci k suboptimálním podmínkám.

Jelikož je ENFA jednou z metod, která nepožaduje absenční data, byla jí věnována studie s názorným příkladem na Kozorožci alpském (*Capra ibex ibex*) (Hirzel a kol., 2002). Tento druh byl reintrodukován ve Švýcarsku a cílem studie bylo předpovědět jeho pravděpodobnou distribuci, neboť po navrácení druh ihned nekolonizoval celou švýcarskou oblast. V práci Huga Rebela a Jonese Garetha (2010) jsou srovnávány dvě *presence-only* metody. Použitím algoritmů ENFA a Maxent je modelována potenciální distribuce jednoho z nejvzácnějších netopýrů *Barbastella barbastellus*. Protože algoritmus ENFA přijímá pouze spojité environmentální proměnné, byl první model Maxent spuštěn za použití stejných proměnných a do druhého modelu Maxentu byla přidána jedna kategorická proměnná – landcover (využití krajiny). Ve finále algoritmus ENFA předpověděl oblast výskytu v širším spektru než oba modely Maxent a v jižní oblasti území došlo k podstatným neshodám ohledně nejvhodnějších environmentálních podmínek pro výskyt druhu. Otázkou porovnání a hodnocení modelovacích algoritmů ENFA, Maxent a GLM včetně výběru nejdůležitějších environmentálních proměnných se zabývá práce s názvem *Modelling potential distribution of the threatened tree species Juniperus oxycedrus* (Ruprecht a kol., 2011). Pokusem vysvětlit faktory zodpovědné za demografické a genetické vyčerpání populací obojživelníků se ve své práci zabýval Dolgener a kol. (2013). S využitím dat ohrožené Kuňky obecné (*Bombina Bombina*) a environmentálními proměnnými (teplota, srážky, půdní vlhkost, hustota vegetace a dopad silničního provozu) byly zjištěny významné korelace mezi silniční disturbancí a pozorovanou interdruhovou diverzitou. Caruso a kol. (2015) modeloval distribuci ohrožených populací pum v Argentině. Metodou ENFA byly zjistěny nejvhodnější lokality pro výskyt pumy, a to vzdálená místa od zemědělské půdy, městských oblastí a také hlavních silnic. Na zúžení ekologické niky byl také potvrzen vliv mezi vzdáleností k silnici a křovinatému porostu. ENFA se ukázala jako vhodný modelovací nástroj pro řízení potenciální distribuce plevelných rostlin na ostrovních systémech bez ohledu na velikost datového souboru (Costa a kol., 2013). Využití této metody je nutno zmínit také v práci Mireiy Vally (Valle a kol., 2011), kde byla pro modelování výskytu mořských řas v estuáriích ve Španělsku jako environmentální proměnné použita data získána ze systému LiDAR (Light Detection And Ranging).

Parametry:

- počet vzorků pozaďových dat (number of background sample points)
    - rozsah hodnot: $\{10, \cdots, \infty\}$
    - typ: přirozená čísla
    - výchozí hodnota: 10 000
    - Výpočtem střední a standardní odchylky a kovariance z každé environmentální vrstvy je umožněno srovnání prezenčních nálezových dat s prostředím. Hodnoty jsou odhadnuty vzorkováním z n počtu bodů z pozaďových dat z důvodu potenciální náročnosti na data velkého objemu.

- počet opakování (number of retries of model)
    - rozsah hodnot: $(1; \infty)$
    - typ: reálná čísla
    - výchozí hodnota: 5
    - Algoritmus invertuje matice, ale v případě inverze singulární matice selhává. Toto se stává, když je vzorek pozaďových dat nereprezentativní nebo podvzorkovaný. Řešením problému je opakování generování modelu (čili převzorkování pozaďových dat).
- metoda vyřazení komponent (method for discarding components)
    - rozsah hodnot: {0,1,2}
    - typ: přirozená čísla
    - výchozí hodnota: 2
    - 0 – zachovat pevně stanovený počet komponent definovaných v proměnné ‚RETAIN_COMPONENTS' (parametr "počet zachovaných komponent")
    - 1 – zachovat prvních n komponent, které kumulativně vysvětlují míru variability definované v proměnné ‚RETAIN_VARIATION'
    - 2 – Porovnání pozorovaného objasnění variace distribuce metodou zlomené hůlky (Matematická biologie, 2015) udržující ty komponenty, které objasňují vyšší úroveň variability.
- počet zachovaných komponent (number o components to retain)
    - rozsah hodnot: $\{1, \cdots, \infty\}$
    - typ: přirozená čísla
    - výchozí hodnota: 2
    - Pokud je metoda vyřazení komponent = 0, pak je tato proměnná použita k určení počtu komponent určených k uchování.
- procento variace pro uchování komponent (percent variation for component retention)
    - rozsah hodnot: $< 0{,}5; 1 >$
    - typ: reálná čísla
    - výchozí hodnota: 0,75
    - Pokud je metoda vyřazení komponent = 1, pak je tato proměnná použita k určení počtu komponent určených k uchování zahrnutím těch komponent, které kumulativně představují nejméně této variace.
- zobrazit nahrávání proměnných pro každý faktor (display variable loadings for each factor)
    - rozsah hodnot: $\{0; 1\}$
    - typ: přirozená čísla
    - výchozí hodnota: 0
    - nastavte na 1 pro zobrazení matice.
    - Var = proměnná
    - Mar = marginalita (faktor 0)
    - Sp-1 = specializace faktoru 1
    - Proměnné jsou číslovány ve stejném pořadí jako v požadavku.
- zobrazit průběh procesu (verbose printing for debugging)
    - rozsah hodnot: $\{0; 1\}$
    - typ: přirozená čísla
    - výchozí hodnota: 0

4.8 Envelope Score

Implementace laxního algoritmus bioklimatické zóny tolerance. Pro každou zadanou environmentální proměnnou x_i najde algoritmus minimum $x_{i_{min}}$ a maximum $x_{i_{max}}$ na všech místech výskytu. V průběhu modelové projekce je pravděpodobnost výskytu definována jako $p = \Sigma l / \Sigma m ; l \in \langle x_{i_{min}}; x_{i_{max}} \rangle$, kde l a m je environmentální vrstva. Při výběru hodnoty prahu $\theta = 1$ bude výstup stejný jako při použití algoritmu BIOCLIM. Tento algoritmus je bez parametrů.

4.9 Environmental Distance

Genetický algoritmus založen na metrikách environmentálních odlišností. Při použití Gowerovy metriky a maximální vzdálenosti = 1 bude výstup z algoritmu stejný jako z algoritmu DOMAIN (Carpenter a kol., 1993).

Parametry:

- metrika (metric)
  - rozsah hodnot: {1,2,3,4}
  - typ: přirozená čísla
  - výchozí hodnota: 1
  - 1 = euklidovská vzdálenost
  - 2 = mahalanobisova vzdálenost
  - 3 = manhattan / gowerova vzdálenost
  - 4 = chebyshevova vzdálenost
- počet nejbližších n bodů (nearest n points)
  - rozsah hodnot: $\{0, \cdots, 1000\}$
  - typ: přirozená čísla
  - výchozí hodnota: 1
  - n počet nejbližších bodů, jejichž průměrná hodnota bude použita jako reference při výpočtu environmentální vzdálenosti.
  - Pokud je proměnná nastavena na hodnotu 1, vzdálenosti budou měřeny k nejbližšímu bodu, pokud je proměnná nastavena na hodnotu 0, budou vzdálenosti měřeny jako průměr všech. Parametr přijímá hodnoty v intervalu $\langle 1; n \rangle$; kde n je celkový počet bodů.
- maximální vzdálenost (maximum distance)
  - rozsah hodnot: $< 0; 1 >$
  - typ: reálná čísla
  - výchozí hodnota: 0,1
  - Udává maximální vzdálenost k referenci v oblasti environmentálního prostoru, nad kterou budou podmínky považovány jako nevhodné pro přítomnost druhu. Vzhledem k tomu, že hodnota 1 odpovídá největší možné vzdálenosti mezi jakýmikoliv dvěma body v environmentálním prostoru, nastavení hodnoty maximální vzdálenosti na tuto hodnotu znamená, že všechny body v environmentálním prostoru budou mít přidruženou pravděpodobnost. Pravděpodobnost přítomnosti pro body, které spadají do rozsahu maximální vzdálenosti je nepřímo úměrná vzdálenosti od referenčního bodu (lineární rozpad). Jedinou výjimkou je případ, kdy bude hodnota maximální vzdálenosti nastavena na 1 a pro určení vzdálenosti bude použita mahalanobisova metrika jejíž pravděpodobnost plyne z χ^2 rozdělení.

4.10 Flexible Discriminant Analysis (FDA)

FDA patří mezi metody založené na klasifikaci a jedná se o rozšíření známé lineární diskriminační analýzy (Linear Discriminant Analysis – LDA). Směsice normál je použita k získání hustoty odhadu pro každou třídu. Ke zvýšení prediktivní přesnosti modelů může být v postprocessingu použita metoda ANN nebo MARS.

Studiem FDA a LDA se ve svých pracích dlouhodobě zabývá Trevor Hastie (Hastie a kol., 1994; Hastie a kol., 1995; Hasti a Tibshirani, 1996). Využití metody v biogeografii a ekologii lze nalézt v práci Manel a kol. (1999), kde je ilustrováno názorné použití na vodním ptactu v Himalájích.

Parametry:

- metoda („method“)
    - výchozí hodnota: ‚mars‘
- přídavné argumenty („add_args“)
    - Další argumenty zadané jako seznam parametrů (odpovídají možnostem funkce FDA).

4.11 GARP (single run)

GARP (Genetic Algorithm for Rule-Set Prediction) zahrnuje několik odlišných algoritmů založených na iterativním a uměle inteligenčním přístupu (Stockwell a Noble, 1992). Individuální algoritmy s různorodými prediktivními přístupy (analýza vícenásobné regrese k predikci pravděpodobnosti přítomnosti nebo průnik rozsahu spolu s environmentálními dimenzemi) jsou použity operativně skrz mnoho generací úpravy pravidla, testování, a začlenění nebo nepřijetí do modelu. Pravidlo vhodnosti (prediktivní přesnosti) je testováno porovnáváním sad bodů převzorkovaných ze známých nálezových bodů a z pozaďových dat. Výsledkem genetického algoritmu je sada 5 až 50 různých pravidel, které společně definují dimenze ekologické niky druhu (Peterson a Vieglas, 2001) se schopností poradit si s málo strukturovanými daty nevhodnými pro klasické statistické metody (Sánchez-Flores, 2007).

Nejprve je zvolena jedna z možných variant (logistická regrese, bioklimatické pravidlo, …) a ta je aplikována na výběr trénovacích bodů. Na základě této aplikace je vytvořeno pravidlo, které je otestováno na prezenčních a pseudoabsenčních bodech. Změna v přesnosti predikce mezi iteracemi je pak využita k vyhodnocení zda dané pravidlo zahrnout do modelu či nikoliv (Brych, 2009). Výpočet končí po zadaném počtu iterací nebo když dojde ke konvergenci.

GARP modely byly podrobeny několika testům přesnosti a robustnosti. Počáteční testy posuzující odolnost ke změně v hustotě dat a velikost vzorku výskytových bodů ukazují, že 4–8 sad environmentálních dat a 10–30 výskytových bodů jsou obecně dostatečným počtem pro dosažení maximální přesnosti predikce pro dané druhy (Peterson a Cohoon, 1999).

Program OpenModeller nabízí dvě verze této metody, a to vlastní vylepšenou implementaci algoritmu a původní implementaci algoritmu z programu DesktopGARP. V nové implementaci programu OpenModeller se jednalo o kompletní přepsání kódu Desktop GARP s následujícími změnami (OpenModeller: Documentation, 2015):

1. Hodnoty genu byly změněny z přirozených číslel $\{1, \cdots, 253\}$ na proměnné s plovoucí desetinnou čárkou $\langle -1; 1 \rangle$. Tím bylo zabráněno problémům v environmentálních hodnotách během projekce (například pokud má

nějaká environmentální proměnná hodnotu 2,56 v buňce jednoho rastru a v buňce jiného rastru má hodnotu 2,76, obě tyto hodnoty Desktop GARP zaokrouhlil na 3).

2. Ve srovnání s jinými pravidly byla pro jejich malý význam odstraněna atomická pravidla.
3. Protože během prvních iterací heurestické parametry operátora (procento mutace a křížení v průběhu iterace) konvergovaly k fixním hodnotám, byly změněny na statické.
4. Byla opravena chyba při řazení pravidel. Při nahrazování pravidla jiným pravidlem bylo toto pravidlo zařazeno na nesprávné místo.

Obecně platí, že několik výskytových bodů je vybráno k náhodné stavbě distribučního modelu zatímco zbytek je dán stranou k vyhodnocení modelu. Správným nastavením parametrů se například ve své práci zabýval Anderson a kol. (2003). Modelování distribuce, kde místo přesných lokalit výskytu byla použita data ze starých a současných map znázorňující distribuci druhů se zabývá práce Milana Koreně a kol. (2001). Tento postup byl vytvořen k modelování vhodnosti prostředí Medvěda hnědého (*Ursus Arctos*) na Slovensku. Ve studii Estrada-Contrease a kol. (2015) autoři použitím metody GARP zkoumají pravděpodobné změny ve složení hlavních typů vegetace v Mexiku v důsledku budoucí klimatické změny. V práci Abrahamyana a Barsevskise (2015) je metodou GARP modelována budoucí distribuce Dobromyslu obecného (*Origanum vulgare L.*) v Arménské republice. Autoři uvádí, že se v důsledku zhoršení životního prostředí a změny klimatu v roce 2050 sníží distribuce druhu hlavně v oblasti centrální Arménie.

Parametry:

- maximální počet iterací (max generations)
  - rozsah hodnot: $\{1; \cdots; \infty\}$
  - typ: přirozená čísla
  - výchozí hodnota: 400
- limit konvergence (convergence limit)
  - rozsah hodnot: $< 0; 1 >$
  - typ: reálná čísla
  - výchozí hodnota: 0,01
  - Definuje limit konvergence, kdy se algoritmus zastaví (pokud do té doby nedosáhne maximálního počtu iterací.
- velikost populace (population size)
  - rozsah hodnot: $\{1; \cdots; 500\}$
  - typ: přirozená čísla
  - výchozí hodnota: 50
  - Udává maximální počet pravidel, které mají být uchovány ve výsledném řešení.
- testovací množina (resamples)
  - rozsah hodnot: $\{1; \cdots; 100\ 000\}$
  - typ: přirozená čísla
  - výchozí hodnota: 2 500
  - Počet bodů použitých k testování pravidel.

4.12 GARP with best subsets

V tomto algoritmu proběhne několik GARP modelů a na základě vyhodnocení uživatelských a systémových chyb jsou nejlepší modely vybrány a spojeny v jeden.

Parametry:

- trénovací podíl (training proportion)
  - rozsah hodnot: $< 0; 100 >$
  - typ: reálná čísla
  - výchozí hodnota: 50
  - Procento výskytových dat, jež budou použita k trénování.
- počet pokusů (total runs)
  - rozsah hodnot: $\{0, \cdots, 10\,000\}$
  - typ: přirozená čísla
  - výchozí hodnota: 20
  - Maximální počet provedení běhu algoritmu.
- práh chyby z opomenutí (hard omission threshold)
  - rozsah hodnot: $< 0; 100 >$
  - typ: reálná čísla
  - výchozí hodnota: 100
  - Maximální přijatelná chyba z opomenutí. Pro použití pouze lehké chyby z opomenutí je vhodné nastavit na 100 %.
- modely pod prahem chyby z opomenutí (models under omission treshold)
  - rozsah hodnot: $\{0, \cdots, 10\,000\}$
  - typ: přirozená čísla
  - výchozí hodnota: 20
  - Minimální počet modelů pod prahem chyby z opomenutí.
- práh pro systémovou chybu (commission treshold)
  - rozsah hodnot: $< 0; 100 >$
  - typ: reálná čísla
  - výchozí hodnota: 50
  - Určuje procento distribučních modelů, které mají být přijaty ohledně systémových chyb.
- velikost vzorku pro výpočet systémové chyby (commission sample size)
  - rozsah hodnot: $(1; \infty)$
  - typ: reálná čísla
  - výchozí hodnota: 10 000
  - Počet vzorků, které budou použity k výpočtu systémové chyby.
- maximální počet vláken (maximum number of threads)
  - rozsah hodnot: $\{1, \cdots, 1024\}$
  - typ: přirozená čísla
  - výchozí hodnota: 1
  - Maximální počet vláken provedení, která běží současně.
- počet iterací (max generations)
  - rozsah hodnot: $\{1, \cdots, \infty\}$
  - typ: přirozená čísla
  - výchozí hodnota: 400
  - Maximální počet iterací.

- limit konvergence (convergence limit)
    - rozsah hodnot: $< 0; 1 >$
    - typ: reálná čísla
    - výchozí hodnota: 0,01
    - Definuje limit konvergence, kdy se algoritmus zastaví (pokud do té doby nedosáhne maximálního počtu iterací).
- velikost populace (population size)
    - rozsah hodnot: $\{1, \cdots, 500\}$
    - typ: přirozená čísla
    - výchozí hodnota: 50
    - Udává maximální počet pravidel, které mají být uchovány ve výsledném řešení.
- validační množina (resamples)
    - rozsah hodnot: $\{1, \cdots, 100\,000\}$
    - typ: přirozená čísla
    - výchozí hodnota: 2500
    - Udává počet bodů použitých k testování pravidel.

4.13 Generalized Additive Model (GAM)

Generalizované aditivní modely (GAM) jsou neparametrickým rozšířením generalizovaných lineárních modelů (GLM). Hlavní výhodou těchto modelů oproti ostatním regresním modelům je jejich schopnost modelovat nelinearitu s použitím neparametrického vyhlazování (Hastie a Tibshirani, 1990 v Palialexis, 2011). Spolu s GLM využívají „link funkci" k propojení průměru závislé proměnné a vyhlazení funkce vysvětlujících proměnných (Guisan a kol., 2002).

Generalizované aditivní modely jsou užitečné při složitějších vztazích mezi proměnnými, které nelze snadno zjednodušit do standardních lineárních nebo nelineárních modelů. Cílem modelu je zobrazit graf hodnot závislých proměnných (výskytů) podél jedné environmentální proměnné a následně co nejšetrněji vypočítat hladkou křivku co nejvěrněji odpovídající datům. Algoritmus každou proměnnou proloží hladkou křivkou a pak aditivně kombinuje výsledky. Program Biomod používá kubický spline, což je kolekce polynomů stupně menšího nebo rovno 3 a podobně jako GLM používá automatizovaný proces postupného výběru nejvýznamnějších proměnných (1).

$$Y = s(X_1, 4) + s(X_2, 4) + s(X_3, 4) \qquad (1)$$

Generalizované aditivní modely jsou při modelování velmi často použity s lineárními modely, tudíž mnoho publikací má tyto dva modely společné. Příkladem je studie s cílem vytvořit modely regionální citlivosti pro disturbance permafrostu (Rudy a kol., 2016) nebo studie modelování kolísání hojnosti ryb v několika estuáriích v Portugalsku (França a kol., 2015). Snahou modelovat dynamiku živin v australských estuáriích se zabývá také práce Richardse a kol. (2014).

Parametry:

- funkce („algo")
    - výchozí hodnota: ‚GAM_gam'
    - Výběr vhodné funkce. Mezi možnosti patří ‚GAM_gam', ‚GAM_mgcv' a ‚BAM_mgcv'.

- rovnice („myFormula“)
    - Určuje použití typického vzorce objektu. Pokud není NULL, jsou argumenty interakce úrovně (interaction.level args) a typ vypnuty. Taktéž je možnost výběru automatického generování GAM vzorce použitím typu a typu argumentů úrovně interakce (nejhladší funkce vygeneruje vzorec). Dalšími hodnotami jsou jednoduchá či polynomická rovnice.
    - Úroveň interakce (výchozí hodnota je 0) je celé číslo odpovídající úrovni interakce mezi uvažovanými proměnými. Stojí za zvážení, že interakce rychle rozšiřují počet efektivních proměnných použitých při GLM, nebo konstruují specifické vzorce objektu.
- vyhlazení („k“)
    - výchozí hodnota: {−1,4}
    - Popis vysvětlované proměnné hladkou funkcí. Toto vyhlazování je možno provést pomocí metody loess nebo metodou kubického spline.
- rodina(„family“)
    - Parametr určující typ distribuce. Mezi argumenty patří: ‚binomial‘, ‚poisson‘, ‚negative.binomial‘, ‚Gamma‘, ‚gaussian‘

4.14 Generalized Boosted Model (GBM)

Zatímco GLM se snaží přizpůsobit do jednoho nejšetrnějšího modelu, který nejlépe vysvětluje vztah mezi výskytem druhu a environmentálními proměnnými, GBM fitují velké množství relativně jednoduchých modelů, jejichž odhady jsou následně spojeny a je získán robustnější odhad.

Algoritmus implementovaný v programu BIOMOD je Boosted Regression Tree – BRT (Friedman a kol., 2001), kde je každý z jednotlivých modelů tvořen buď jednoduchým klasifikátorem nebo regresními stromy. Definováním jednoduchého pravidla založeného na jediné vysvětlující proměnné je opakovaným rozdělením dat vytvořen strom. V každém procesu rozdělení jsou data rozdělena do dvou skupin, z nichž každá je maximálně homogenní. Pro vytvoření konečného modelu jsou iterační metodou postupně přidávány stromy do modelu, přičemž statistické vážení dat zdůrazňuje špatné předpovědi předchozích stromů.

Určení maximálního počtu stromů maximalizuje schopnost modelu dělat přesné předpovědi na nových a nezávislých místech a zároveň zabraňuje nadměrné složitosti modelu. Určit optimální počet stromů lze nastavením počtu provedení křížové validace. Uživatel má také možnost definovat maximální počet stromů které budou fitovány. Neexistuje žádný způsob, jak zjistit, jaký počet je nejlepší, ale Thuiller a kol. (2010) uvádí, že ideální kompromis je mezi 2000 a 5000.

GBM byl použit v analýze vztahů druhové hojnosti ryb žijících při dně a životního prostředí na Novém Zélandu (Leathwick, 2006) nebo při předpovídání vzorů druhové bohatosti rostlin v jižní Africe (Thuiller, 2006).

Parametry:

- rozdělení („distribution“)
    - výchozí hodnota: 'bernoulli'
    - Buď jako řetězec znaků s udáním konkrétního rozdělení nebo seznam s názvy komponent upravujícími rozdělení a všechny další potřebné parametry. Pokud není tento parametr přesně určen, bude se model snažit rozdělení sám odhadnout: v případě, že má odpověď pouze dvě jedinečné

hodnoty, předpokládá se Bernoulliho rozdělení (taktéž Binomické rozdělení). Na základě takovýchto odpovědí model dále rozpoznává multinomické rozdělení, Gaussovské, či Coxův model proporcionálních rizik.
    - Aktuální dostupné možnosti jsou: „gaussian“ (čtvercová chyba), „laplace“ (absolutní ztráta), „tdist“ (t-rozložení), „Bernoulli“ (logistická regrese pro 0-1 výstupů), „huberized“ (hinge ztráta pro 0–1 výstupů), „multinomální“ (klasifikace, pokud existuje více než 2 třídy), „adaboost“ (Adaptive Boosting, jehož výstupem je klasifikátor dvou tříd), „poisson“ (počítá výstupy), „coxph“ (Coxův model proporcionálních rizik), „quantile“, „pairwise“ (měření pořadí).
- počet iterací („n.trees“)
    - výchozí hodnota: 2500
    - Značí ekvivalent počtu iterací a počtu základních funkcí v přídavném rozšíření.
- maximální hloubka („interaction.depth“)
    - výchozí hodnota: 7
    - Určuje maximální hloubku variabilních interakcí, přičemž „1“ implikuje model aditivní, „2“ model s obousměrnými interakcemi, atd.
- počet pozorování („n.minobsinnode“)
    - výchozí hodnota: 5
    - Minimální počet pozorování v koncových uzlech stromů. Udává skutečný počet pozorování, ne jejich celkovou váhu.
- rychlost učení („shrinkage“)
    - výchozí hodnota: 0.001
    - Parametr známý jako rychlost učení či zmenšení velikosti kroku.
- trénovací množina („bag.fraction“)
    - výchozí hodnota: 0.5
    - Náhodně vybraná část trénovací množiny pozorování k navržení dalšího stromu v rozšíření. Tento parametr přináší do modelu jistou míru neurčitosti. Pokud je hodnota menší než 1, pak spuštění toho samého modelu dvakrát povede k dosažení podobného výsledku ovšem s rozdílným přizpůsobením.
- velikost vzorku („train.fraction“)
    - výchozí hodnota: 1
    - Určuje velikost vzorku určeného pro trénování.
- křížová validace („cv.folds“)
    - výchozí hodnota: 3
    - Počet provedení křížové validace. Pokud je větší než 1, pak model kromě obvyklého fitování bude provádět i křížovovou validaci a odhadne chybu zjednodušení.
- zachování dat („keep.data“)
    - výchozí hodnota: FALSE
    - Logická proměnná udávající, zda se s uloženým objektem mají zachovávat i data a jejich indexy.
- průběh procesu („verbose“)
    - výchozí hodnota: FALSE
    - Pokud bude tento parametr pravdivý, model bude zobrazovat průběh procesu.

- odhad počtu iterací („perf.method“)
    - výchozí hodnota: 'cv'
    - Označuje metodu používanou pro odhad optimálního počtu iterací za účelem zlepšení výkonu modelu. Metoda „cv“ extrahuje optimální počet iterací pomocí křížové validace, metoda „OOB“ odhaduje chyby na testovacím souboru (*out-of-bag* odhad) a metoda „test“ používá validaci nebo test pro odhad out-of-sample.

4.15 Generalized linear model (GLM)

Zobecněné lineární modely (GLM) jsou matematických rozšířením obecných linárních modelů, založené na vztahu průměru vysvětlované proměnné a lineární kombinaci vysvětlujících proměnných (tzv. link funkce) (Brych, 2009). Data mohou pocházet z různých rozdělení včetně normálního, binomického, Poissonova, negativního binomického, či gamma (Guisan a kol., 2002).

GLM je méně omezující metodou než klasické vícenásobné regrese, protože poskytuje rozdělení chyb pro závislou proměnnou pomocí nekonstantních variačních funkcí. Pokud není reakce s nezávislou proměnnou lineární, pak mohou být zahrnuty transormace (pokud podmínky dovolí simulaci nevyváženou a bimodální odpovědí).

Nedostatkem modelu je jeho nutnost znát povahu vztahu mezi druhy a jejich environmentálními podmínkami. Mimoto, GLM není vždy natolik flexibilní, aby aproximovalo skutečný regresní povrch.

Pro výběr nejvíce šetrného modelu je použit automatický výběr postupného modelu. Funkce „stepAIC“ staví modely tím, že postupně přidává nové výrazy a testuje, nakolik se zlepšilo přizpůsobení modelu. Statistické kritérium používané výběru modelu s rostoucím přizpůsobením může být buď Akaikovo informační kritérium (AIC) nebo Bayesovo informační kritérium (BIC). Postupná procedura umožňuje odstranění redundance v proměnných a redukuje multikolinearitu.

Metoda GLM byla například použita pro modelování distribuce Eukalyptu (*Eucalyptus cypellocarpa*) v Austrálii (Austin a Meyers, 1996) nebo modelování distribuce čtyř druhů středomořkých dřevin (Thuiller a kol., 2003). Srovnáním 16 algoritmů včetně GLM se na 226 druzích v 6 regionech ve své práci věnuje J.Elith a kol. (2006).

Parametry:

- rovnice („myFormula“)
    - Určuje použití typického vzorce objektu. Pokud není NULL, jsou argumenty interakce úrovně (interaction.level args) a typ vypnuty. Taktéž je možnost výběru automatického generování GLM vzorce použitím typu a typu argumentů úrovně interakce (výchozí hodnotou je kvadratická rovnice (2)). Dalšími hodnotami jsou lineární (1) či polynomická rovnice (3).

$$Y_1 = X_1 + X_2 + X_3 + (X_1 \cdot X_2) + (X_2 \cdot X_3) \qquad (2)$$

$$Y_1 = X_1 + X_1^2 + X_1^3 + X_2^2 + X_3^3 \qquad (3)$$

$$Y_1 = f(X_1 + X_1^2 + X_1^3) + f(X_2 + X_2^2 + X_2^3) \qquad (4)$$

- test („test“)
    - Výběr optimálního substitučního modelu na základě informačních kritérií. Výchozí hodnotou je „AIC“ reprezentující Akaikovo informační kritérium ‚druhou možností je „BIC“ pro Bayesovo informační kritérium. Použití „none“, což je také podporovaná hodnota, vede k posouzení pouze celého modelu bez postupného výběru.
- rodina („family“)
    - Parametr určující typ distribuce. Mezi argumenty patří: ‚binomial‘, ‚poisson‘, ‚negative.binomial‘, ‚Gamma‘, ‚gaussian‘.
- kontrola („control“)
    - Seznam parametrů pro kontrolu fitovacího procesu.

4.16 Maximum Entropy (MAXENT)

Principem maximální entropie je metoda pro analýzu dostupných kvalitativních údajů s cílem určit rozdělení pravděpodobnosti. Uvádí se, že i neúplná distribuce, která kóduje určité zadané údaje je ta, která maximalizuje informaci entropie (či nejistotu) (Thuiller, 2002).

Tato implementace se řídí stejným přístupem jako sotware Maxent, který byl vyvinut za účelem použití metody Maximum Entropy (Phillips a kol., 2004). Standardním experimentem byla metoda ze software OpenModeller srovnávána s Maxent 3.3.3. Za použití veškerých možných kombinací parametrů a vytváření modelů se stejným počtem iterací, distribuce map s korelací (*r*) byla vyšší než 0,999 a bez rozdílu v konečném výsledku (OpenModeller: Documentation, 2015). Avšak předchozí verze algoritmu (1.0) generovaly podstatně odlišné výsledky (Muñoz a kol., 2009).

První verze byly založeny na existující knihovně Maximum Entropy z třetí strany, která ve srovnání s ostatními algoritmy produkovala nekvalitní modely. Následně byl algoritmus několikrát upravován Elisangelem Rodriguesem v rámci jeho doktorského studia. Celkové kompatibility s Maxentem bylo dosaženo díky financování zbývajících prací projektem Brazil-OpenBio. Je třeba uvést, že ne všechny dostupné funkce z Maxent jsou dostupné i v OpenModeller, a to zejména možnost využití sběru systematické odchylky (*bias*), stejně jako mnoho specifických parametrů určených pokročilým uživatelům. Nicméně by měl být běžný uživatel schopen získat kompatibilní výsledky pro všechny ostatní dostupné parametry.

Pojem třída prvků (features) je rozšířená sada původních proměnných a algoritmus Maxent je zhodnocuje jako proměnné na vstupu. V aktuální verzi jich Maxent zhodnocuje šest, a to: lineární, kvadratická, prahová, kategorická, „*hinge*“ a „*product*“ (Elith a kol., 2011). Funkce „product“ umožňuje párové kombinace všech možných proměnných a kroková funkce umožňuje nastavit práh a tím získat odlišnou odpověď pod a nad prahovou hodnotou. Podobné krokové funkci jsou „*hinge*“ features, které navíc umožňují změnu sklonu odpovědi.

Využití metody v oblasti biogeografie a ekologie lze nalézt v práci Reddyho a kol. (2015), Carvalha a kol. (2015) nebo Khosraviho a kol. (2016). Po důkladné rešerši bylo zjištěno, že algoritmus se stal v poslední době zřejmě velmi oblíbeným v Číně, viz práce Yi (2016), Zhang (2016), Yuan (2016), Su (2016) či Xu (2015).

Parametry (Open Modeller):

- počet pozaďových bodů (number of background points)
    - rozsah hodnot: $\{0,\cdots,10000\}$
    - typ: přirozená čísla
    - výchozí hodnota: 10 000
    - Počet pozaďových bodů určených ke generování.
- použití absenčních bodů jako pozaďových bodů (use absence points as background)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 0
    - Pokud jsou k dispozici absenční body, může být tento parametr použit jako pokyn k jejich použití jako pozaďových bodů. Tím se zabrání jejich náhodnému generování a také bude umožněno následné porovnání mezi různými algoritmy.
- zahrnout vstupní body do pozadí (include input points in the background)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - 0 = Ne, 1 = Ano
- počet iterací (number of iterations)
    - rozsah hodnot: $\{1,\cdots,\infty\}$
    - typ: přirozená čísla
    - výchozí hodnota: 500
- ukončení tolerance (terminate tolerance)
    - rozsah hodnot: $(0;\infty)$
    - typ: reálná čísla
    - výchozí hodnota: 0,00001
    - Tolerance pro detekování konvergence modelu.
- výstupní formát (output format)
    - rozsah hodnot: {1,2}
    - typ: přirozená čísla
    - výchozí hodnota: 2
    - 1 = raw, 2 = logistic
- kvadratická funkce (quadratic features)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - Použít kvadratickou funkci?
    - 0 = ne, 1 = ano
- product features
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - 0 = ne, 1 = ano
    - Popisuje párové interakce mezi indikátory.

- hinge features
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - 0 = ne, 1 = ano
    - Kombinuje lineární a krokovou funkci.
- prahová funkce (treshold features)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - 0 = ne, 1 = ano
    - Zahrnuje základní krokové funkce.
- povolení automatizace (auto features)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 1
    - 0 = ne, 1 = ano
    - Tato možnost automatizuje úlohu výběru prvků pomocí empirického algoritmu v závislosti na velikosti vzorku.
- product/threshold treshold
    - rozsah hodnot: $\{1,\cdots,\infty\}$
    - typ: přirozená čísla
    - výchozí hodnota: 80
- kvadratický práh (quadratic treshold)
    - rozsah hodnot: $\{1,\cdots,\infty\}$
    - typ: přirozená čísla
    - výchozí hodnota: 10
    - Počet vzorků, od jejichž dosažení dojde k použití kvadratické funkce (lze použít pouze při povolené automizaci).
- hinge treshold
    - rozsah hodnot: $\{1,\cdots,\infty\}$
    - typ: přirozená čísla
    - výchozí hodnota: 15
    - Počet vzorků, od jejichž dosažení dojde k použití „hinge“ funkce (lze použít pouze při povolené automizaci).

Parametry (Biomod):

- odkaz („path_to_maxent.jar“)
    - Odkaz na maxent.jar soubor. Defaultně je nastaven výchozí adresář.
- alokace paměti („memory_allocated“)
    - rozsah hodnot: $\{64,128,256,512,1024,2048,\cdots\}$
    - typ: přirozená čísla
    - výchozí hodnota: 512
    - Množství rezervované paměti. Pokud je parametr ‚NULL‘, dojde k výchozímu limitnímu parametru JAVA paměti.
- maximální počet iterací („maximumiterations“)
    - typ: přirozená čísla
    - výchozí hodnota: 200
    - Udává maximální počet iterací.

- viditelnost („visible“)
    - typ: logická hodnota
    - výchozí hodnota: ‚FALSE‘
    - Zviditelnění uživatelského rozhraní.
- lineární funkce („linear“)
    - typ: logická hodnota
    - výchozí hodnota: ‚TRUE‘
    - Povolení k použití lineární funkce.
- kvadratická funkce („quadratic“)
    - typ: logická hodnota
    - výchozí hodnota: ‚TRUE‘
    - Povolení k použití kvadratickéi funkce.
- product funkce („product“)
    - typ: logická hodnota
    - výchozí hodnota: ‚TRUE‘
    - Povolení k použití „product“ funkce.
- prahová funkce („treshold“)
    - typ: logická hodnota
    - výchozí hodnota: ‚TRUE‘
    - Povolení k použití krokové funkce.
- hinge funkce („hinge“)
    - typ: logická hodnota
    - výchozí hodnota: ‚TRUE‘
    - Povolení k použití „hinge“ funkce.
- počet vzorků k product a prahové funkci („lq2lqptthreshold“)
    - typ: přirozená čísla
    - výchozí hodnota: 80
    - Určení počtu vzorků od nichž se použije product a prahová funkce.
- Počet vzorků k hinge funci („hingetreshold“)
    - typ: přirozená čísla
    - výchozí hodnota: 15
    - Určení počtu vzorků od nichž se použije hinge funkce.
- parametr prahové funkce („beta_treshold“)
    - typ: reálná čísla
    - výchozí hodnota: -1
    - Regulující parametr všech prahových funkcí. Záporná hodnota indikuje automatické nastavení.
- parametr kategorizující funkce („beta_categorical“)
    - typ: reálná čísla
    - výchozí hodnota: -1
    - Regulující parametr všech kategorizujících funkcí. Záporná hodnota indikuje automatické nastavení.
- parametr lineární, kvadratické a product funkce („beta_lgp“)
    - typ: reálná čísla
    - výchozí hodnota: -1
    - Regulující parametr všech lineárních, kvadratických a product funkcí. Záporná hodnota indikuje automatické nastavení.

- parametr hinge funkce („beta_hinge“)
    - typ: reálná čísla
    - výchozí hodnota: -1
    - Regulující parametr všech hinge funkcí. Záporná hodnota indikuje automatické nastavení.
- prevalence („default_prevalence“)
    - typ: reálná čísla
    - výchozí hodnota: 0,5
    - Defaultní prevalence druhu, udává pravděpodobnost výskytu v prezenčních bodech.

4.17 Multivariate Adaptive Regression Splines (MARS)

Hlavním předpokladem každého lineárního procesu je stabilita koeficientů skrz všechny úrovně vysvětlujících proměnných (v případě časových řad také skrz všechna časová období). Pokud mají koeficienty různé optimální hodnoty a stabilní nejsou, stává se MARS velmi užitečnou metodou analýzy (např. oblast financí, energetiky, ekonomických věd, společenských věd). Metoda MARS, zavedena Friedmanem v roce 1991 (Friedman, 1991), systematicky odhaduje a identifikuje takový model, jehož koeficienty se liší v závislosti na úrovni vysvětlujících proměnných. Zlomové body (*breakpoints*) nebo prahové hodnoty, které definují změnu koeficientu modelu se označují jako spline uzel a lze ho chápat jako podobný částečné regresi.

Algoritmus automaticky volí velikost vyhlazení (jenž je požadováno pro každý prediktor) a také pořadí interakcí prediktorů. To je považováno za projekční metodu, kde sice výběr proměnné není problémem, ale je potřeba stanovit její maximální možnou míru interakce.

MARS jsou kombinací GAM a regresních stromů. Gradient odpovědi závislé proměnné na vysvětlující proměnné je při výpočtu nejprve automaticky rozdělen na jednotlivé úseky pomocí CTA, pro které je pak samostatně spočítán regresní model. To umožňuje zachytit i velmi komplexní odpovědi a interakce či stavy, kdy se koeficient funkce odpovědi závislé proměnné s průběhem vysvětlujících proměnných mění (Friedman, 1991).

Spolu s dalšími metodami byla metoda MARS použita v práci Elith a kol. (2006). Využití statistických přístupů pro prediktivní geomorfologické mapování demonstroval ve své práci Luoto a Hjort (2005). Metoda byla také použita pro předpovídání vlastností lesa (Moisen a Frescino, 2002).

Parametry:

- („degre“)
    - výchozí hodnota: 2
    - Výběr vhodné funkce. Mezi možnosti patří ‚GAM_gam‘, ‚GAM_mgcv‘ a ‚BAM_mgcv‘.
- („nk“)
    - výchozí hodnota: ‚NULL‘
    - Volitelné přirozené číslo určující maximální počet modelových podmínek. Pokud je ‚NULL‘, pak je implicitně použita hodnota MARS funkce definována jako $max(21, 2 \cdot nb\_expl\_var + 1)$)
- („penalty“)
    - výchozí hodnota: 2

- („thresh“)
    - výchozí hodnota: 0,001
- („prune“)
    - výchozí hodnota: ‚TRUE‘

4.18 Niche Mosaic

Tento algoritmus je stále experimentální a není dovoleno ho používat v publikacích bez svolení autora.

Parametry:

- počet iterací (number of iterations)
    - rozsah hodnot: $\{1000, \cdots, \infty\}$
    - typ: přirozená čísla
    - výchozí hodnota: 2 000

4.19 Random Forests (RF)

Náhodné lesy lze použít jak pro klasifikaci, tak pro regresi a pro jejich složitost je na ně nahlíženo jako na černé skříňky. Klasifikační les je klasifikační model vytvořený kombinací určitého počtu klasifikačních stromů, kde každý strom přiřazuje hodnotě vektoru prediktorů nějakou třídu a výsledná klasifikace je dána hlasováním nebo jako průměr pravděpodobnosti (Komprdová, 2012). Oproti tomu je výsledná regresní funkce regresního lesa definována jako vážený průměr regresních funkcí několika stromů. Stromy uvnitř lesa jsou nezávislé a rozhodnutí činí na základě jejich vlastních a nezávislých informací.

Výhodou RF je efektivní běh na velkých databázích a vypořádání se s tisíci vstupními proměnnými bez nutnosti je mazat. RF také disponuje odhady, které proměnné budou pro klasifikaci důležité a vytváří vnitřní nezkreslený odhad chyby generalizace během stavby lesa.

Náhodné lesy generují dva datové objekty. Když je na základě odebraných vzorků nastavena trénovací množina pro aktuální strom, přibližně třetina vzorků je vynechána. Tyto vynechané vzorky jsou nazývány OOB (out-of-bag) a jsou použity k získání nezkresleného odhadu chyby klasifikace v průběhu přidávání stromů do lesa.

Princip metody včetně jejího matematického rozboru ve svém článku uvedl Breiman (2001). Z oblasti ekologie a biogeografie lze uvést následující publikace: Elith a kol. (2006), Prasad a kol. (2006), Fukuda a kol. (2014) či Vezza a kol. (2015). Využití metody RF pro předpověď výkonu elektrárny je uvedena v práci Janouška a kol. (2016) z Vysoké školy Báňské – Technická univerzita Ostrava.

Parametry (Open Modeller):

- počet stromů (number of trees)
    - rozsah hodnot: $\{1, \cdots, 1000\}$
    - typ: přirozená čísla
    - výchozí hodnota: 10

- počet proměnných na jeden strom (number of variables per tree)
    - rozsah hodnot: {±∞}
    - typ: přirozená čísla
    - výchozí hodnota: 0
    - Počet proměnných na každý strom (0 výchozí druhé odmocnině počtu vrstev).
- síla učení bez učitele (force unsupervised learning)
    - rozsah hodnot: {0,1}
    - typ: přirozená čísla
    - výchozí hodnota: 0
    - Pokud jsou poskytnuty absenční body, tento parametr může být použit k ignorování těchto bodů bez nutnosti učení bez učitele.
    - Pokud absenční body k dispozici nejsou, toto učení bude provedeno v každém případě.

Parametry (Biomod):

- do-classif
    - výchozí hodnota: ‚TRUE'
    - typ: přirozená čísla
- počet stromů („ntree")
    - výchozí hodnota: 500
- velikost uzlu („nodesize")
    - výchozí hodnota: 5
- maximální počet uzlů („maxnodes")
    - výchozí hodnota: ‚NULL'

4.20 Support Vector Machines (SVM)

Jedním z algoritmů využívajících ke klasifikaci nepravděpodobnostní výběr je Support Vector Machines. SVM jsou algoritmy určené primárně pro klasifikaci a regresi a jsou pozoruhodné pro jejich přirozenou reprezentaci neznámých lineárních vztahů a také stabilitě a robustnosti při aplikaci vysokodimenzionálních dat (Drake 2014). Mohou efektivně vyřešit problémy nelineární klasifikace použitím „*kernel trick*", kdy je u měření vzdálenosti použito implicitní zobrazení proměnných do vyššího dimenzního prostoru v němž jsou problémy jako identifikace či separace nadrovin značně zjednodušeny (Drake 2014).

SVM zobrazuje vstupní vektory ve vyšší dimenzi prostoru, v němž je zkonstruována maximální separační nadrovina. Na každé straně nadroviny jsou vytvořeny dvě paralelní nadroviny které separují data. Separační nadrovina je taková nadrovina, která maximalizuje vzdálenost mezi dvěma rovnoběžnými nadrovinami. Předpokladem je, že čím větší rozpětí nebo vzdálenost mezi těmito rovnoběžnými rovinami bude, tím lepší bude také chyba zobecnění klasifikátoru. Model SVM závisí pouze na podmnožině trénovacích dat, protože funkce pro stavbu modelu se nestará o tréninkové body ležící mimo meze.

Drake ve své práci (2006) srovnával přesnost tří metod lišící se svým přístupem ke snižování složitosti modelu. Modely byly testovány na nezávislých pozorování jak výskytu, tak absence druhu a bylo zjištěno, že nejlepší metodou je ta, která používá všechny dostupné proměnné a nedochází k předběžnému zpracování pro snížení korelace. Uvádí, že teoreticky tudíž modely SVM předčí modely založené na simulaci pseudoabsenčních dat. Srovnáním one-class SVM algoritmu multi-class klasifikací

se zabývá studie Kanga a Cho (2015). Na výskytu mořských organismů (*Pseudo-nitzschia*) sledovaných v průběhu 8 let je založena práce Gonzálese a kol. (2014). Studie v závěru demonstruje použití přístupu SVM pro jeho velkou schopnost přesné predikce, čímž může model poskytnou včasné varování před hrozící květenou druhu *Pseudo-nitzschia*. Tina Tirelli a kol. (2012) se zabývá domácím druhem ryby (*Alburnus alburnus alborella*) v severní Itálii, jejíž populace během posledních dvaceti let prudce klesla v důsledku vlivu člověka. Proto byl tento druh vybrán k následné reintrodukci a metody SVM bylo využito pro předpověď budoucího výskytu či absence druhu. Algoritmus byl pro modelování ve vodním prostředí dále použit v práci Sadeghi a kol. (2012). Konkrétně se jednalo o invazivní druh vodního kapradí (*Azolla filiculoides*). Modelováním distribuce suchozemských rostlin ve Francouzské Polynésii se zabývá práce s názvem *Support vector machines to map rare and endangered native plants in Pacific islands forests* (Pouteau a kol., 2012). Metody SVM a RF byly srovnány mezi sebou pomocí souborů dat tří druhů, které s ohledem velmi malý počet nálezových dat autoři považovali za vzácné – druh *Lepinia taitensis* s 28 pozorovanými výskyty, druh *Pouteria tahitensis* s 20 výskyty a druh *Santalum insulare var. Raiateense* s 80 výskyty. Je uvedeno, že na základě Kappa statistiky metoda SVM stále mírně překonává metodu RF v oblasti predikce distribuce ohrožených druhů.

Parametry:

- typ SVM (SVM type)
  - rozsah hodnot: {0,1,2}
  - typ: přirozená čísla
  - výchozí hodnota: 0
  - 0 = C-SVC, 1 = Nu-SVC, 2 = one-class SVM
- typ kernel funkce (kernel type)
  - rozsah hodnot: {0,1,2}
  - typ: přirozená čísla
  - výchozí hodnota: 2
  - 0 = lineární: $u' \cdot v$
  - 1 = polynomická: $gamma \cdot u' \cdot v + coef0^{degree}$
  - 2 = radiální základní: $exp(-gamma \cdot |u - v|^2)$
- degree
  - rozsah hodnot: $\{0, \cdots, \infty\}$
  - typ: přirozená čísla
  - výchozí hodnota: 3
  - Pouze pro polynomické kernel funkce.
- gamma
  - rozsah hodnot: $\{\pm\infty\}$
  - typ: reálná čísla
  - výchozí hodnota: 0
  - pouze pro polynomické a radiální kernel funkce
  - Pokud je gamma nastavena na 0, bude výchozí hodnota $1/k$, kde k je počet vrstev.
- coef0
  - rozsah hodnot: $\{\pm\infty\}$
  - typ: reálná čísla
  - výchozí hodnota: 0
  - Pouze pro polynomické kernel funkce

- cost
  - rozsah hodnot: (0,001; +∞)
  - typ: reálná čísla
  - výchozí hodnota: 1
  - pouze pro typ C-SVC
- nu
  - rozsah hodnot: (0,001; +∞)
  - typ: reálná čísla
  - výchozí hodnota: 0,5
  - pouze pro typ Nu-SVC a one-class SVM
- výstup pravděpodobnosti (probability output)
  - rozsah hodnot: {0,1}
  - typ: přirozená čísla
  - výchozí hodnota: 1
  - pouze pro typ C-SVC a Nu-SVC
  - Místo binárního výstupu bude výstup pravděpodobnostní.
- počet pseudoabsenčních bodů (number of pseudoabsences)
  - rozsah hodnot: $\{0, \cdots, \infty\}$
  - typ: přirozená čísla
  - výchozí hodnota: 0
  - Pouze pro typ C-SVC a Nu-SVC, kde nebyly poskytnuty žádné absenční body.

4.21 Surface Range Envelope (SRE)

Algoritmus je totožný s algoritmem Bioclim dostupným v software OpenModeller (viz Bioclim).

Parametry:

- kvantil („quant“)
  - výchozí hodnota: 0,025
  - Kvantil hodnot environmentálních proměnných určených k vyloučení z bioklimatické zóny tolerance.

4.22 Virtual Niche Generator

Tento algoritmus funguje na principu vytvoření virtuálních nik. Nika je zastoupena multivariační Gaussovou křivkou jejíž střední hodnota je stanovena na základě optimálních podmínek a náhodné standardní odchylky (OpenModeller: Documentation, 2015).

Vhodné prostředí je vypočteno na základě předpokladu vzájemné nezávislosti všech proměnných, což znamená, že konečná hodnota je výsledkem individuální vhodnosti zadané každou proměnnou.

Jednotlivé vhodnosti prostředí jsou vypočteny jako výsledek Gaussovské pravděpodobnostní hustotní funkce zmenšené o takový faktor, aby optimální podmínky odpovídaly hodnotě 1.

Směrodatné odchylky pro každou proměnnou jsou zvoleny náhodně v rozmezí $[X \cdot S, S]$, kde S je směrodatná odchylka z celé nativní oblasti (vypočítána na základě pozaďových bodů) a X je směrodatná odchylka faktoru parametru mezi 0 a 1.

Parametry:

- počet pozaďových bodů (number of backround points)
  - rozsah hodnot: $\{0, \cdots, 10000\}$
  - typ: přirozená čísla
  - výchozí hodnota: 10000
  - Počet background bodů, které mají být generovány, což bude použito pro odhad směrodatné odchylky pro každou proměnnou v dané oblasti zájmu
- použití absenčních bodů jako pozaďových bodů (use absence points as background)
  - rozsah hodnot: {0,1}
  - typ: přirozená čísla
  - výchozí hodnota: 0
  - Pokud jsou absenční body k dispozici, tento parametr může sloužit jako k pokyn k jejich použití také jako pozaďových bodů.
- práh vhodnosti prostředí (suitability treshold)
  - rozsah hodnot: $< 0; 1 >$
  - typ: reálná čísla
  - výchozí hodnota: 1
  - Práh vhodnosti prostředí k získání binární niky.
  - K zachování průběžné niky je třeba zadat hodnotu 1
- faktor směrodatné odchylky (standard deviation factor)
  - rozsah hodnot: $< 0; 1 >$
  - typ: reálná čísla
  - výchozí hodnota: 0
  - Faktor (x) slouží ke kontrole minimálního limitu náhodné směrodatné odchylky pro každou proměnnou. Hodnota náhodné směrodatné odchylky bude v rozmezí $[x \cdot S; S]$, kde S je směrodatná odchylka z celé nativní oblasti. Při použití mnoha environmentálních proměnných je potřeba tento faktor zvýšit pro zvětšení nik.

5 TVORBA ROZHODOVACÍHO STROMU

Goodchild v roce 1992 (Goodchild, 1992) popsal dvě složky reálné geografie relevatní pro mapování prostorové distribuce druhů: pole a entitu. Jako entita je brána existence diskrétních geografických objektů roztroušených v geografickém prostoru, který je jinak prázdný – při aplikaci na oblast SDM jsou entitou výskytové body. Druhá složka, pole, zastupuje geografické (tedy environmentální) proměnné. Ty mohou být kvalitativního či kvantitativního charakteru, ale oproti entitě mají hodnotu (čili mohou být měřeny) na každém místě. Veškeré vstupní aspekty modelování jsou případně detailněji popsány v práci Hartmannové (2016) nebo Brycha (2009).

Z rozhodovacího procesu byly vynechány dva algoritmy: algoritmus Niche Mosaic, protože je stále experimentální a v žádých publikacích dosud nebyl použit a algoritmus Consensus, protože pouze spojuje již existující algoritmy v jeden.

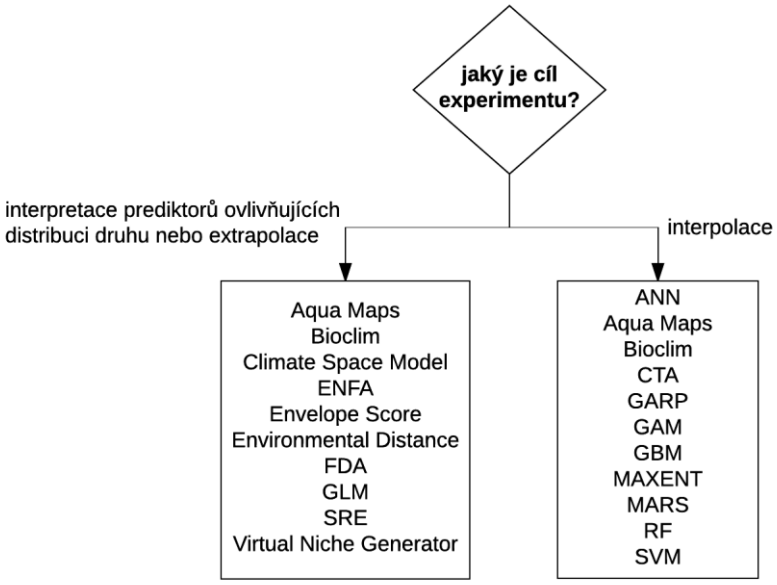
V následujících podkapitolách jsou prezentovány jednotlivé aspekty uvažované jako vstupy do rozhodovacího stromu. Výsledná tabulka pro rozhodovací strom je ke shlédnutí jako volná příloha č. 1.

5.1 Cíl experimentu

Prvním krokem při výběru správného modelu by vždy mělo být následující ujasnění:

1. Je cílem experimentu interpretovat environmentální faktory (pole), jež ovlivňují distribuci druhu (tzn. nalezení těch prediktorů, které nejvíce ovlivňují jeho distribuci)?
2. Je cílem experimentu interpolovat výskyt druhu (entitu) na neprozkoumaných lokalitách?
3. Je cílem experimentu extrapolovat výskyt druhu do nového území?

Obecně lze říci, že pro 1. a 3. je lepší použít parametrické metody (mají nízký rozptyl a vysokou systematickou chybu – bias). Naopak pro 2. se doporučuje použít pro jejich vysoký rozptyl a nízkou systematickou chybu neparametrické nebo samoučící metody (Obr. 5.1) (Eshafani, 2008).



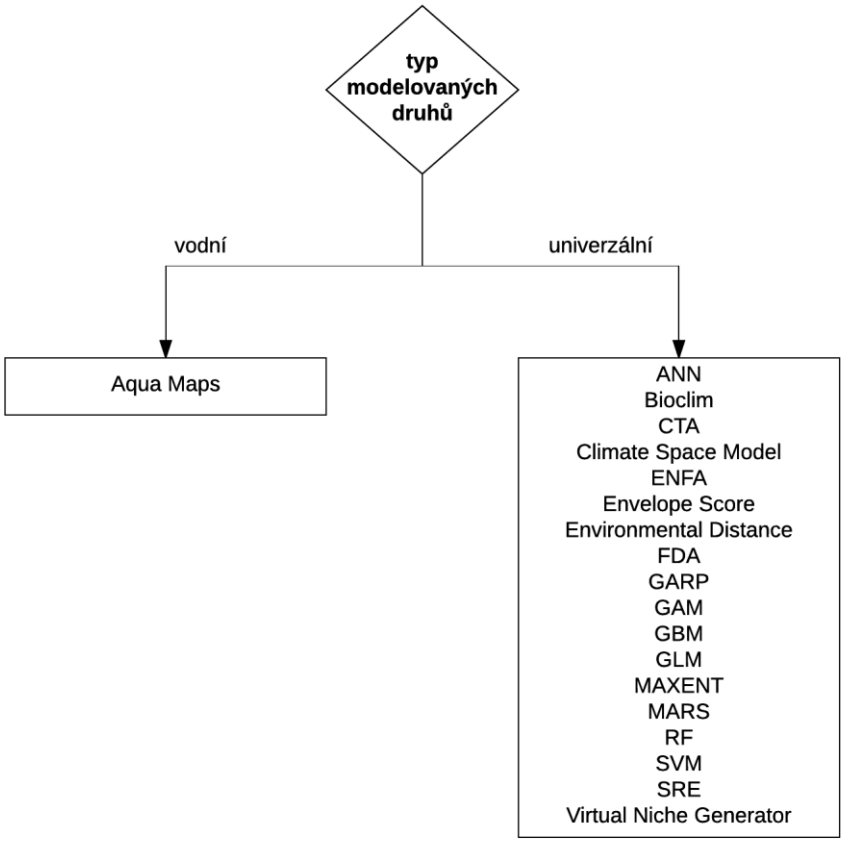
Obr. 5.1 Diagram výběru metody podle cíle experimentu

5.2 Účel experimentu

V prvotní fázi výběru by se měl uživatel také zaměřit na typ modelovaných druhů (účel experimentu). Budou modelovány:

1. vodní nebo
2. suchozemské druhy?

Na základě dostupné literatury a referencí algoritmů (kapitola 3) byly metody rozděleny na hydrologicky zaměřené (Fyhr a kol., 2013; Reiss a kol., 2011), na metody modelující suchozemské druhy, případně na metody, jež lze použít univerzálně. Jak lze vidět na obrázku (Obr. 5.2), kromě algoritmu Aqua Maps bylo ke všem ostatním nalezeno využití pro modelování jak mořských, tak suchozemských druhů. Tudíž byly klasifikovány jako „univerzální“.



Obr. 5.2 Diagram výběru metody podle účelu experimentu

5.3 Vstupní data

Modely vhodnosti habitatu mohou být generovány použitím metod, které požadují informace o přítomnosti druhů (*presence-only*) nebo o přítomnosti a absenci druhů (*presence-absence*). Na základě typu vstupních dat byl rozhodovací strom rozdělen na dva, a to právě podle typu požadovaných vstupních dat. Data požadovaná jednotlivými algoritmy shrnuje tabulka 5.1 (Tab. 5.1).

5.4 Pseudoabsenční body

Pokud model požaduje na vstupu absenční body a uživatel je nemá k dispozici, je možné nechat body automaticky vygenerovat algoritmem. Takové body se nazývají pseudoabsenční. S metodami Maximum Entropy a ENFA je potřeba jednat obezřetně, protože byť se prezentují jako *presence-only* metody, pseudoabsenční body generují taky. Není tedy pravidlem, že pseudoabsenční data používají pouze *presence-absence* metody.

Problém pseudoabsenčních bodů nastává v programu Biomod. Protože generuje soubor 9 různých algoritmů, které se řídí jinými teoretickými přístupy a které zároveň mají jiné požadavky na vstupní data, požaduje program generování pseudoabsenčních dat při každém spuštění. Nekteré algoritmy je ovšem budou interpretovat jako reálné absenční body, jiné jako pseudoabsenční body a další jako pozaďová data.

Prakticky každá *presence-absence* metoda při neobdržení absenčních dat dokáže vygenerovat svá vlastní pseudoabsenční data. Je na expertním uvážení, jaký budou mít pseudoabsenční data, jež budou použita jako absenční, na výsledný model vliv.

Tab. 5.1 Požadavky na vstupní data

metoda	typ vstupních dat	vyžaduje absenční body?
ANN	P/A	NE
Aqua Maps	P	NE
Bioclim	P	NE
Classification Tree Analysis	P/A	NE
Climate Space Model	P	NE
ENFA	P	NE
Envelope Score	P	NE
Environmental Distance	P	NE
Flexible Discriminant Analysis	P	NE
GARP	P	ANO
GAM	P/A	ANO
GBM	P/A	NE
GLM	P/A	ANO
Maximum Entropy	P	NE
MARS	P/A	NE
Random Forest	P/A	NE
SVM (One-class)	P	NE
SVM (multiple-class)	P/A	NE
Surface Range Envelope	P	NE
Virtual Niche Generator	P/A	NE

** P = prezenční, P/A = prezenční+absenční*

5.5 Kvalita dat

Pod pojmem „kvalita dat“ si lze představit citlivost algoritmu na úplnost dat, protože ne vždy se podaří shromáždit dostatečně kvalitní a reprezentativní vzorek pozorovaného druhu (např. nálezová data obsahují pouze hnízdící ptactvo).

Mezi algoritmy požadující dobrou kvalitu prezenčních a absenčních dat patří metody generující statistické funkce nebo rozlišovací pravidla, která umožňují seřadit vhodnost habitatu na základě distribuce přítomnosti nebo absence druhu. Mezi ně patří například analýzy klasifikačních a regresních stromů a neuronové sítě.

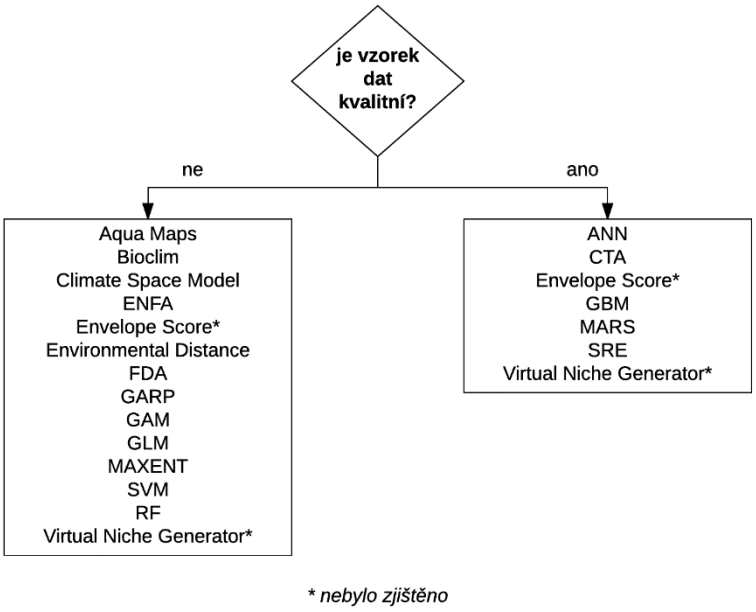
V případě, kdy jsou informace o nepřítomnosti druhu nedostatečné, či nejsou k dispozici, lze použít algoritmy vyžadující pouze prezenční data. Tyto metody (např. ENFA, Bioclim nebo Environmental Distance) vycházejí z definice bioklimatických zón tolerance okolo míst výskytu druhu, které jsou následně porovnány s pozaďovými daty.

Limitním faktorem výběru vhodného algortimu je počet vzorků. U modelování distribuce např. kriticky ohrožených druhů, kde není vždy zaručen dostatečný počet výskytových bodů, je lepší použít samoučící metody, kterým stačí 10 vzorků k maximální predikci. Navíc, při modelování vzácných druhů může být model kombinací mála výskytových bodů a mnoha prediktorů snadno přeučen.

Mnoho studií se zabývá optimálním počtem vzorků, které zaručí co největší přesnost modelu. Stockwell a Peterson (2002) například uvádí, že k dosažení optimálního výsledku stačí 50–100 vzorků. Pro regresní modely pro maximální predikci Coudon a Gégout (Coudon a G´egout, 2007) zmiňují kompromis okolo 50 bodů. Tyto studie mají velmi praktický význam, protože ačkoliv některé datové soubory používané při SDM zahrnují tisíce pozorování, často je počet pozorování přítomnosti pro daný druh mnohem menší.

Jak se tedy vyhnout přeučení modelu při modelování distribuce vzácných druhů? Frank T. Breiner (2015) uvádí, že problém lze vyřešit redukcí počtu environmentálních proměnných. Obecným pravidlem je, že počet výskytových bodů (tzn. velikost vzorku) by měl být desetkrát větší než počet prediktorů. Při malé velikosti vzorku je ale tento poměr velmi těžké udržet, neboť pro 20 výskytových bodů by to znamenalo zahrnutí pouze dvou prediktorů.

Citlivost algoritmů na kvalitu dat shrnuje následující diagram (Obr. 5.3). Skutečnost je ovšem taková, že možnost použití algoritmu na nekvalitních datech neznamená jeho vyloučení ze skupiny algoritmů vhodných k modelování na kvalitních datech.



Obr. 5.3 Vývojový diagram výběru metody kvality vstupních dat

5.6 Environmentální faktory

Problematice environmentálních dat není v literatuře věnováno tolik pozornosti jako jiným aspektům SDM (nálezová data a modelovací metody). Otázka prostorového rozlišení a kvality dat prediktorů byla zkoumána v práci Aspinalla a Pearsona (1996) a studie Susany Suárez-Seoane (2004) poukazuje na zlepšení predikce použitím klimatických proměnných a landuse získaných z družicových dat. Přesto vybrání správných vrstev korelujících s výskytem druhu závisí na uživateli a je často konzultováno s odborníky. Přehled prostorových dat reprezentující základní environmentální režimy (klima, nadmořská výška, orientace, sklon, vegetace, geologie, disturbance, data z metod dálkového průzkumu země, apod.) je uveden v práci Franklin a Miller (2009).

5.6.1 Limit vstupních vrstev

Limit vstupních vrstev pro jednotlivé algoritmy zjištěn nebyl, neboť žádná publikace se tomuto tématu explicitně nevěnuje. Jediný limit je uveden u algoritmu Bioclim, kde je stanoveno maximum vstupních vrstev na 35. Pokud by byl u některého algoritmu přesažen limitní počet environmentálních proměnných, lze velmi snadno najít ty, které s výskytem druhu korelují nejméně a následně je z modelu vyloučit.

5.6.2 Formát vstupních vrstev

Jak již bylo řečeno v úvodu kapitoly, environmentální vrstvy lze dělit na kvalitativní (1) a kvantitativní (2). Kvalitativní se dále dělí na nominální a ordinální a lze s nimi provádět téměř všechny matematické operace. Kvantitativní data se dělí na intervalová a poměrová. Příkladem jsou nadmořská výška (kvantitativní proměnná) a typ vegetace (kvalitativní proměnná).

V regresních modelech (Tab. 5.1) může být využito tzv. umělých proměnných, které se používají, pokud chceme do modelu zahrnout proměnné, které se nedají přímo

kvantifikovat (např. využití krajiny, typ půdy). Umělé (*dummy*) proměnné jsou obvykle binární, ale lze použít i jinou škálu. Pokud je použito více kategorií, může být pro regresní metody obtížné odhadnout a interpretovat významné parametry pro všechny kategorie, které se vyskytují v datech. V tom případě je lepší použít rozhodovací stromy, které mají navíc výhodu automatické interakce mezi proměnnými (v regresním modelování musí být interakce stanoveny předem).

S ohledem na kvalitativní a kvanitativní data, většina algoritmů si je schopná poradit s oběma typy. Jedinou výjimkou jsou umělé neuronové sítě a metoda ENFA, které na vstupu požadují pouze data kvalitativního charakteru.

Dalším aspektem stojícím za zvážení je projekce vrstev, které jsou však limitovány spíše vývojovým prostředím, než samotným algoritmem. V SW OpenModeller je na výběr několik souřadnicových systémů (pro Česko je k dispozici systém WGS84/UTM Zone 33N). Samotný OpenModeller ale souřadnicové systémy převést neumí a v případě nutnosti změny projekce vrstev je nutné použít dostupný GIS software či knihovnu PROJ4. Výhodou programu je jeho schopnost poradit si s rastry s jiným rozlišením (velikostí buňky) a s rozdílnou velikostí území jednotlivých vrstev.

Program Biomod žádné souřadnice nenačítá a při modelování žádnou geografickou informaci nerozpoznává. V tom případě musí uživatel zajistit, aby veškeré datové soubory byly uloženy ve správném pořadí, tzn. aby každá informace o výskytu či nevýskytu druhu byla správně propojena s prediktorem. Lze tedy usuzovat, že všechny vstupní vrstvy musí mít stejnou rozlohu zabírajícího území a stejnou velikost buňky.

Protože změna rozlišení, projekce a další úpravy environmentálních vrstev jsou plně v kompetenci uživatele a neměly by rozhodovat o výběru vhodného algoritmu pro modelování, do výsledného rozhodovacího stromu tento aspekt zahrnut nebyl.

5.7 Míra predikce

Protože pochopení ekologických požadavků druhů na jejich potenciální distribuci je stěžejním faktorem, použití různých modelovacích technik vyžaduje další zkoumání, jako je například posouzení prediktivního výkonu algoritmu a jeho stabilita za různých podmínek.

Míra predikce je testována jako soulad mezi pozorovanou a simulovanou distribucí; stabilita modelu je určena standardní odchylkou, variačním koeficientem, Kappa statistikou, či pomocí hodnot AUC (Area Under the ROC Curve). Prakticky každá studie potenciální distribuce druhu, ať už za použití jedné nebo více metod, ve svém závěru statisticky hodnotí míru predikce a stabilitu vybrané metody.

Pro rozhodovací strom byla míra predikce jednotlivých algoritmů stanovena na základě dostupných publikací, ve kterých bylo hodnoceno minimálně 5 algoritmů na minimálně dvou datových souborech (Duan a kol., 2014; Peterson, 2011; Elith a kol., 2008; Eshafani, 2008) (Tab. 5.2)

Tab. 5.2 Míra predikce jednotlivých metod

metoda	míra predikce[1]	metoda	míra predikce[1]
ANN	2–3	Generalized Additive Model	2–3
Aqua Maps	1-3*	Generalized Boosted Model	3
Bioclim	1	Generalized Linear Model	2–3
Classification Tree Analysis	1–2	Maximum Entropy	3
Climate Space Model	1	MARS	2–3
ENFA	2	Random Forest	3
Envelope Score	1-2	SVM (one-class)	3
Environmental Distance	1	SVM (multiple-class)	3
Flexible Discriminant Analysis	2	Surface Range Envelope	1
GARP	1–3	Virtual Niche Generator	2

[1] *1 = nízká, 2 = střední, 3 = vysoká*

** nebylo zjištěno*

5.8 Povaha výstupu algoritmu

Posledním uvažovaným aspektem je povaha výstupu algoritmu – co vlastně algoritmus dělá? Obecně lze říci, že algoritmy počítající bioklimatickou zónu tolerance (Bioclim, SRE, Virtual Niche Generator, ...) každou buňku zařadí do třídy vhodnosti pro daný druh a výstupem jsou diskrétní hodnoty (např. 0%, 50% a 100% pravděpodobnost výskytu). Výstupy spojitého charakteru mají neuronové sítě, RF a další.

Povaha výstupu algoritmu v rozhodovacím stromu zohledněna nebyla, nicméně je uvedena v tabulce v příloze (Příloha č. 1).

6 SIMULACE

Pro simulaci bylo vybráno jen několik algoritmů reprezentující svou kategorii danou výpočetní metodou (Tab. 6.1). Prostorová distribuce je v této práci názorně demonstrována pouze na příkladě perleťovce velkého (*Argynnis aglaja)*. Výstupy algoritmů pro další druhy motýlů lze shlédnout v přílohách a na přiloženém DVD.

Pokud není uvedeno jinak, byla u algoritmů použita defaultní nastavení.

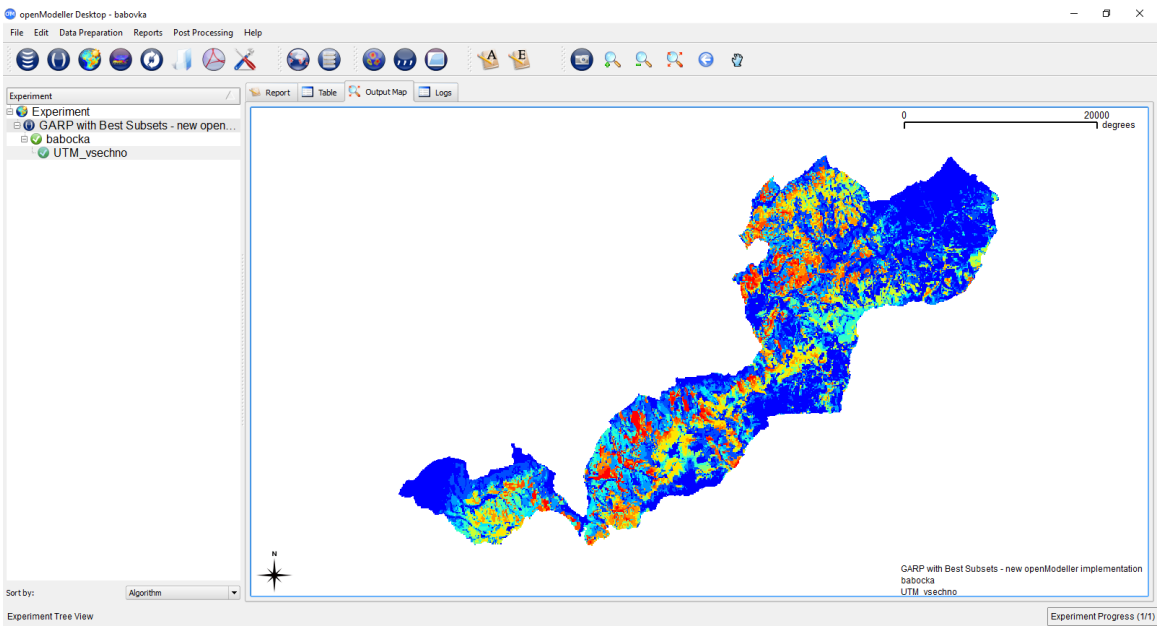
Tab. 6.1 Typy metod a jejich simulace

metoda	typ metody	simulováno
ANN	samoučící (regrese)	ne
Aqua Maps	statistická (environmentální obálka)	ne
Bioclim	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	ano
Classification Tree Analysis	samoučící (stromy)	ano
Climate Space Model	podobnostní a expertní pravidla (faktorová analýza)	ano
ENFA	podobnostní a expertní pravidla (faktorová analýza)	ano
Envelope Score	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	ne
Environmental Distance	podobnostní a expertní pravidla (podobnostní pravidlo)	ano
Flexible Discriminant Analysis	podobnostní a expertní pravidla (faktorová analýza)	ano
GARP	samoučící (genetický algoritmus)	ano
Generalized Additive Model	statistický (regrese)	ne
Generalized Boosted Model	samoučící (stromy, regrese)	ano
Generalized Linear Model	statistický (regrese)	ano
Maximum Entropy	samoučící (entropie)	ano
Multivariate Adaptive Regression Splines	samoučící (regrese)	ano
Random Forest	samoučící (stromy)	ano
Support Vector Machines	samoučící (kernel funkce)	ano
Surface Range Envelope	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	ne
Virtual Niche Generator	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	ano

6.1 Simulace v programu OpenModeller

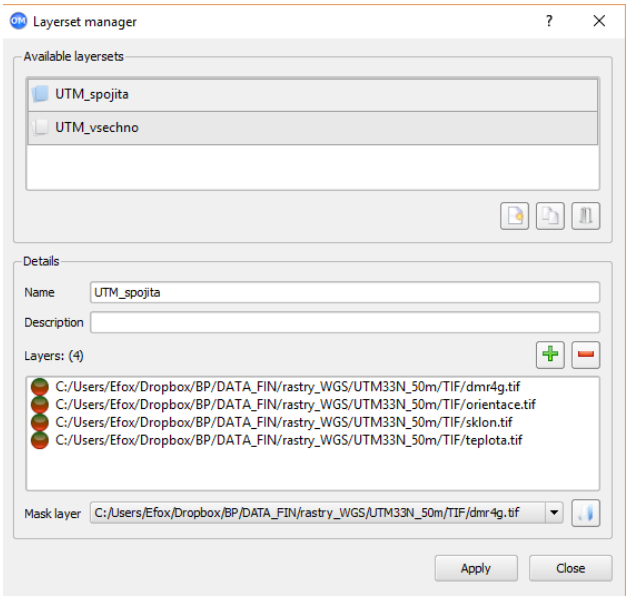
V SW OpenModeller bylo testováno 7 algoritmů. Výsledné mapy predikce perleťovce velkého jsou k dispozici v přílohách č. 2–8, pro další druhy motýlů v přílohách č. 12, 14 a 16.

Uživatelské rozhraní programu OpenModeller lze shlédnout na obrázku níže (Obr. 6.1).



Obr. 6.1 Uživatelské rozhraní programu OpenModeller

V prvním kroku byly nahrány všechny dostupné environmentální proměnné do Správce vrstev. Layersety byly vytvořeny celkem dva, a to jednak se všemi dostupými vrstvami a jednak s vrstvami majícími pouze kvalitativní charakter (Obr. 6.2).



Obr. 6.2 Správce uživatelských vrstev

V případě nestejného rozlišení lze nastavit jednu vrstvu jako masku, která bude sloužit jako vzor k převzorkování.

Následně byl pro modelování založen nový experiment, kde byla nahrána nálezová data a připojeny environmentální vrstvy. Výhodou programu je možnost spuštění všech algoritmů najednou, ovšem za cenu časové náročnosti celého procesu.

Pro perleťovce většího byly na základě výstupu z jednotlivých algoritmů sestaveny mapy predikce výskytu jednotlivě. Pro ostatní druhy motýlů (bělopáska topolového, babočky jilmové a modráska hnědoskvrnného) byly výstupy poskládány do modelu jednoho.

Pro následné vizuální srovnání (příloha č. 9) byly hodnoty pravděpodobnosti výskytu jednotlivých metod reklasifikovány na dva intervaly s mezní hodnotou 80 %. Jelikož měla metoda ENFA maximální hodnotu pravděpodobnosti výskytu 20 %, byla z reklasifikace vynechána.

Program také skýtá možnost exportu všech proměnných do formátu CSV, kde jsou výskytovým bodům (v tabulce jsou reprezentovány souřadnicemi) jednotlivě přiřazovány hodnoty prediktorů (Tab. 6.2).

Tab. 6.2 Tabulka entit a prediktorů

#Num	Lat	Long	dmr4g	geologie	orientace	sklon	teplota	topografie
1	17.5498	48.8208	551.18	25	305.916	5.86915	24.3254	3
2	17.5348	48.8233	568.68	25	344.162	3.4518	24.3746	3
3	17.5283	48.8156	478.23	25	306.795	9.12417	24.4087	3
7	17.5223	48.8206	451.73	7	213.384	5.06799	24.4373	3
9	17.3837	48.8236	399.37	29	358.47	4.20819	25.4384	3
12	17.5768	48.8279	594.13	25	343.263	4.73432	24.2174	3
13	17.5461	48.8457	454.89	7	25.5654	13.8926	24.3136	3
14	17.5358	48.8429	514.94	25	297.684	11.9058	24.395	3
15	17.5388	48.8496	435.01	25	5.82452	9.92629	24.359	3

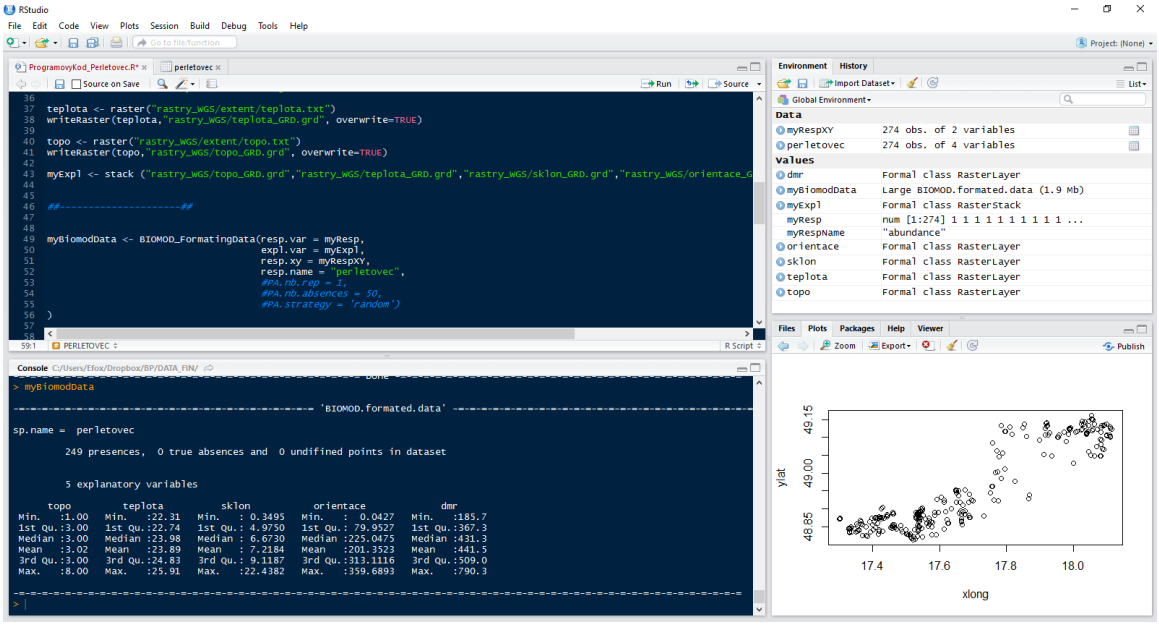
6.2 Simulace v programu BIOMOD

Program BIOMOD (Obr. 6.3) obsahuje sérii funkcí, které umožňují modelování prostorové distribuce včetně generování pseudoabsenčních bodů, počítání hustoty pravděpodobnosti výskytu, vykresování grafů funkcí apod.

Funkce „Models“ spouští různé implementované modely spolu s jejich vyhodnocením pomocí tří technik (Kappa statistika, True Skill Statistic – TSS a ROC křivka – Relative Operating Characteristic). Volba každého modelu je provedena zadáním T (True) nebo F (False).

Jelikož se program neumí vypořádat s různým rozlišením rastrů a také s různě velkým územím, je nutno vždy environmentální data sjednotit na stejné rozlišení a na stejné území.

Programové kódy napsané pro modelování jsou k dispozici na přiloženém DVD.



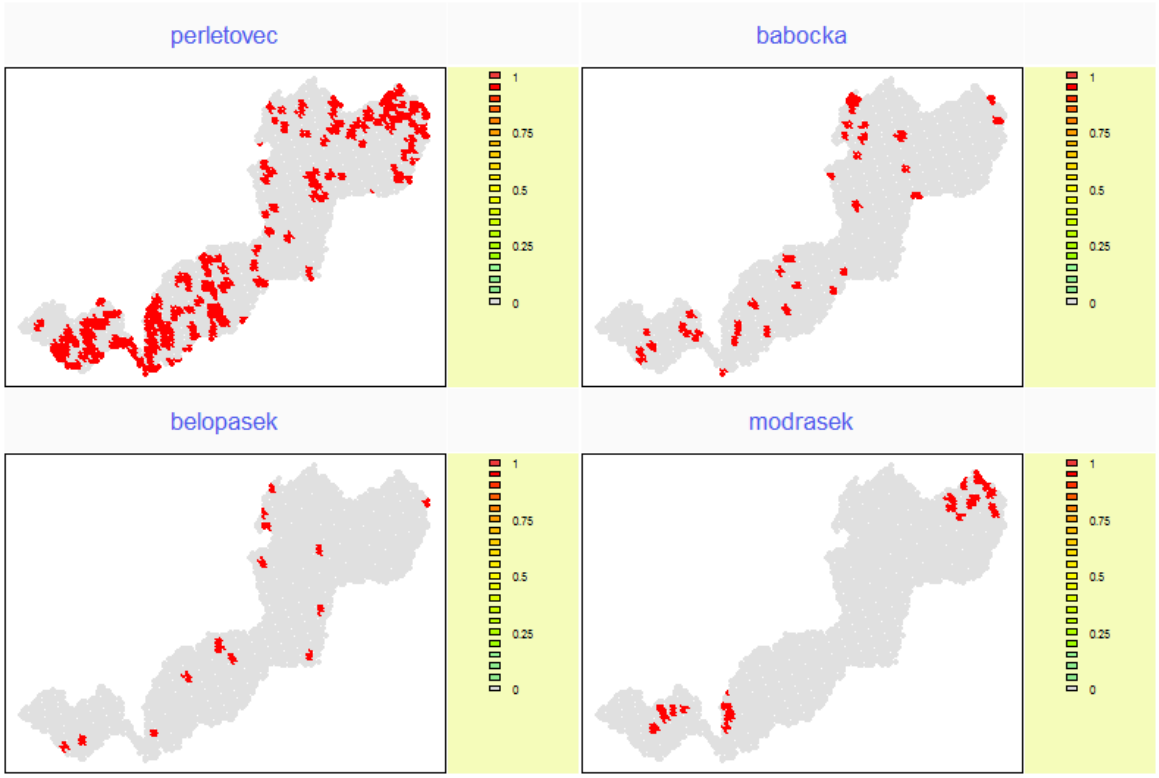
Obr. 6.3 Uživatelské prostředí programu Biomod

Data byla předpřipravena v software ArcMap 10.2.2. Napříč celým územím byla vygenerována čtvercová polygonová síť o velikosti čtverce 500 metrů. Současně byla vygenerována bodová vrstva centroidů jednotlivých čtverců. Každému centroidu byla poté přiřazena hodnota ze všech šesti environmentálních vrstev na kterých bod ležel. V posledním kroku byl každému bodu přiřazen výskyt či nevýskyt jednoho ze čtyř druhů motýlů, a to na základě vzdálenosti od zaznamenaného výskytu. Pokud ležel bod v okolí 250 metrů od zaznamenaného výskytu, byla mu přiřazena hodnota 1. V opačném případě 0. Výsledná tabulka, jež byla naimportována do programu Biomod je na obrázku níže (Obr. 6.4)

```
> head(Sp.Env)
  Idw        X        Y      dmr orientace topo  teplota geologie     sklon perletovec babocka
1   1 17.38792 48.81744 482.6168 334.19452    3 25.40963       29  3.657900          0       0
2   2 17.39472 48.81730 450.9000 310.19031    3 25.38305       10  4.201595          0       0
3   3 17.52401 48.81454 437.7000 288.08154    3 24.39643        7 10.076011          1       1
4   4 17.36091 48.82250 440.8300  12.77608    3 25.58939       25  6.310195          1       0
5   5 17.36771 48.82236 417.9200  15.60199    3 25.55081       26  4.664931          1       0
6   6 17.37452 48.82222 400.9349 320.89694    3 25.51263       11  7.040876          1       0
  belopasek modrasek
1         0        0
2         0        0
3         0        0
4         0        0
5         0        0
6         0        0
>
```

Obr. 6.4 Hlavička tabulky naimportované do RStudia

Na obrázku (Obr. 6.4) lze vidět jednotlivé vstupní proměnné, kde „Idw“ zastupuje roli jednoznačného identifikátoru a „X“ a „Y“ zeměpisné souřadnice, které pro samotné modelování nejsou potřeba, ovšem své využití nalézají při zakreslování do grafů apod. Pole „dmr“, „orientace, „topo“, „teplota“, „geologie“ a „sklon“ obsahují údaje environmentálních proměnných. Zbylé hodnoty prezentují výskyt / nevýskyt druhu (Obr. 6.5)



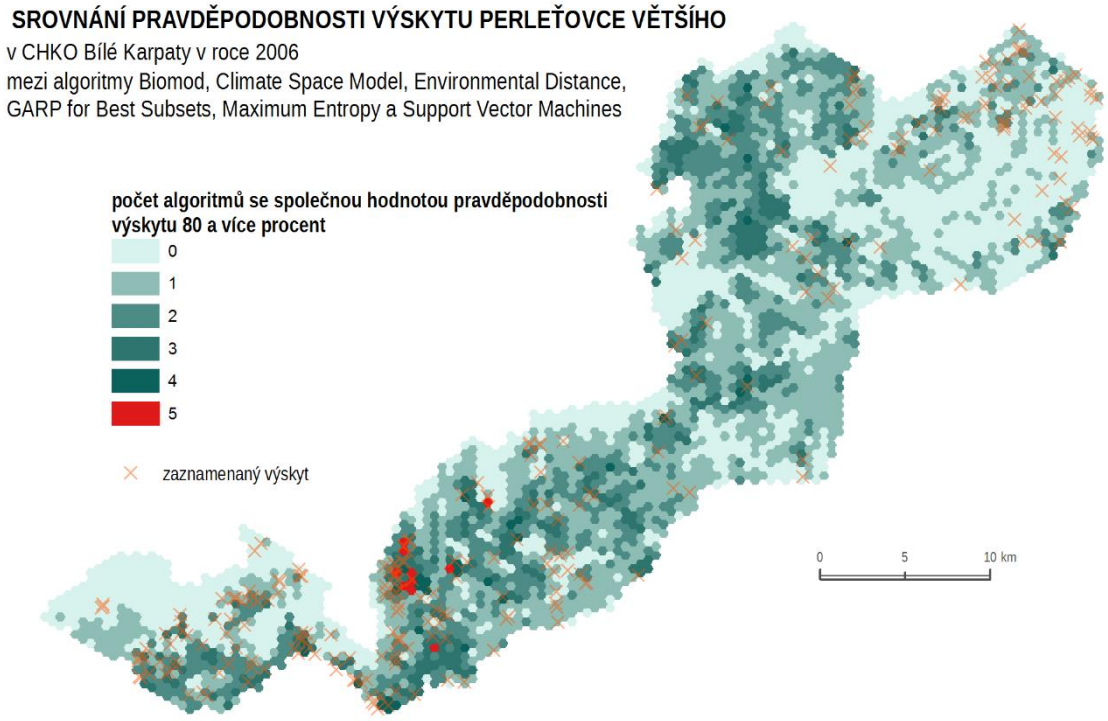
Obr. 6.5 Zobrazení nálezových dat v SW RStudio

Spuštěno bylo celkem sedm modelů (ANN, GBM, GLM, MARS, FDA, RF) na 4 druzích motýlů, s jedním opakováním (Thuiller a kol., 2012). Celkový počet modelů byl tedy ve výsledku 48 a je dostupný na přiloženém DVD. Pro testování byla data rozdělena na trénovací a testovací množinu, kde trénovací množina obsahovala 80 % dat a testovací 20 %. Jednotný výstup ze všech algoritmů pro perleťovce většího je v příloze č. 10 a pro ostatní druhy motýlů je k dispozici na přiloženém DVD.

7 VÝSLEDKY

7.1 Statistický report simulací

Statistický report byl vytvořen na základě výstupů jednotlivých algoritmů z programu OpenModeller (viz kapitola 5.2). Pro potřebnou analýzu bylo všech 6 výstupních vrstev nahráno do programu ArcMap 10.2.2. Následně byly jednotlivě překlasifikovány, a to na hodnotu 0 pro interval 0–79 % a 1 pro interval 80–100 %. Tím bylo docíleno unifikace všech vrstev – byly akcentovány hodnoty s pravděpodobností výskytu větší než 80 % a naopak hodnoty nižší než 80 % byly potlačeny. V dalších kroku byly vrstvy agregovány do předem vytvořené hexagonální sítě na základě průměru hodnot, jež spadaly do daného hexagonu. Konečným sečtením všech vrstev bylo zjištěno, kolik algoritmů, z maximálního počtu šesti, predikovalo pravděpodobnost výskytu 80 % a více. Jak lze vidět na obrázku níže (Obr. 7.1), algoritmy v žádné buňce neměly tuto hodnotu společnou, tudíž maximální pravděpodobnost výskytu byla dosažena pouze pomocí pěti algoritmů.

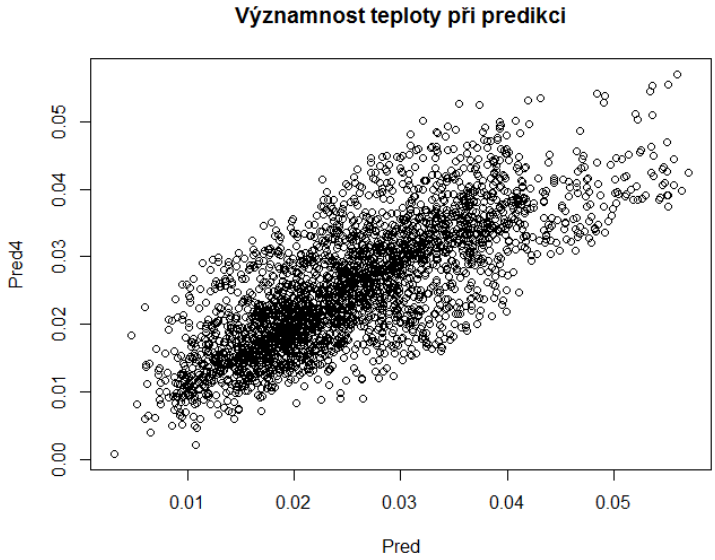


Obr. 7.1 Srovnání pravděpodobnosti výskytu perleťovce velkého

Vizuální srovnání algoritmů s mírou predikce více než 80 % lze shlédnout v příloze č. 9 pro perleťovce většího, v příloze č. 13 pro babočku jilmovou, v příloze č. 15 pro bělopáska topolového a v příloze č. 17 pro modráska hnědoskvrnného. V těchto vizuálních srovnáních se jednalo hlavně o zvýraznění predikce výskytu s více než 80% praděpodobností.

Porovnávat předpovědi různých modelovacích technik je velmi obtížné, protože modely jsou závislé na odlišných algoritmech a na odlišných předpokladech ohledně distribuce druhu v závislosti na environmentálních proměnných. Jednou z výhod modelování v programu Biomod je jeho schopnost určení významnosti prediktorů a tím

i ulehčení porovnání mezi modely. Tento odhad významnosti proměnné v modelu je určen jako hodnota korelace mezi standardním predikčním modelem (tj. model, do kterého vstupovaly všechny proměnné) a mezi novým predikčním modelem, kde byla zkoumaná proměnná randomizována. Pokud se korelace mezi oběma predikcemi jen mírně liší, ukazuje, že randomizovaná proměnná má na predikci malý vliv a nelze ji tedy považovat za důležitého prediktora. To znamená, že čím menší korelace, tím větší vliv na distribuci. Na obrázku (Obr. 7.2) je znázorněna viditelná korelace s hodnotou 0,75 mezi predikcemi perleťovce většího, kde je randomizovanou proměnnou „teplota“. Příkazem „VarImportance“ lze zjistit důležitost pro všechny proměnné, veškeré modely a skrz všechny studované druhy.



Obr. 7.2 Korelace mezi predikcemi

7.2 Rozhodovací stromy

Na základě aspektů důležitých pro predikci distribuce byly vytvořen rozhodovací strom. Z důvodu velkého počtu informací bylo rozhodnuto o rozdělení stromu na dva a to pro: *presence-only* metody a *presence/absence* metody, tedy na základě typu vstupních dat. Rozhodovací stromy jsou vloženy v práci jako volné přílohy č. 18 a č. 19 a jsou také k dispozici na přiloženém DVD.

Mezi aspekty uvažovanými jako vstupy do rozhodovacího stromu patřilo následující :

1. cíl experimentu,
2. účel experimentu,
3. typ vstupních dat,
4. generování pseudoabsenčních bodů algoritmem,
5. kvalita vstupních dat,
6. limit vstupních environmentálních vrstev,
7. formát vstupních environmentálních vrstev,
8. míra predikce a
9. povaha výstupu algoritmu.

Každá z těchto položek je popsána v kapitole č.5 včetně odůvodnění proč byla či nebyla zahrnuta do tvorby rozhodovacího stromu.

Dodatečným zdrojem pro výběr adekvátního algoritmu je také tabulka (příloha č.1), která vznikala soubežně s kapitolou č. 5. Jsou v ní obsaženy veškeré uvažované faktory důležité při výběru adekvátního algoritmu včetně dodatečných poznámek.

8 DISKUZE

Samotné téma prostorové distribuce druhů vyžaduje odborný přístup a nebylo v možnostech autora všechna tato specifika obsáhnout, přesto je práce vhodným úvodem do celé problematiky včetně náležitého přehledu algoritmů dostupných v programech OpenModeller a Biomod zahrnující i jejich požadované vstupní parametry.

Výběr programu pro modelování prostorové distribuce druhů zavisí z velké části na zaměření uživatele. Pro geoinformatika bude jistě vhodnější program OpenModeller, už jenom z toho důvodu, že výstupy z programu Biomod nejsou interoperabilní s gisovým software, což může být pro následné analýzy značnou nevýhodou. Oproti tomu, matematik či statistik najde více výhod při práci v programu Biomod, částečně i kvůli velkému množství následných statistických analýz, které program OpenModeller nenabízí.

S ohledem na provedené simulace lze uvažovat o navázání na tuto práci, jež se bude věnovat právě možnostem hodnocení predikce jednotlivých modelů. Byť se v programu Biomod tyto možnosti vyloženě nabízejí, bylo nutné je v práci vynechat. Jednak kvůli již dostatečně velké obsáhlosti tématu a jednak kvůli nutnosti ovládat programovací jazyk R více, než jen elementárně. Autor si taktéž plně uvědomuje neplnohodnotnost provedených simulací, ze kterých nelze vyvozovat žádné reálné závěry o prostorové distribuci zkoumaných druhů motýlů. Algoritmy byly testovány formálním způsobem pouze jako ukázka jejich použití a potenciálních výsledků. Pro adekvátní výsledky ohledně potenciální distribuce motýlů je třeba jednat více odborněji a zohlednit také biologické cykly zkoumaných druhů a v nemalé míře se také zaměřit na faktory, které daný druh v rozmístění ovlivňují nejvíce, což v v této práci zkoumáno nebylo a autor vycházel pouze z dodaných dat pěti environmentálních proměnných.

Rozhodovací stromy by měly sloužit primárně pro začínající uživatele, kteří ještě nemají dostatečný odborný přehled o všech možnostech modelování prostorové distribuce druhů. Nicméně by měli mít alespoň základní přehled o vícerozměrných statistických metodách a o statistickém hodnocení biodiverzity, protože v práci tyto základní pojmy vysvětlovány nejsou.

Kartograf by mohl poznamenat, že výsledné simulace nejsou zpracovány natolik vhodně, aby uživateli podaly informace o potenciální distribuci zkoumaných druhů motýlů, nicméně cílem nebylo tuto informaci ani poskytnout. Jednalo se hlavně o demonstraci potenciální distribuce tak, jak ji predikují jednotlivé algoritmy.

9 ZÁVĚR

Cílem bakalářské práce bylo vytvoření rozhodovacího stromu, který bude sloužit k výběru adekvátního algoritmu v závislosti na cíli experimentu a kvalitě vstupních dat. Při výběru správné metody pomocí rozhodovacího stromu, jenž souhlasí s výběrovými požadavky uživatele je eliminována nutnost studia všech teoretických východisek algoritmů a usnadněna možnost studia jen vybrané metody.

Rozhodovací strom byl v závěru pro větší přehlednost rozdělen, a to podle typu vstupních dat (*presence-only* a *presence-absence*). Mezi další uvažované výběrové aspekty patřil cíl a účel experimentu, typ vstupních dat, generování pseudoabsenčních bodů algoritmem, kvalita vstupních dat, limit a formát vstupních environmentálních vrstev, míra predikce algoritmu a povaha výstupu algoritmu. Byť všechny tyto aspekty při tvorbě rozhodovacího stromu zohledněny nebyly, pro větší přehlednost byla dodatečně vytvořena tabulka obsahující algoritmy se všemi uvažovanými faktory a při výběru adekvátního algoritmu pomocí rozhodovacího stromu je doporučováno přihlédnou i na ni.

V rámci simulace byly algoritmy otestovány v programech OpenModeller a Biomod. Pro program Biomod je k dispozici i programový kód v jazyce R, jenž byl pro modelování distribuce napsán. Dle zadání bakalářské práce byly výsledné simulace perleťovce většího statisticky zhodnoceny pomocí hexagonální sítě a vizuálně srovnány. Bylo vytvořeno 7 map, jež slouží jako výstupy z predikce jednotlivých algoritmů pro druh perleťovce většího. Dále byly vytvořeny dvě mapy pro každý druh sloužící jako vizuální srovnání predikce algoritmů.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

ABRAHAMYAN, Armine a Arvids BARSEVSKIS. Environmental Niche Modelling with Desktop GARP for Wild Origanum vulgare L . (Lamiaceae) in Armenia. *Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference*. 2015, **3**, 7-. DOI: 10.17770/etr2013vol3.869. ISSN 2256-070x.

ABRAHART, Robert J., Pauline K. KNEALE a Linda M. SEE. *Neural networks for hydrological modelling*. Leiden: A.A. Balkema, 2004. ISBN 90-580-9619-X.

ANDĚL, Petr, Tereza MINÁRIKOVÁ a Michal ANDREAS (eds.). *Ochrana průchodnosti krajiny pro velké savce*. Liberec: Evernia, 2010. ISBN 978-80-903787-5-9.

ANDERSON, Robert P, Daniel LEW a A.Townsend PETERSON. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*. 2003, **162**(3), 211-232. DOI: 10.1016/S0304-3800(02)00349-6. ISSN 03043800.

AUSTIN, M.P. a J.A. MEYERS. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecology and Management*. 1996, **85**(1-3), 95-106. DOI: 10.1016/S0378-1127(96)03753-X. ISSN 03781127.

BARBET-MASSIN, Morgane, Walter JETZ a Risto HEIKKINEN. A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions*. Australian Government Publishing Service, 2014, **20**(11), 1285-1295. DOI: 10.1111/ddi.12229. ISSN 13669516.

BOMBI, Pierluigi, Daniele SALVI, Leonardo VIGNOLI a Marco A. BOLOGNA. Modelling Bedriaga's rock lizard distribution in Sardinia: An ensemble approach. *Amphibia-Reptilia*. 2009, **30**(3), 413-424. DOI: 10.1163/156853809788795173. ISSN 01735373.

BREIMAN, Leo. Random Forests. *Machine Learning*. 2001, **45**(1), 5-32. DOI: 10.1023/A:1010933404324. ISSN 08856125.

BREINER, Frank T., Antoine GUISAN, Ariel BERGAMINI, Michael P. NOBIS a Barbara ANDERSON. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*. 2015, **6**(10), 1210-1218.

BROTONS, Lluís, Wilfried THUILLER, Miguel B. ARAÚJO a Alexandre H. HIRZEL. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*. 2004, **27**(4), 437-448. DOI: 10.1111/j.0906-7590.2004.03764.x. ISSN 09067590.

BRYCH, Pavel. *Modelování potenciálního šíření invazitních druhů rostlin v ČR: Porovnání metod a jejich implementací, dostupnost dat a vliv ekologie druhu na přesnost predikce*. České Budějovice, 2009. Magisterská práce. Katedra botaniky, Jihočeská Univerzita v Českých Budějovicích. Vedoucí práce RNDr. Stanislav Mihulka, PhD.

CARPENTER, G., A. N. GILLISON a J. WINTER. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*. 1993, **2**(6), 667-680. DOI: 10.1007/BF00051966. ISSN 0960-3115.

CARUSO, N., M. GUERISOLI, E.M. LUENGOS VIDAL, D. CASTILLO, E.B. CASANAVE a M. LUCHERINI. Modelling the ecological niche of an endangered population of Puma concolor: First application of the GNESFA method to an elusive carnivore. *Ecological Modelling*. 2015, **297**, 11-19. DOI: 10.1016/j.ecolmodel.2014.11.004. ISSN 03043800.

CARVALHO, Bruno M., Elizabeth F. RANGEL, Paul D. READY, Mariana M. VALE a Nigel BEEBE. Ecological Niche Modelling Predicts Southward Expansion of Lutzomyia (Nyssomyia) flaviscutellata (Diptera: Psychodidae. *PLOS ONE*. Australian Government Publishing Service, 2015, **10**(11), e0143282-. DOI: 10.1371/journal.pone.0143282. ISSN 1932-6203.

CECCARELLI, Soledad, Agustín BALSALOBRE, María SUSEVICH, María ECHEVERRIA, David GORLA a Gerardo MARTI. Modelling the potential geographic distribution of triatomines infected by Triatoma virus in the southern cone of South America. *Parasites*. 2015, **8**(1), 153-. DOI: 10.1186/s13071-015-0761-1. ISSN 1756-3305. Dostupné také z: http://www.parasitesandvectors.com/content/8/1/153

CIVÍN, Lukáš. *Vrstevnaté neuronové sítě a jejich aplikace při dobývání znalostí*. Praha, 2006. Diplomová práce. Katedra Softwarového Inženýrství, Univerzita Karlova v Praze. Vedoucí práce RNDr. Iveta Mrázová, CSc.

CORO, Gianpaolo, Chiara MAGLIOZZI, Anton ELLENBROEK, et al. Automatic classification of climate change effects on marine species distributions in 2050 using the AquaMaps model. *Environmental and Ecological Statistics*. 2016, **23**(1), 155-180. DOI: 10.1007/s10651-015-0333-8. ISSN 1352-8505.

COSTA, H, V MEDEIROS, E B AZEVEDO, L SILVA a José GONZALEZ-ANDUJAR. Evaluating ecological-niche factor analysis as a modelling tool for environmental weed management in island systems. *Weed Research*. 2013, **53**(3), 221-230. DOI: 10.1111/wre.12017. ISSN 00431737.

COUDUN, Christophe a Jean-Claude GÉGOUT. Quantitative prediction of the distribution and abundance of Vaccinium myrtillus with climatic and edaphic factors. *Journal of Vegetation Science*. 2007, **18**(4), 517-524.

DAUBNER, Lukáš. *Deep learning*. Brno, 2015. Bakalářská práce. Fakulta Informatiky, Masarykova Univerzita. Vedoucí práce Doc. RNDr. Lubomír Popelínský, Ph.D.

DE'ATH, Glenn, Katharina E. FABRICIUS, Mathieu ROUGETI, et al. Classification and regression trees: a powerfull yet simple technique for ecological data analysis. *Ecology*. 2000, **81**(11), 3178-3192. DOI: 10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2. ISSN 0012-9658.

DOLGENER, N., L. FREUDENBERGER, M. SCHLUCK, N. SCHNEEWEISS, P. L. IBISCH a Ralph TIEDEMANN. Environmental niche factor analysis (ENFA) relates environmental parameters to abundance and genetic diversity in an endangered amphibian, the fire-bellied-toad (Bombina bombina). *Conservation Genetics*. 2014, **15**(1), 11-21. DOI: 10.1007/s10592-013-0517-4. ISSN 1566-0621.

DORMANN, Carsten F., Stanislaus J. SCHYMANSKI, Juliano CABRAL, et al. Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*. 2012, **39**(12), 2119-2131. DOI: 10.1111/j.1365-2699.2011.02659.x. ISSN 03050270.

DRAKE, JOHN M., CHRISTOPHE RANDIN a ANTOINE GUISAN. Modelling ecological niches with support vector machines. *Journal of Applied Ecology*. 2006, **43**(3), 424-432. DOI: 10.1111/j.1365-2664.2006.01141.x. ISSN 0021-8901.

DRAKE, John M. Ensemble algorithms for ecological niche modeling from presence-background and presence-only data. *Ecosphere*. 2014, **5**(6), art76-. DOI: 10.1890/ES13-00202.1. ISSN 2150-8925.

DUAN, Ren-Yan, Xiao-Quan KONG, Min-Yi HUANG, Wei-Yi FAN, Zhi-Gao WANG a Enrique HERNANDEZ-LEMUS. The Predictive Performance and Stability of Six Species Distribution Models. *PLoS ONE*. 2014, **9**(11), e112764-. DOI: 10.1371/journal.pone.0112764. ISSN 1932-6203.

ELITH, Jane, Catherine H. GRAHAM, Robert P. ANDERSON, et al. Novel methods improve prediction of species' distributions from occurrence data: bridging a dichotomy. *Ecography*. 2006, **29**(2), 129-151. DOI: 10.1111/j.2006.0906-7590.04596.x. ISSN 09067590. Dostupné také z: http://doi.wiley.com/10.1111/j.2006.0906-7590.04596.x

ELITH, Jane, Steven J. PHILLIPS, Trevor HASTIE, Miroslav DUDÍK, Yung En CHEE a Colin J. YATES. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*. 2011, **17**(1), 43-57. DOI: 10.1111/j.1472-4642.2010.00725.x. ISSN 13669516.

ELITH, Jane a Robert J. HIJMANS. Species distribution modeling with R. *R project* [online]. 2011 [cit. 2016-07-21]. Dostupné z: https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf

ESFAHANI, Mostafa Tarkesh. *Predictive Vegetation Modelling: Comparison of Methods, Effect of Sampling Design and Application on Different Scales*. Jena, 2008. Disertační práce. Harmazeutischen Fakultät der Friedrich-Schiller-Universität Jena.

ESTRADA-CONTRERAS, Israel; EQUIHUA, Miguel; CASTILLO-CAMPOS, Gonzalo and ROJAS-SOTO, Octavio. Climate change and effects on vegetation in Veracruz, Mexico: an approach using ecological niche modelling. *Act. Bot. Mex* [online]. 2015, n.112 [cited 2016-07-31], pp.73-93. ISSN 0187-7151.

FAYYAD, Usama, Gregory PIATETSKY-SHAPIRO a Padhraic SMYTH. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. 1996, **17**(3), 37–54.

FRANKLIN, Janet a Jennifer A. MILLER. *Mapping species distributions: spatial inference and prediction*. Cambridge: Cambridge University Press, 2009. Ecology, biodiversity and conservation. ISBN 978-0-521-70002-3.

FRANÇA, Susana, Henrique N. CABRAL, Paul TREITZ, et al. Predicting fish species richness in estuaries: Which modelling technique to use? *Environmental Modelling*. 2015, **66**(2), 17-26. DOI: 10.1016/j.envsoft.2014.12.010. ISSN 13648152.

FRIEDMAN, Jerome H., Miroslav DUDÍK a Robert E. SCHAPIRE. Multivariate Adaptive Regression Splines. *The Annals of Statistics*. New York, New York, USA: ACM Press, 1991, **19**(1), 1-67. DOI: 10.1214/aos/1176347963. ISBN 1581138285. ISSN 0090-5364. Dostupné také z: http://projecteuclid.org/euclid.aos/1176347963

FRIEDMAN, Jerome H., Milani CHALOUPKA, Darrell STRAUSS, et al. Machine: Which modelling technique to use? *The Annals of Statistics*. 2001, **29**(5), 1189-1232. DOI: 10.1214/aos/1013203451. ISSN 0090-5364.

FUKUDA, Shinji, Eriko YASUNAGA, Marcus NAGLE, Kozue YUGE, Vicha SARDSUD, Wolfram SPREER a Joachim MÜLLER. Modelling the relationship between peel colour and the quality of fresh mango fruit using Random Forests: Bagging and Random Forests for Ecological Prediction. *Journal of Food Engineering*. 2014, **131**(2), 7-17. DOI: 10.1016/j.jfoodeng.2014.01.007. ISSN 02608774.

FYHR, Frida, Åsa NYLSSON a Antonia NYSTRÖM SANDMAN. *A review of Ocean Zoning tools and Species distribution modelling methods for Marine Spatial Planning*. Litva: Marmoni, 2013.

GOODCHILD, Michael F a Yang SHIREN. A hierarchical spatial data structure for global geographic information systems. *CVGIP: Graphical Models and Image Processing*. 1992, **54**(1), 31-44.

GONZÁLEZ VILAS, Luis, Evangelos SPYRAKOS, Jesus M. TORRES PALENZUELA a Yolanda PAZOS. Support Vector Machine-based method for predicting Pseudo-nitzschia spp. blooms in coastal waters (Galician rias, NW Spain). *Progress in Oceanography*. 2014, **124**, 66-77. DOI: 10.1016/j.pocean.2014.03.003. ISSN 00796611.

GUISAN, Antoine, Thomas C EDWARDS a Trevor HASTIE. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*. 2002, **157**(2-3), 89-100. DOI: 10.1016/S0304-3800(02)00204-1. ISSN 03043800

HARTMANNOVÁ, Sylvie. *Modelování výskytu živočichů*. Olomouc, 2016. Diplomová práce. Přírodovědecká fakulta, Univerzita Palackého v Olomouci. Vedoucí práce RNDr. Jan Brus, Ph.D.

HASTIE, Trevor, Robert TIBSHIRANI, Andreas BUJA, et al. Flexible Discriminant Analysis by Optimal Scoring: An alternative non-parametric approach for predicting species distributions. *Journal of the American Statistical Association*. 1994, **89**(428), 1255-1270. DOI: 10.1080/01621459.1994.10476866.

HASTIE, Trevor, Andreas BUJA a Robert TIBSHIRANI. Penalized Discriminant Analysis. *The Annals of Statistics*. Institute of Mathematical Statistics, 1995, **23**(1), 73–102.

HASTIE, Trevor a Robert TIBSHIRANI. Discriminant Analysis by Gaussian Mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)*. Institute of Mathematical Statistics, 1996, **58**(1), 155–176.

HIRZEL, A. H., J. HAUSSER, D. CHESSEL a N. PERRIN. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*. 2002, **83**(7), 2027-2036. DOI: 10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2. ISSN 0012-9658.

HIRZEL Alexandre H. a Raphael ARLETTAZ. Modeling Habitat Suitability for Complex Species Distributions by Environmental-Distance Geometric Mean. *Environmental Management*. 2003, **32**(5), 614-623. DOI: 10.1007/s00267-003-0040-3. ISSN 0364-152x.

JANOUŠEK, Jan, Petr GAJDOŠ, Pavel DOHNÁLEK a Michal RADECKÝ. Towards power plant output modelling and optimization using parallel Regression Random Forest. *Swarm and Evolutionary Computation*. 2016, **26**, 50-55. DOI: 10.1016/j.swevo.2015.07.004. ISSN 22106502.

JONES, Miranda C., Stephen R. DYE, John K. PINNEGAR, et al. Modelling commercial fish distributions: Prediction and assessment using different approaches. *Ecological Modelling*. 2012, **225**(1), 133-145. DOI: 10.1016/j.ecolmodel.2011.11.003. ISSN 03043800.

KANG, Seokho a Sungzoon CHO. A novel multi-class classification algorithm based on one-class support vector machine. *Intelligent Data Analysis*. 2015, **19**(4), 713-725. DOI: 10.3233/IDA-150741. ISSN 1088467x.

KEARNEY, Michael R., Brendan A. WINTLE, Warren P. PORTER, et al. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change: bridging a dichotomy. *Conservation Letters*. 2010, **3**(3), 203-213. DOI: 10.1111/j.1755-263X.2010.00097.x. ISSN 1755263x. Dostupné také z: http://doi.wiley.com/10.1111/j.1755-263X.2010.00097.x

KESNER-REYES, K., KULLANDER, S., C. GARILAO, J. BARILE a J. FROESE a K. KASCHNER (eds.). AquaMaps: algorithm and data sources for aquatic organisms. In: FROESE, R. a D. PAULY. *Fish Base* [World Wide Web electronic publication]. AquaMaps, 2012 [cit. 2016-07-22]. Dostupné z: http://www.aquamaps.org/main/fb_book_kreyes_aquamaps_jg.pdf

KHOSRAVI, Rasoul, Mahmoud-Reza HEMAMI, Mansoureh MALEKIAN, Alan L. FLINT a Lorraine E. FLINT. Maxent modeling for predicting potential distribution of goitered gazelle in central Iran: the effect of extent and grain size on performance of the model. *TURKISH JOURNAL OF ZOOLOGY*. 2016, **40**, 574-585. DOI: 10.3906/zoo-1505-38. ISSN 13000179.

KOMPRDOVÁ, Klára. *Rozhodovací stromy a lesy*. Brno: Akademické nakladatelství CERM, 2012. ISBN 978-80-7204-785-7.

KOREŇ, Milan, Slavomír FIND'O, Michaela SKUBAN a Matúš KAJBA. Habitat suitability modelling from non-point data. *Ecological Informatics*. 2011, **6**(5), 296-302. DOI: 10.1016/j.ecoinf.2011.05.002. ISSN 15749541.

KUMAR, Sunil, Lisa G. NEVEN a Wee L. YEE. Evaluating correlative and mechanistic niche models for assessing the risk of pest establishment. *Ecosphere*. 2014, **5**(7), art86-. DOI: 10.1890/ES14-00050.1. ISSN 2150-8925.

LEATHWICK, J. R., J. ELITH, M. P. FRANCIS, et al. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*. 2006, **321**(5), 267-281. DOI: 10.3354/meps321267. ISSN 0171-8630.

MANEL, Stephanie, J.M. DIAS, S.T. BUCKTON a S.J. ORMEROD. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*. 1999, **36**(5), 734-747. DOI: 10.1046/j.1365-2664.1999.00440.x. ISSN 0021-8901.

MASAILA, Aleh. *Regresní stromy*. Praha, 2012. Diplomová práce. Katedra pravděpodobnosti a matematické statistiky, Univerzita Karlova v Praze. Vedoucí práce Mgr. Tomáš Hanzák.

MACEČEK, A. *Rychlost učení vícevrstvé sítě*. Brno, 2011. Bakalářská práce. Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně. Vedoucí bakalářské práce doc. Ing. Václav Jirsík, CSc..

Matematická biologie: e-learningová učebnice [online]. Brno: Masarykova univerzita, 2015 [cit. 2016-08-09]. Dostupné z: http://portal.matematickabiologie.cz/

MISKA, Luoto a Hjort JAN. Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology*. Institute of Mathematical Statistics, 2005, **67**(3-4), 299-315. DOI: 10.1016/j.geomorph.2004.10.006. ISSN 0169555x.

MOISEN, Gretchen G. a Tracey F. FRESCINO. Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*. 2002, **157**, 209–225.

MULLER, Berndt, J. REINHARDT a M. T. STRICKLAND. *Neural networks: an introduction*. 2nd updated and corr. ed. New York: Springer, c1995. ISBN 35-406-0207-0.

MUÑOZ, Mauro Enrique, Renato DE GIOVANNI, Marinez Ferreira DE SIQUEIRA, Tim SUTTON, Peter BREWER, Ricardo Scachetti PEREIRA, Dora Ann Lange CANHOS a Vanderlei Perez CANHOS. OpenModeller: a generic approach to species' potential distribution modelling. *GeoInformatica* [online]. 2009, **15**(1), 111-135 [cit. 2016-07-15]. DOI: 10.1007/s10707-009-0090-7. ISSN 1384-6175.

NIX, H. A. A biogeographic analysis of Australian elapid snakes. *Australian Flora and Fauna Series*. Australian Government Publishing Service, 1986, **7**, 4–15.

OLDEN, Julian D. a Donald A. JACKSON. A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*. 2002, **47**, 1976–1995.

OpenModeller: Documentation [online]. 2015 [cit. 2016-08-09]. Dostupné z: http://openmodeller.sourceforge.net/documentation.html

PALIALEXIS, A., S. GEORGAKARAKOS, I. KARAKASSIS, K. LIKA a V. D. VALAVANIS. Fish distribution predictions from different points of view: comparing associative neural networks, geostatistics and regression models. *Hydrobiologia*. The Cooper Ornithological Society, 2011, **670**(1), 165-188. DOI: 10.1007/s10750-011-0676-6. ISSN 0018-8158.

PETERSEN, M.B., A. TOLVER, L. HUSTED, T.H. TØLBØLL a T.H. PIHL. Repeated measurements of blood lactate concentration as a prognostic marker in horses with acute colitis evaluated with classification and regression trees (CART) and random forest analysis. *The Veterinary Journal*. 2016, **213**, 18-23. DOI: 10.1016/j.tvjl.2016.03.012. ISSN 10900233.

PETERSON, A. Townsend, Kevin P COHOON a Trevor HASTIE. Sensitivity of distributional prediction algorithms to geographic data completeness: A robust and informative method of data analysis. *Ecological Modelling*. 1999, **117**(1), 159-164. DOI: 10.1016/S0304-3800(99)00023-X. ISSN 03043800. Dostupné také z: http://linkinghub.elsevier.com/retrieve/pii/S030438009900023X

PETERSON, A. Townsend. Predicting species' geographic distributions based on ecological niche modeling. *Condor*. The Cooper Ornithological Society, 2001, **103**(3), 599–605. ISSN 0010-5422. Dostupné také z: http://dx.doi.org/10.1650/0010-5422(2001)103[0599:PSGDBO]2.0.CO;2

PETERSON, A. Townsend. *Ecological niches and geographic distributions*. Oxford: Princeton University Press, c2011. Monographs in population biology, 49. ISBN 978-0-691-13688-2.

PETERSON, A. Townsend a David A. VIEGLAIS. Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem. *Bioscience*. 2001, **51**(5), 363–371.

PHILLIPS, Steven J., Miroslav DUDÍK a Robert E. SCHAPIRE. A maximum entropy approach to species distribution modeling. *Twenty-first international conference on Machine learning - ICML '04*. New York, New York, USA: ACM Press, 2004, , 83-. DOI: 10.1145/1015330.1015412. ISBN 1581138285.

POUTEAU, Robin, Jean-Yves MEYER, Ravahere TAPUTUARAI a Benoît STOLL. Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics*. 2012, **9**, 37-46. DOI: 10.1016/j.ecoinf.2012.03.003. ISSN 15749541.

READY, Jonathan, Kristin KASCHNER, Andy B. SOUTH, et al. Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*. 2010, **221**(3), 467-478. DOI: 10.1016/j.ecolmodel.2009.10.025. ISSN 03043800.

REBELO, Hugo a Gareth JONES. Ground validation of presence-only modelling with rare species: a case study on barbastelles Barbastella barbastellus (Chiroptera. *Journal of Applied Ecology*. 2010, **47**(2), 410-420. DOI: 10.1111/j.1365-2664.2009.01765.x. ISSN 00218901.

REDDY, Medagam Thirupathi, Hameedunnisa BEGUM, Neelam SUNIL, Pandravada Someswara RAO, Natarajan SIVARAJ a Shashi KUMAR. Predicting Potential Habitat Distribution of Sorrel (Rumex vesicarius L.) in India from Presence-Only Data Using Maximum Entropy Model. *OALib*. 2015, **02**(06), 1-11. DOI: 10.4236/oalib.1101590. ISSN 2333-9721.

REISS, H, S CUNZE, K KÖNIG, H NEUMANN a I KRÖNCKE. Species distribution modelling of marine benthos: a North Sea case study. *Marine Ecology Progress Series*. 2011, **442**, 71-86. DOI: 10.3354/meps09391. ISSN 0171-8630.

RUPPRECHT, Franziska, Jens OLDELAND a Manfred FINCKH. Modelling potential distribution of the threatened tree species Juniperus oxycedrus: how to evaluate the predictions of different modelling approaches? *Journal of Vegetation Science*. 2011, **22**(4), 647-659. DOI: 10.1111/j.1654-1103.2011.01269.x. ISSN 11009233.

PRASAD, Leo, Louis R. IVERSON a Andy LIAW. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*. 2006, **9**(2), 181-199. DOI: 10.1007/s10021-005-0054-1. ISSN 1432-9840.

RECKNAGEL, Friedrich. ANNA – Artificial Neural Network model for predicting species abundance and succesion of blue-green algae. *Hydrobiologia*. 1997, **349**(1/3), 47-57. DOI: 10.1023/A:1003041427672. ISSN 00188158. Dostupné také z: http://link.springer.com/10.1023/A:1003041427672

RICHARDS, Russell, Milani CHALOUPKA, Darrell STRAUSS, et al. Using Generalized Additive Modelling to Understand the Drivers of Long-Term Nutrient Dynamics in the Broadwater Estuary (a Subtropical Estuary), Gold Coast, Australia: Which modelling technique to use? *Journal of Coastal Research*. 2014, **298**(2), 1321-1329. DOI: 10.2112/JCOASTRES-D-12-00190.1. ISSN 0749-0208.

ROBERTSON, M. P., N. CAITHNESS a M. H. VILLET. A PCA-Based Modelling Technique for Predicting Environmental Suitability for Organisms from Presence Records. *Diversity and Distributions*. 2001, **7**(1/2), 15–27.

RUDY, Ashley C.A., Scott F. LAMOUREUX, Paul TREITZ, et al. Transferability of regional permafrost disturbance susceptibility modelling using generalized linear and generalized additive models: Predicting spatial distributions of plant species at different scales. *Geomorphology*. 2016, **264**(2), 95-108. DOI: 10.1016/j.geomorph.2016.04.011. ISSN 0169555x.

SADEGHI, Roghayeh, Rahmat ZARKAMI, Karim SABETRAFTAR a Patrick VAN DAMME. Use of support vector machines (SVMs) to predict distribution of an invasive water fern Azolla filiculoides (Lam.) in Anzali wetland, southern Caspian Sea, Iran. *Ecological Modelling*. 2012, **244**, 117-126. DOI: 10.1016/j.ecolmodel.2012.06.029. ISSN 03043800.

SÁNCHEZ-FLORES, Erick, Ian R. NOBLE a Trevor HASTIE. GARP modeling of natural and human factors affecting the potential distribution of the invasives Schismus arabicus and Brassica tournefortii in 'El Pinacate y Gran Desierto de Altar' Biosphere Reserve: A robust and informative method of data analysis. *Ecological Modelling*. 2007, **204**(3-4), 457-474. DOI: 10.1016/j.ecolmodel.2007.02.002. ISSN 03043800. Dostupné také z: http://linkinghub.elsevier.com/retrieve/pii/S0304380007000555

STOCKWELL, David R.B., Ian R. NOBLE a Trevor HASTIE. Induction of sets of rules from animal distribution data: A robust and informative method of data analysis. *Mathematics and Computers in Simulation*. 1992, **33**(5-6), 385-390. DOI: 10.1016/0378-4754(92)90126-2. ISSN 03784754. Dostupné také z: http://linkinghub.elsevier.com/retrieve/pii/0378475492901262

SU, W., D. WU, M. ZHANG, F. JIANG, R. ZHANG a H. WU. Upscaling method for corn canopy LAI using MaxEnt model. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*. 2016, **32**(7), 165-172. DOI: 10.11975/j.issn.1002-6819.2016.07.023. ISSN 10026819.

SUÁREZ-SEOANE, Susana, Patrick E. OSBORNE a Andries ROSEMA. Can climate data from METEOSAT improve wildlife distribution models? *Ecography*. 2004, **27**(5), 629-636.

SUTTON, T., R. GIOVANNI a M. F. SIQUEIRA. Introducing openModeller – A fundamental niche modelling framework. *OSGeo Journal*. 2007, **1**. ISSN 1994-1897.

THUILLER, Wilfried, Miguel B. ARAÚJO a Sandra LAVOREL. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*. 2003, **14**(5), 669-680. DOI: 10.1111/j.1654-1103.2003.tb02199.x. ISSN 11009233.

THUILLER, Wilfried, Kevin P COHOON a Trevor HASTIE. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change: A robust and informative method of data analysis. *Global Change Biology*. 2003, **9**(10), 1353-1362. DOI: 10.1046/j.1365-2486.2003.00666.x. ISSN 1354-1013.

THUILLER, Wilfried, Guy F. MIDGLEY, Mathieu ROUGETI, et al. Predicting patterns of plant species richness in megadiverse South Africa: an analysis using boosted regression trees. *Ecography*. 2006, **29**(5), 733-744. DOI: 10.1111/j.0906-7590.2006.04674.x. ISSN 09067590.

THUILLER, Wilfried, Bruno LAFOURCADE, Robin ENGLER a Miguel B. ARAÚJO. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*. 2009, **32**(3), 369-373. DOI: 10.1111/j.1600-0587.2008.05742.x. ISSN 09067590.

THUILLER, Wilfred, Bruno LAFOURCADE a Miguel ARAUJO. *ModOperating Manual for BIOMOD* [online]. Francie, 2009, s. 1–90 [cit. 2016-08-06]. Dostupné z: http://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/inst/doc/Biomod%20Manual.pdf?revision=67&root=biomod&pathrev=218

THUILLER, Wilfred, Damien GEORGES, Robin ENGLER a Bruno LAFOURCADE. *BIOMOD : Tutorial* [online]. 2012, s. 1–88 [cit. 2016-08-06]. Dostupné z: http://www.will.chez-alice.fr/pdf/BiomodTutorial.pdf

TIRELLI, Tina, Marco GAMBA a Daniela PESSANI. Support vector machines to model presence/absence of Alburnus alburnus alborella (Teleostea, Cyprinidae) in North-Western Italy: Comparison with other machine learning techniques. *Comptes Rendus Biologies*. 2012, **335**(10-11), 680-686. DOI: 10.1016/j.crvi.2012.09.001. ISSN 16310691.

TSOAR, Asaf, Omri ALLOUCHE, Ofer STEINITZ, Dotan ROTEM a Ronen KADMON. A comparative evaluation of presence-only methods for modelling species distribution: An ensemble approach. *Diversity and Distributions*. 2007, **13**(4), 397-405. DOI: 10.1111/j.1472-4642.2007.00346.x.

VALLE, Mireia, Ángel BORJA, Guillem CHUST, Ibon GALPARSORO a Joxe Mikel GARMENDIA. Modelling suitable estuarine habitats for Zostera noltii, using Ecological Niche Factor Analysis and Bathymetric LiDAR. *Estuarine, Coastal and Shelf Science*. 2011, **94**(2), 144-154. DOI: 10.1016/j.ecss.2011.05.031. ISSN 02727714.

VAYSSIÈRES, Marc P., Richard E. PLANT, Barbara H. ALLEN-DIAZ, et al. Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*. 2000, **11**(5), 679-694. DOI: 10.2307/3236575. ISSN 11009233. Dostupné také z: http://doi.wiley.com/10.2307/3236575

VÁCLAVÍK, Tomáš, Hansen EVERETT, Alan KANASKIE a Janet OHMANN. Predicting potential and actual distribution of sudden oak death in Oregon: prioritizing landscape contexts for early detection and eradication of disease outbreaks. *Forest Ecology and Management*. 2010, **260**(6), 1026–1035.

VENABLES, W. N. a Brian D. RIPLEY. *Modern applied statistics with S*. 4th ed. New York: Springer, c2002. Statistics and computing. ISBN 0-387-95457-0.

VEZZA, P., R. MUÑOZ-MAS, F. MARTINEZ-CAPEL, A. MOUTON, Vicha SARDSUD, Wolfram SPREER, Joachim MÜLLER a Joachim. Random forests to evaluate biotic interactions in fish distribution models: Bagging and Random Forests for Ecological Prediction. *Environmental Modelling*. 2015, **67**(2), 173-183. DOI: 10.1016/j.envsoft.2015.01.005. ISSN 13648152.

XU, J., B. CAO a C.-K. BAI. Prediction of potential suitable distribution of endangered plant Kingdonia uniflora in China with MaxEnt. *Chinese Journal of Ecology*. 2015, **34**(12), 3354-3359.

YI, Yu-jun, Xi CHENG, Zhi-Feng YANG a Shang-Hong ZHANG. Maxent modeling for predicting the potential distribution of endangered medicinal plant (H. riparia Lour) in Yunnan, China. *Ecological Engineering*. 2016, **92**, 260-269. DOI: 10.1016/j.ecoleng.2016.04.010. ISSN 09258574.

YUAN, Hai-Sheng, Yu-Lian WEI a Xu-Gao WANG. Maxent modeling for predicting the potential distribution of Sanghuang, an important group of medicinal fungi in China. *Fungal Ecology*. 2015, **17**, 140-145. DOI: 10.1016/j.funeco.2015.06.001. ISSN 17545048.

ZHANG, C., L. CHEN, C. M. TIAN, T. LI, R. WANG a Q.-Q. YANG. Predicting the distribution of dwarf mistletoe (Arceuthobium sichuanense) with GARP and MaxEnt models. *Beijing Linye Daxue Xuebao/Journal of Beijing Forestry University*. 2016, **38**(5), 23-32. DOI: 10.13332/j.1000-1522.20150516.

PŘÍLOHY

SEZNAM PŘÍLOH

Vázané přílohy:

Volné přílohy

Popis struktury DVD

Adresáře:

- Data
- Webové stránky
- Text práce
- Přílohy

ASPEKTY VÝBĚRU ADEKVÁTNÍHO ALGORITMU

metoda	typ metody	typ vstupních dat	vyžaduje absenční body?	cíl experimentu	účel experimentu	použití nekvalitního vzorku	nutnost sjednocení environmentálních vrstev	míra predikce[1]	hodnoty jako povaha výstupu	poznámka
Artificial Neural Network (ANN)	samoučící (regrese)	P/A	NE	interpolace	univerzální	ne	ano / ne	2–3	spojitá	Neakceptuje kvantitativní data na vstupu. Metoda vyžaduje odborný přístup a není vhodná pro začínající uživatele.
Aqua Maps	statistická (environmentální obálka)	P	NE	extrapolace nebo interpretace prediktorů	vodní druhy	ano	ne	1-3*	diskrétní	
Bioclim	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	P	NE	extrapolace nebo interpretace prediktorů, interpolace	univerzální	ano	ne	1	diskrétní hodnoty	Akceptuje pouze bioklimatické environmentální proměnné v maximálním počtu 35 vrstev. Při extrapolaci často novou distribuci podceňuje.
Classification Tree Analysis (CTA)	samoučící (stromy)	P/A	NE	interpolace	univerzální	ne	ano	1–2	spojitá / binární	Ve srovnání se statistickými metodami podávají samotné klasifikační stromy velmi slabý výkon. Nicméně jsou vhodné pro grafickou vizualizaci klasifikačních pravidel.
Climate Space Model	podobnostní a expertní pravidla (faktorová analýza)	P	NE	extrapolace nebo interpretace prediktorů	univerzální	ano	ne	1	spojitá	Ve srovnání s ostatními metodami má tendence přeceňovat současnou distribuci.
Ecological Niche Factor Analysis (ENFA)	podobnostní a expertní pravidla (faktorová analýza)	P	NE	extrapolace nebo interpretace prediktorů	univerzální	ano	ne	2	spojitá	Na vstupu požaduje pouze kvalitativní data a byť je prezentována jako presence-only metoda, tak také generuje pozaďové body.
Envelope Score	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	P	NE	extrapolace nebo interpretace prediktorů	univerzální	ano/ne*	ne	1-2	diskrétní	
Environmental Distance	podobnostní a expertní pravidla (podobnostní pravidlo)	P	NE	extrapolace nebo interpretace prediktorů	univerzální	ano	ne	2	spojitá	Výkon často srovnatelný se statistickými metodami.
Flexible Discriminant Analysis (FDA)	podobnostní a expertní pravidla (faktorová analýza)	P	NE	extrapolace nebo interpretace prediktorů	univerzální	ano	ano	2	spojitá	
GARP	samoučící (genetický algoritmus)	P	ANO	interpolace	univerzální	ano	ne	1–3	diskrétní	V porovnání s ostatními metodami podává při modelování špatný výkon. Má sklony přeceňovat aktuální distribuci druhu a podceňovat jeho distribuci v nových podmínkách . Taktéž generuje pseudoabsenční body.
Generalized Additive Model (GAM)	statistický (regrese)	P/A	ANO	interpolace	univerzální	ano	ano	2–3	spojitá	Má podobný, až mírně lepší výkon než metoda GLM. Je užitečný pro možnost vizualizace křivek.
Generalized Boosted Model (GBM)	samoučící (stromy, regrese)	P/A	NE	interpolace	univerzální	ne	ano	3	spojitá	Při dostatečně kvalitním vzorku dat má velmi silnou schopnost predikce.
Generalized Linear Model (GLM)	statistický (regrese)	P/A	ANO	extrapolace nebo interpretace prediktorů	univerzální	ano	ano	2–3	spojitá	Účinná metoda při globálním modelování. S dostatečnými údaji funguje velice dobře , a to i při použití pseudoabsenčních dat.
Maximum Entropy (MAXENT)	samoučící (entropie)	P	NE	interpolace	univerzální	ano	ano / ne	3	spojitá	Funguje dobře na datasety s malým množstvím dat a při projekci v postoru a čase. Stejně jako algoritmus ENFA generuje pseudoabsenční body.
Multivariate Adaptive Regression Splines (MARS)	samoučící (regrese)	P/A	NE	interpolace	univerzální	ne	ano	2–3	spojitá	Výkon algoritmu je adekvátní velikosti vzorku. Na komplexních datech je výpočetně rychlejší než algoritmus GAM.
Random Forest (RF)	samoučící (stromy)	P/A	NE	interpolace	univerzální	ano	ano	3	spojitá	Vhodný na velké datové soubory s velkým počtem environmentálních proměnných.
Support Vector Machines (One-class) (SVM)	samoučící (kernel)	P	NE	interpolace	univerzální	ano	ne	2–3	spojitá	Vhodný na malé datové soubory.
Support Vector Machines (multiple-class) (SVM)	samoučící (kernel)	P/A	NE	interpolace	univerzální	ano	ne	2–3	spojitá	Vhodný na malé datové soubory.
Surface Range Envelope (SRE)	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	P	NE	extrapolace nebo interpretace prediktorů	univerzální	ne	ano	1	diskrétní	
Virtual Niche Generator	podobnostní a expertní pravidla (bioklimatická zóna tolerance)	P/A	NE	extrapolace nebo interpretace prediktorů	univerzální	ano/ne*	ne	1-2*	*	

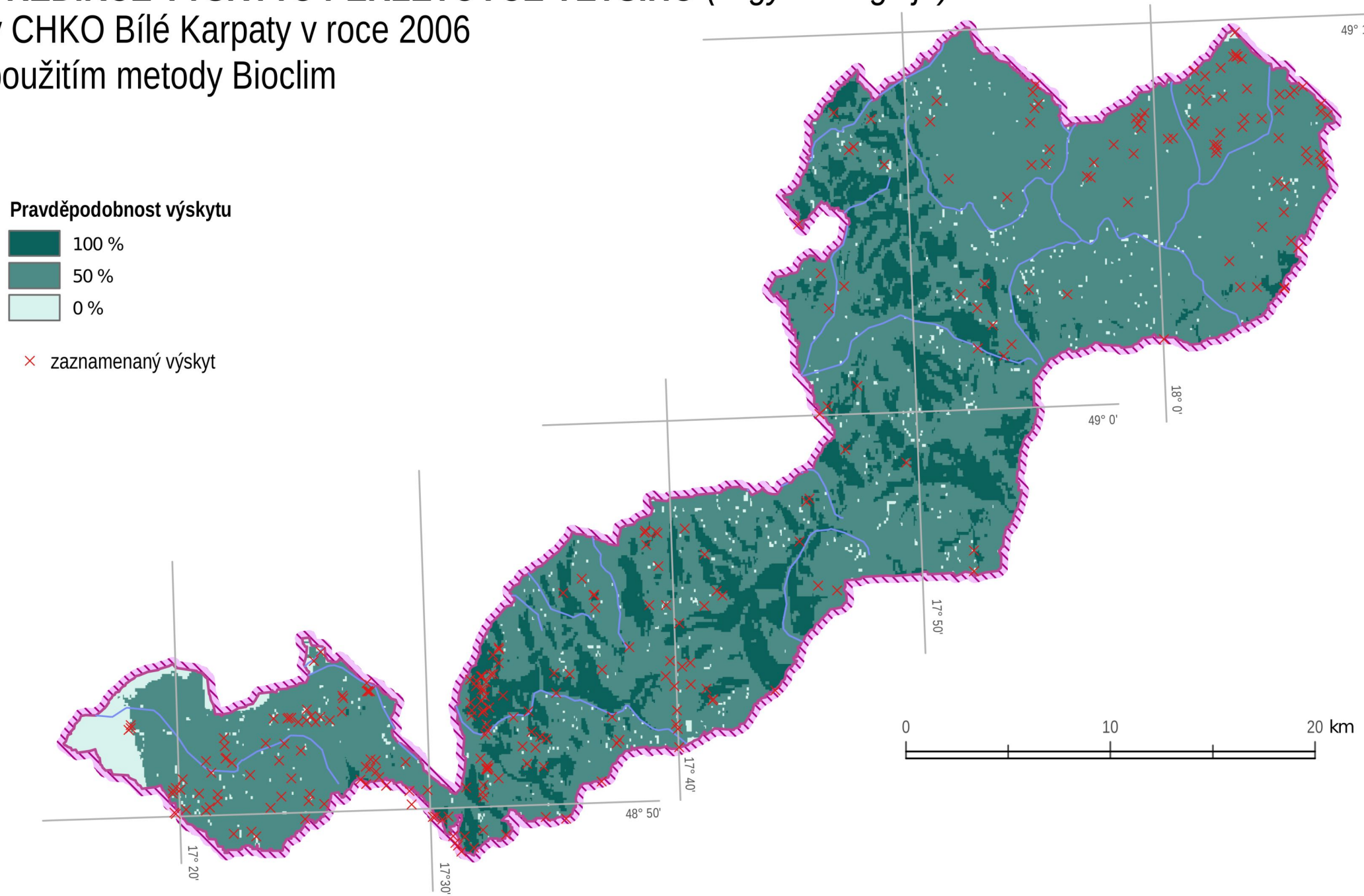
[1] 1 – nízká, 2 – střední, 3 – vysoká

* nebylo zjištěno

Příloha 2

PREDIKCE VÝSKYTU PERLEŤOVCE VĚTŠÍHO *(Argynnis aglaja)*

v CHKO Bílé Karpaty v roce 2006
použitím metody Bioclim



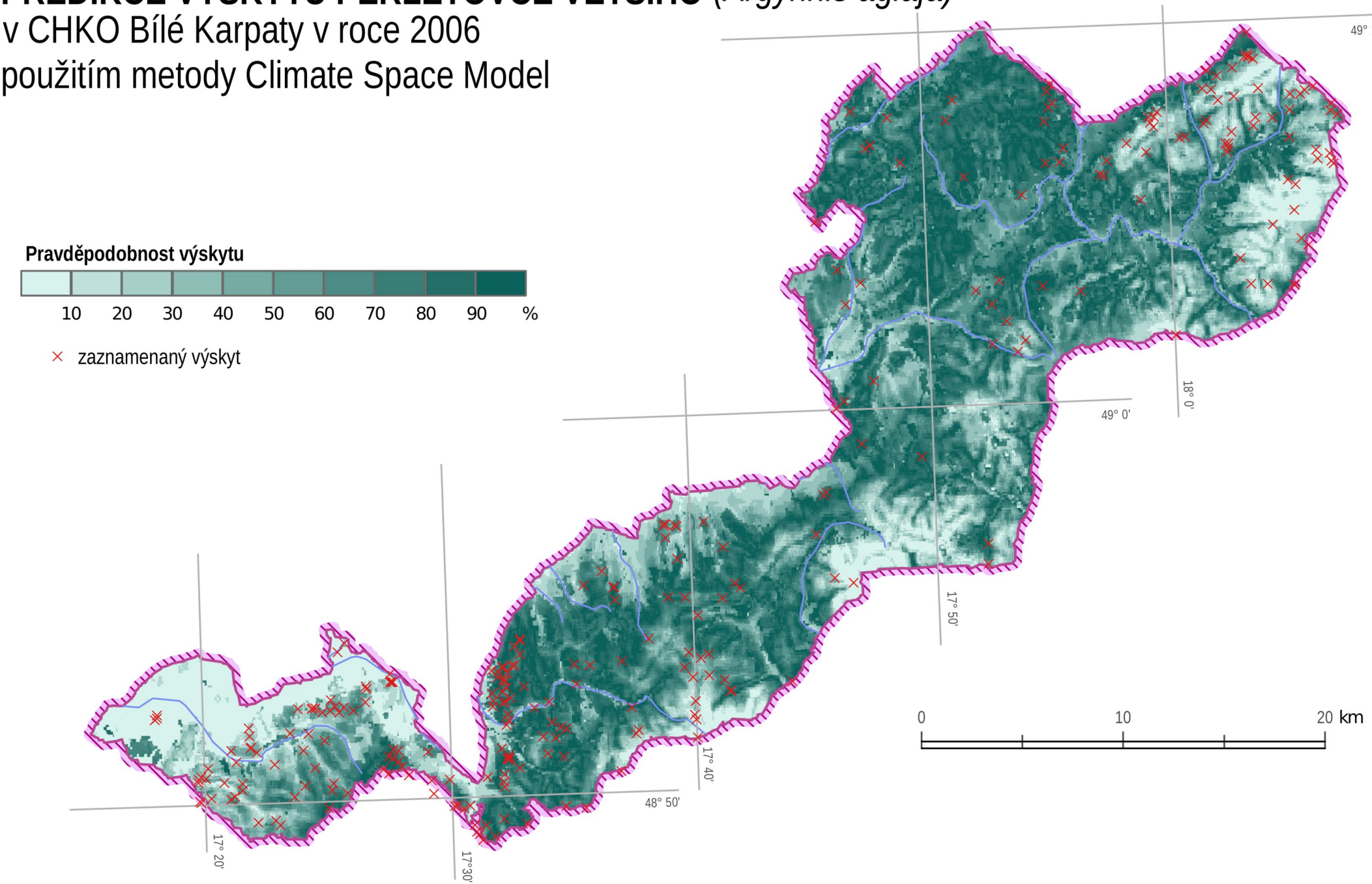
Eliška VLČKOVÁ
Olomouc 2016

Příloha 3

PREDIKCE VÝSKYTU PERLEŤOVCE VĚTŠÍHO *(Argynnis aglaja)*

v CHKO Bílé Karpaty v roce 2006

použitím metody Climate Space Model

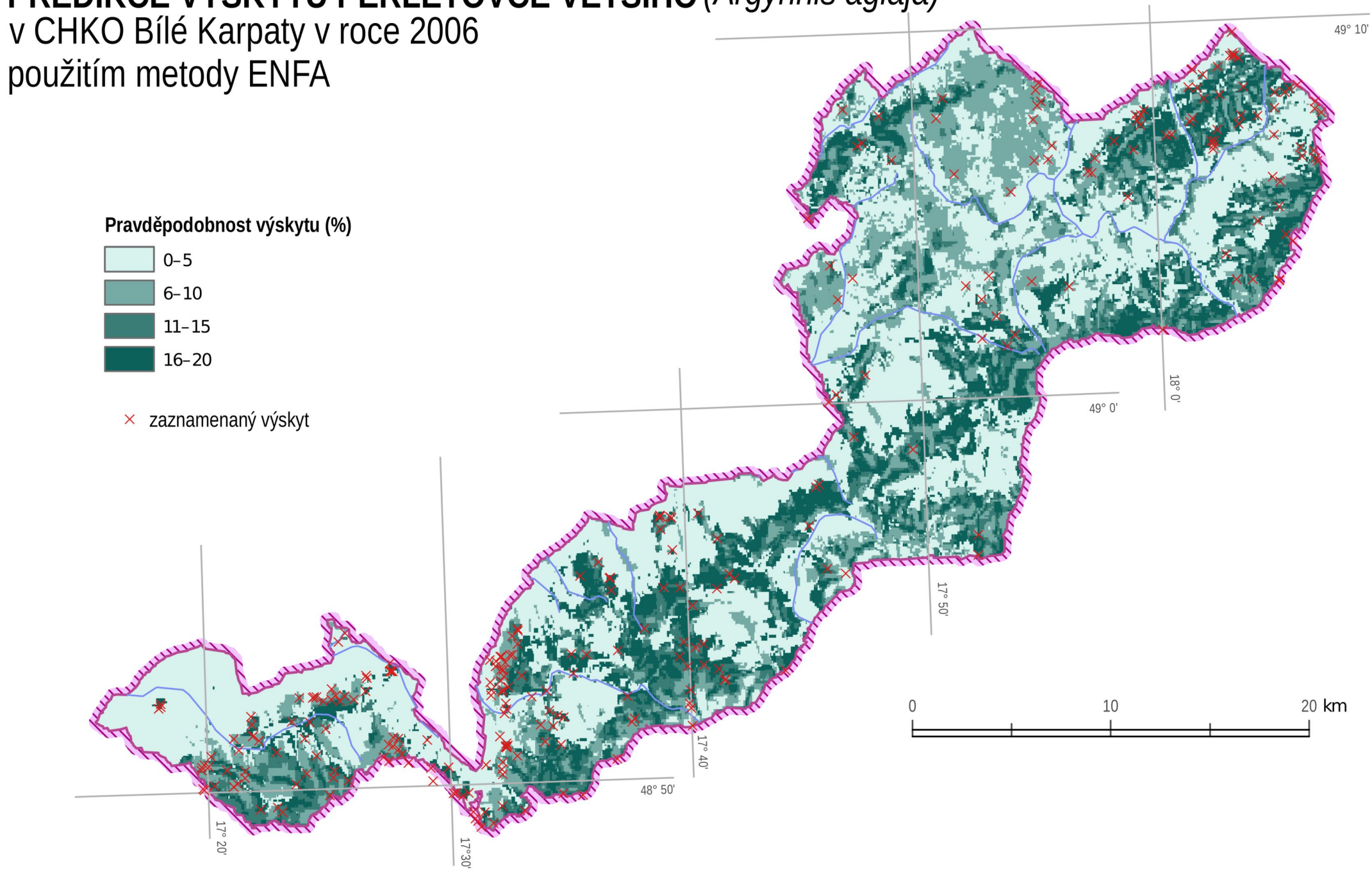


Eliška VLČKOVÁ
Olomouc 2016

Příloha 4

PREDIKCE VÝSKYTU PERLEŤOVCE VĚTŠÍHO *(Argynnis aglaja)*

v CHKO Bílé Karpaty v roce 2006
použitím metody ENFA



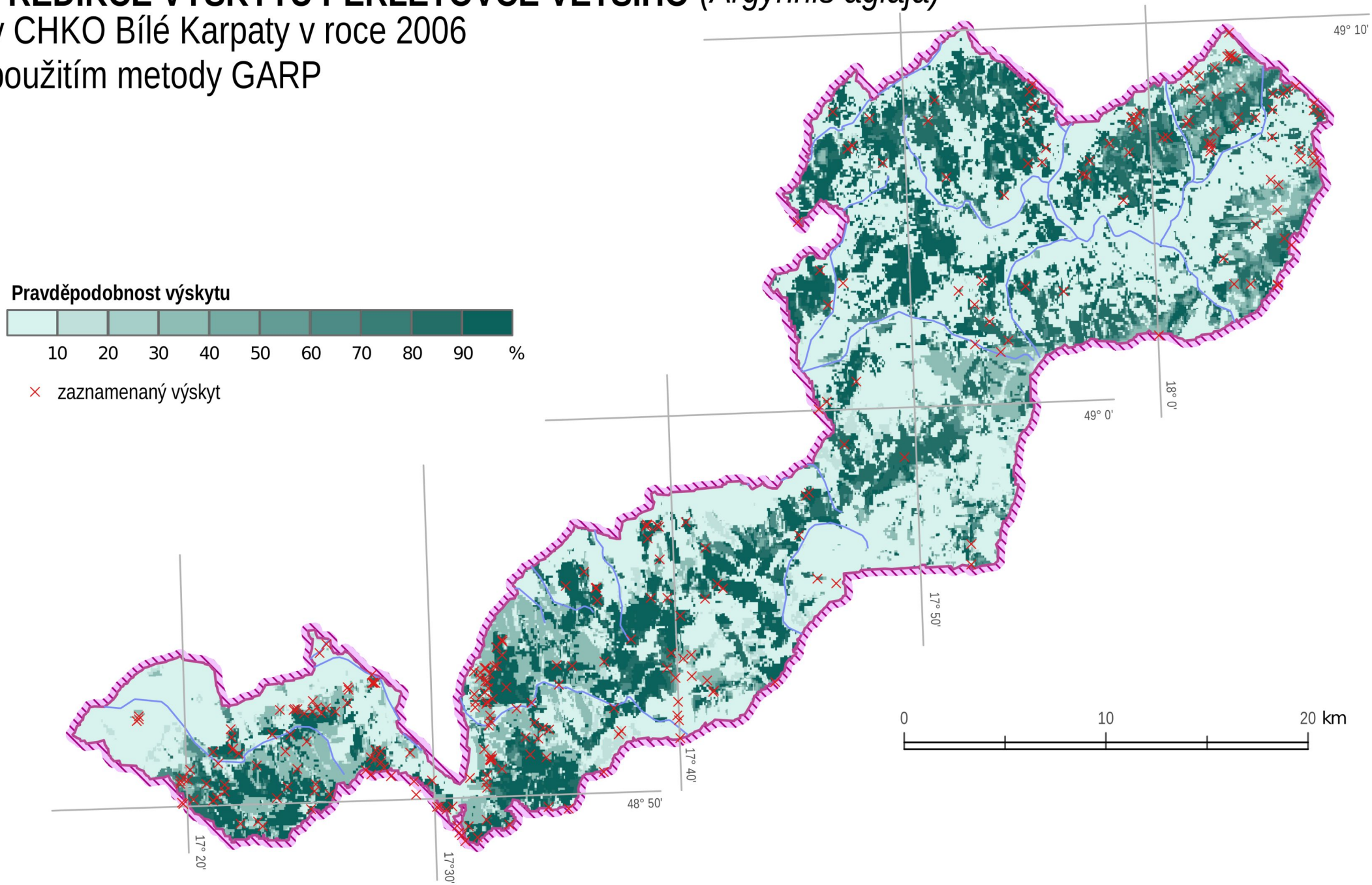
Eliška VLČKOVÁ
Olomouc 2016

Příloha 5

PREDIKCE VÝSKYTU PERLEŤOVCE VĚTŠÍHO *(Argynnis aglaja)*

v CHKO Bílé Karpaty v roce 2006

použitím metody GARP

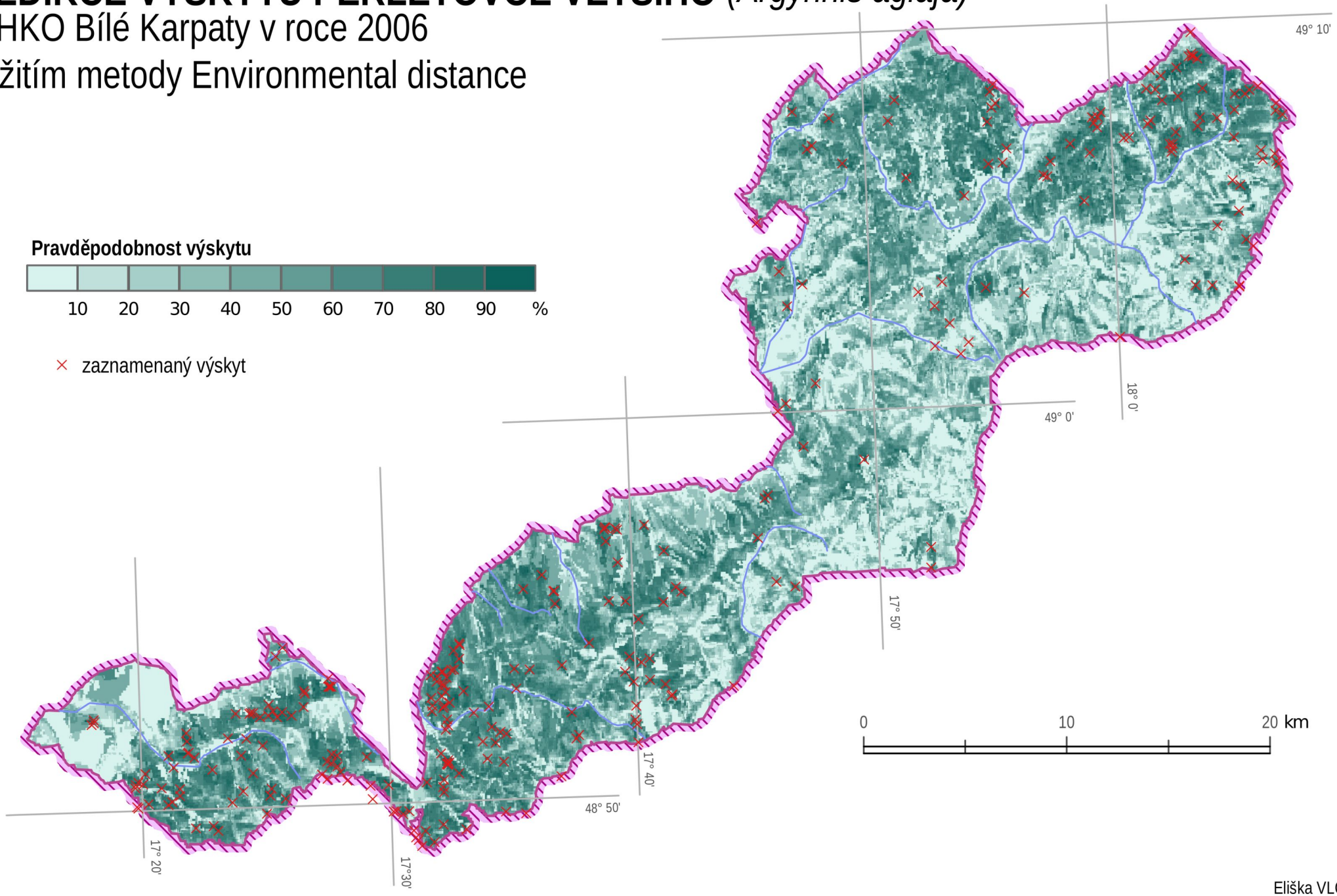


Eliška VLČKOVÁ
Olomouc 2016

PREDIKCE VÝSKYTU PERLEŤOVCE VĚTŠÍHO *(Argynnis aglaja)*

v CHKO Bílé Karpaty v roce 2006

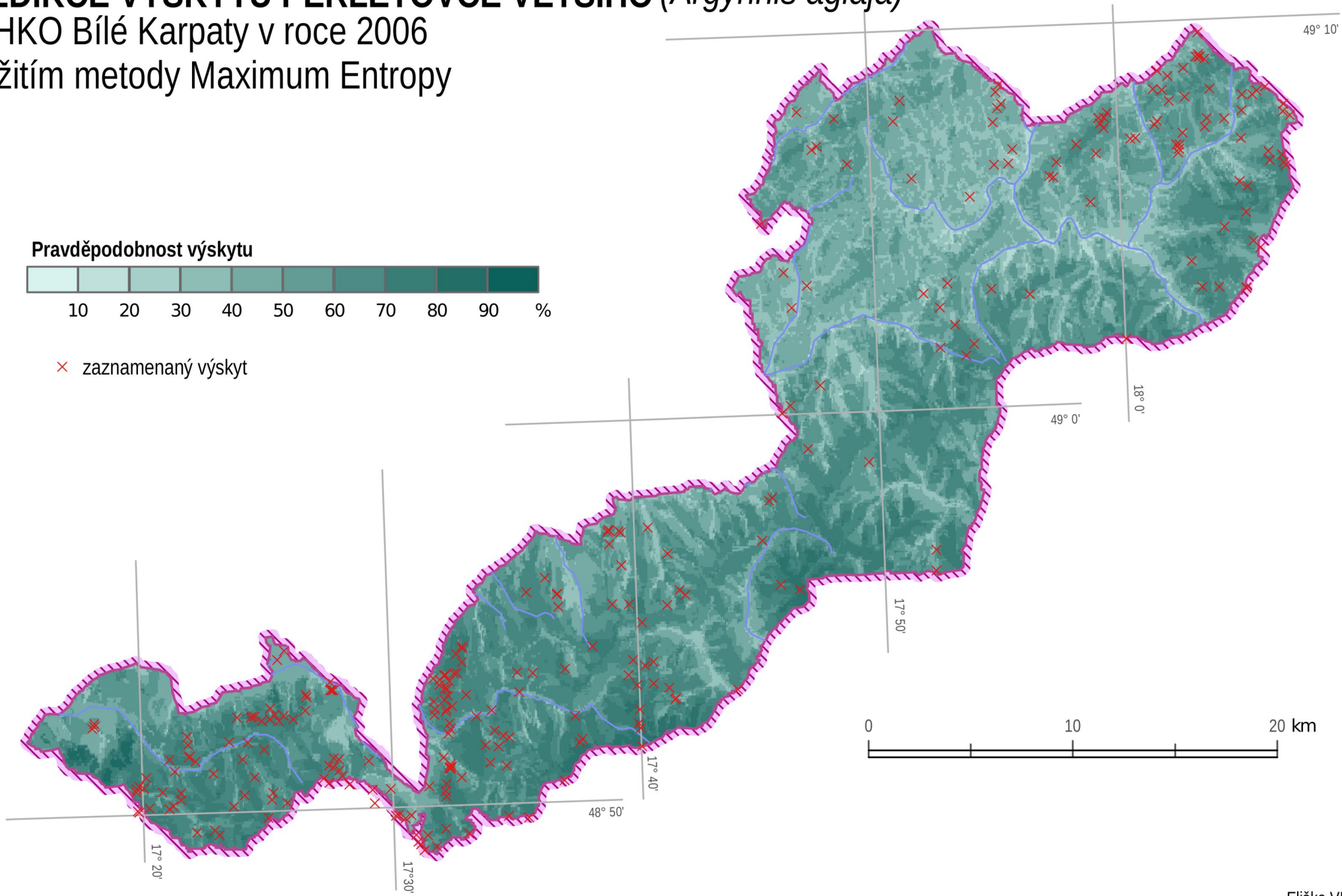
použitím metody Environmental distance



Eliška VLČKOVÁ
Olomouc 2016

PREDIKCE VÝSKYTU PERLEŤOVCE VĚTŠÍHO *(Argynnis aglaja)*

v CHKO Bílé Karpaty v roce 2006
použitím metody Maximum Entropy

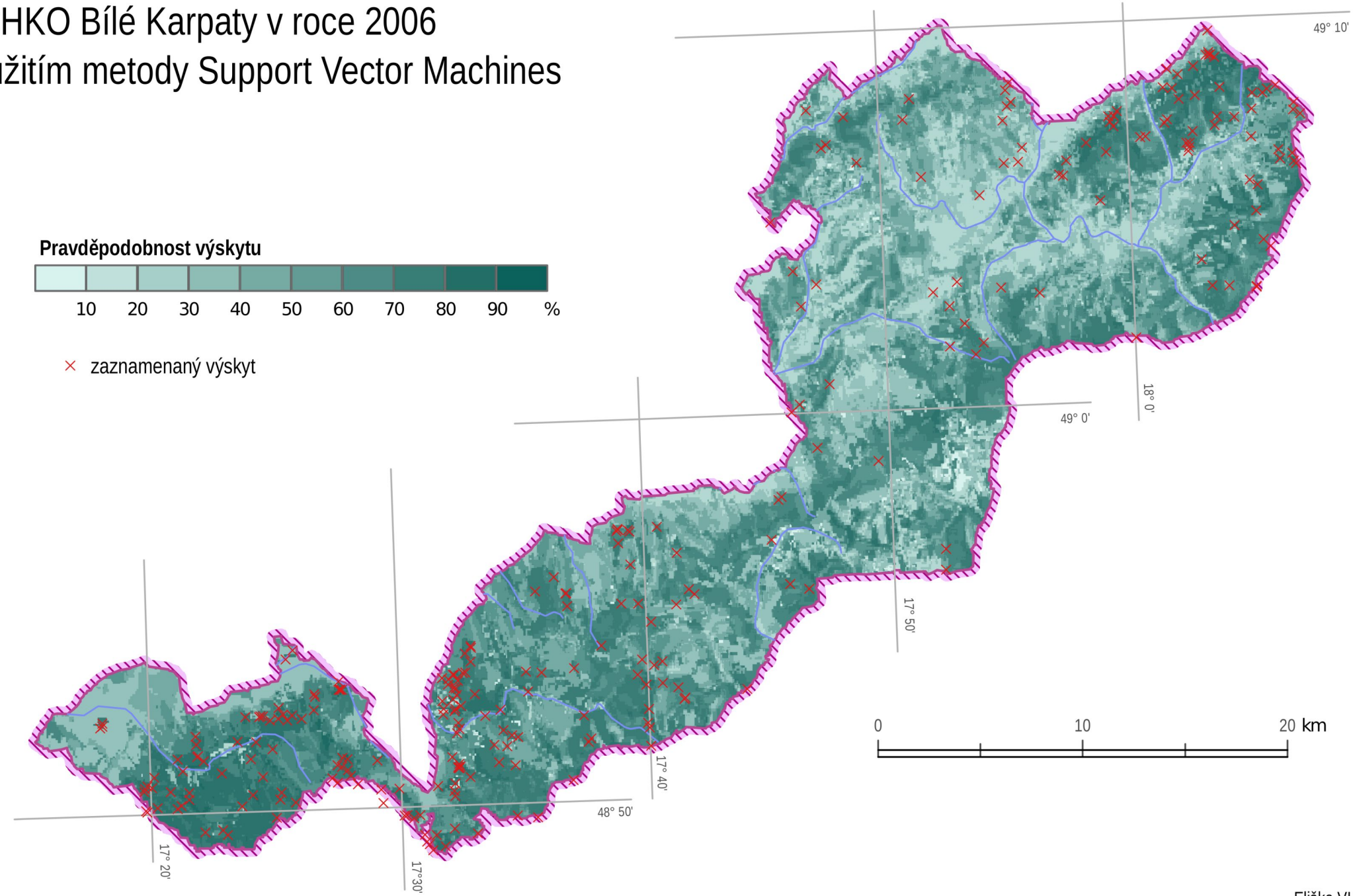


Eliška VLČKOVÁ
Olomouc 2016

PREDIKCE VÝSKYTU PERLEŤOVCE VĚTŠÍHO *(Argynnis aglaja)*

v CHKO Bílé Karpaty v roce 2006

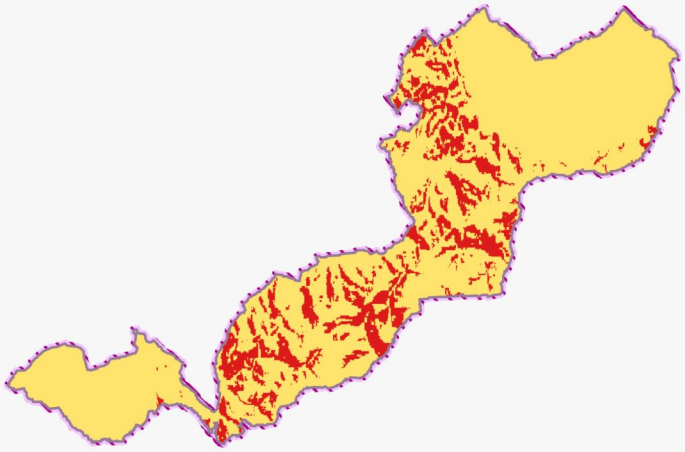
použitím metody Support Vector Machines



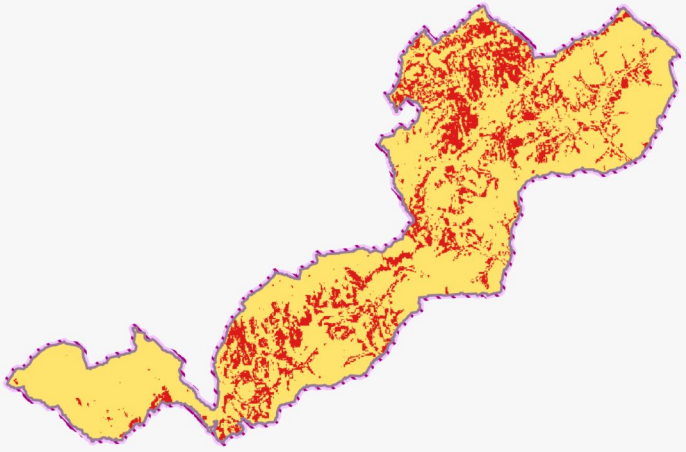
Eliška VLČKOVÁ
Olomouc 2016

VIZUÁLNÍ SROVNÁNÍ METOD PRO PREDIKCI PERLEŤOVCE VĚTŠÍHO

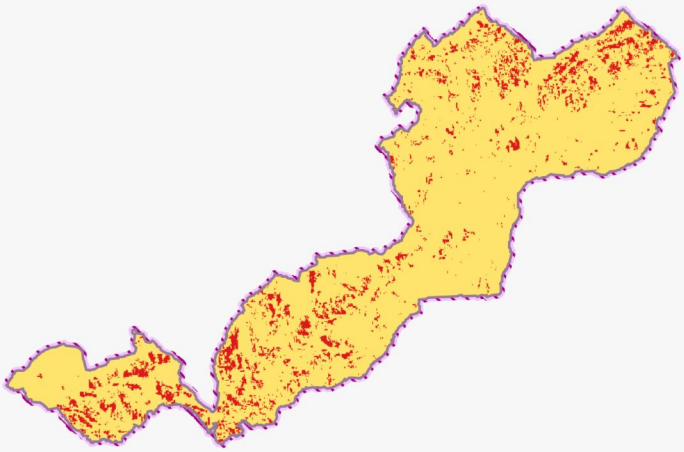
Bioclim



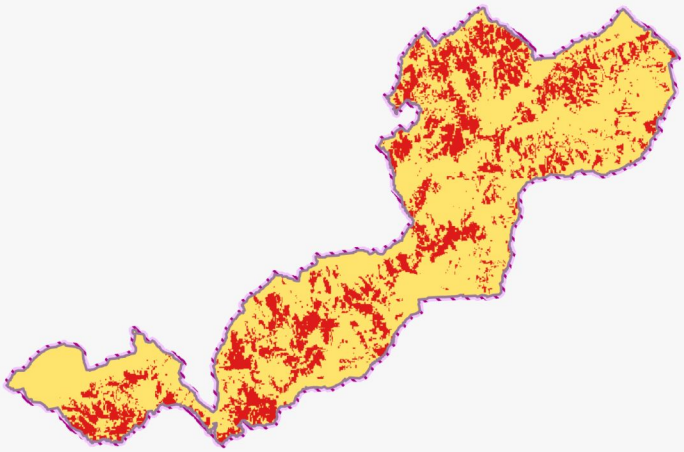
Climate Space Model



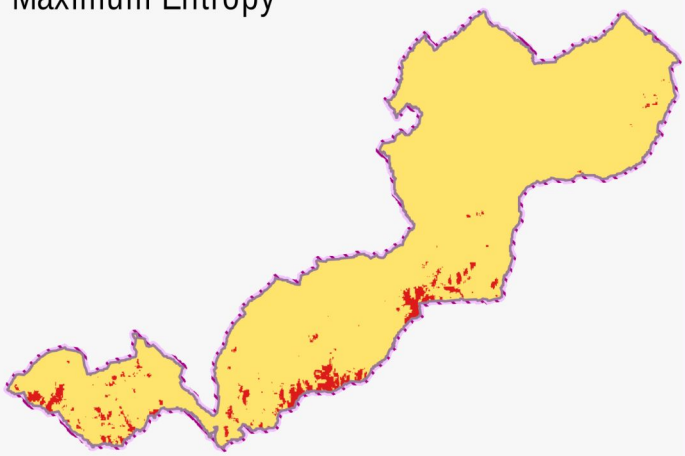
Environmental Distance



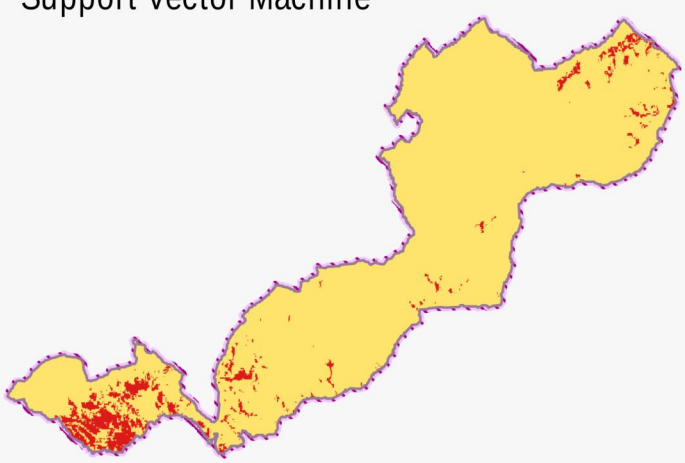
GARP with Best Subsets



Maximum Entropy



Support Vector Machine



Pravděpodobnost výskytu (%)

0–79

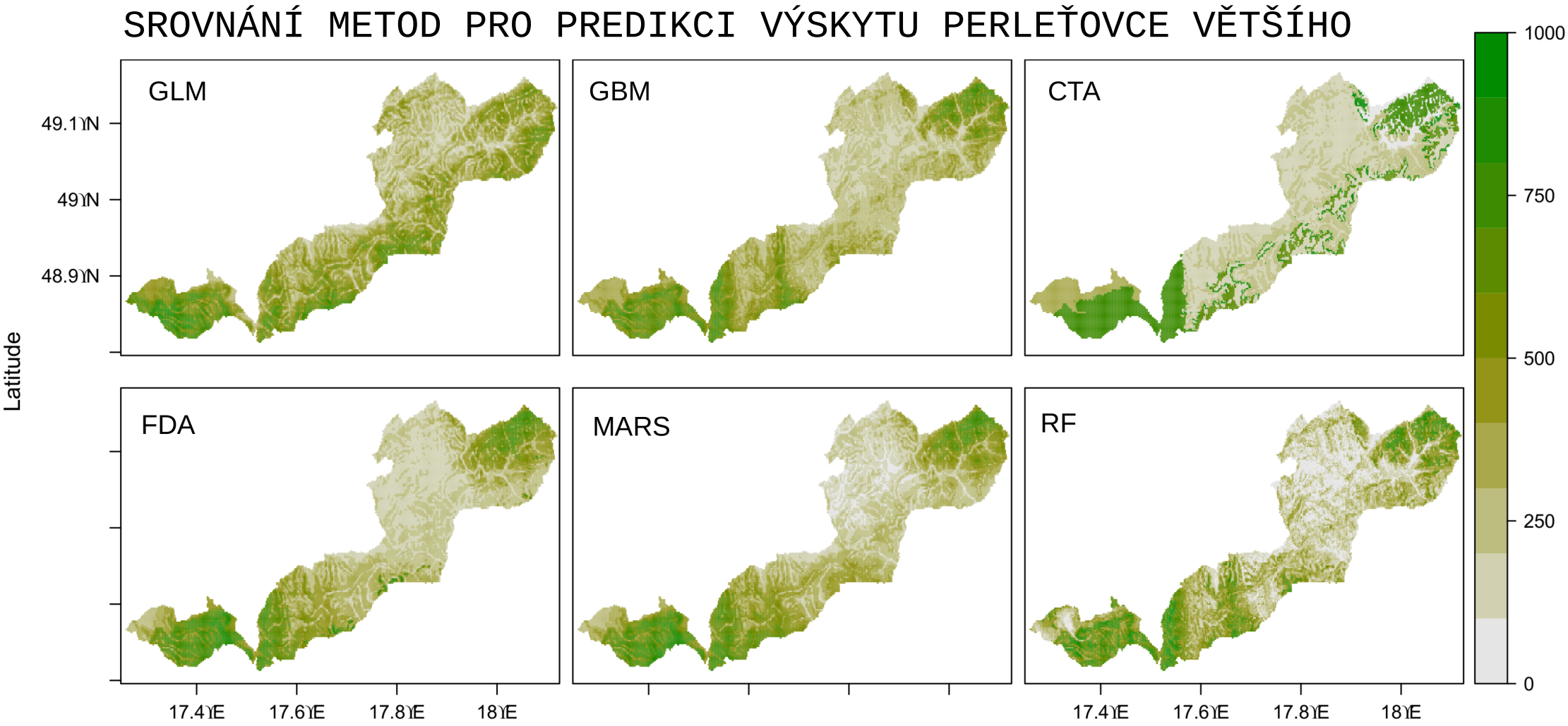
80–100

0 10 20 km

příloha k bakalářské práci s názvem
Srovnání výpočetních algoritmů pro prostorovou distribuci druhů
území: CHKO Bílé Karpaty

Eliška VLČKOVÁ
Olomouc 2016

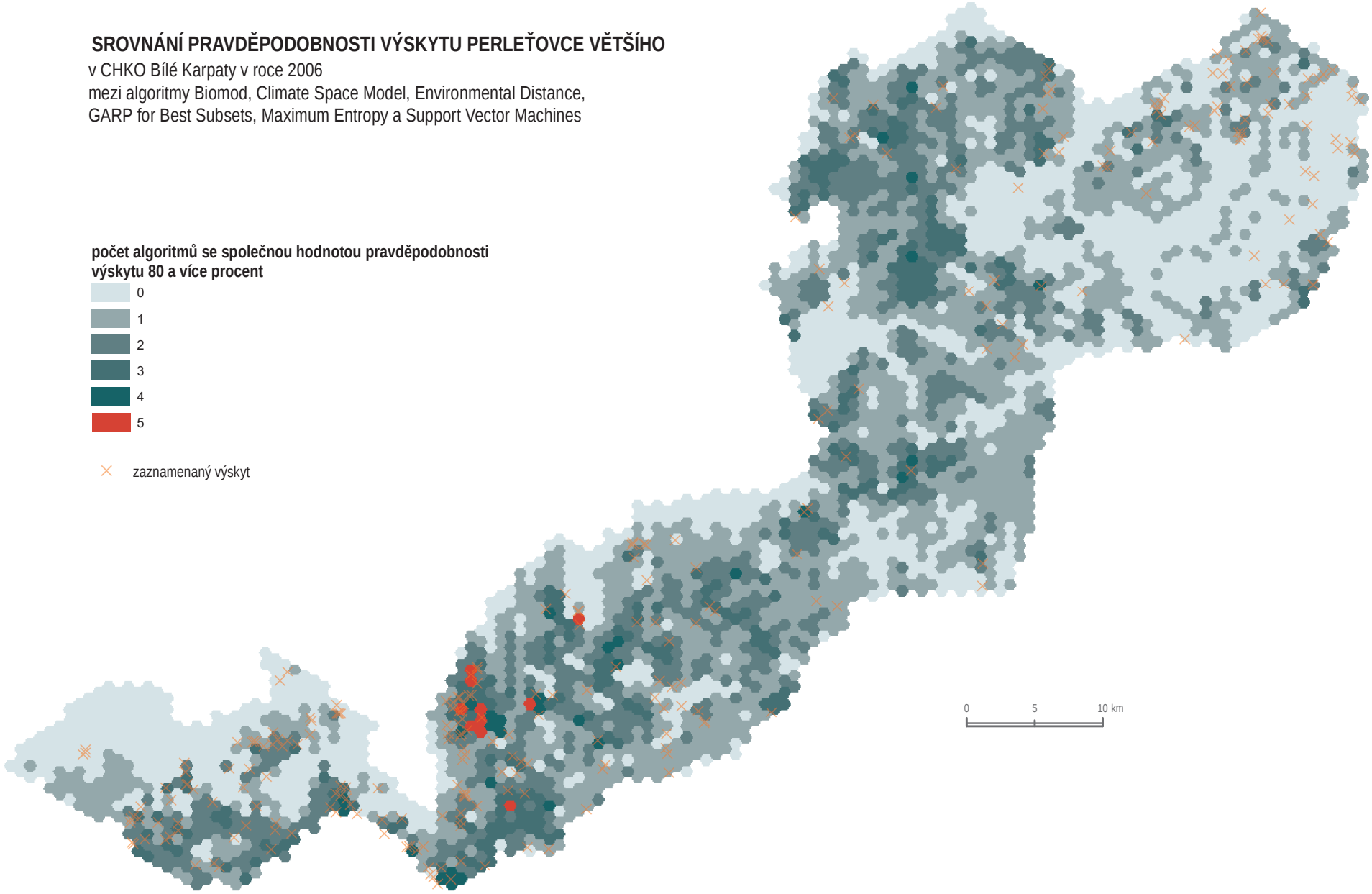
Příloha č. 10


SROVNÁNÍ METOD PRO PREDIKCI VÝSKYTU PERLEŤOVCE VĚTŠÍHO
GLM
GBM
CTA
FDA
MARS
RF
Latitude
49.1°N
49°N
48.9°N
17.4°E
17.6°E
17.8°E
18°E
1000
750
500
250
0

Příloha č. 11

SROVNÁNÍ PRAVDĚPODOBNOSTI VÝSKYTU PERLEŤOVCE VĚTŠÍHO

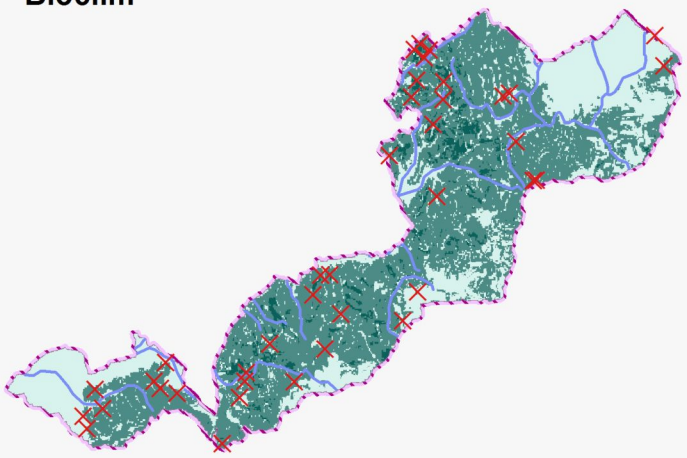
v CHKO Bílé Karpaty v roce 2006
mezi algoritmy Biomod, Climate Space Model, Environmental Distance,
GARP for Best Subsets, Maximum Entropy a Support Vector Machines



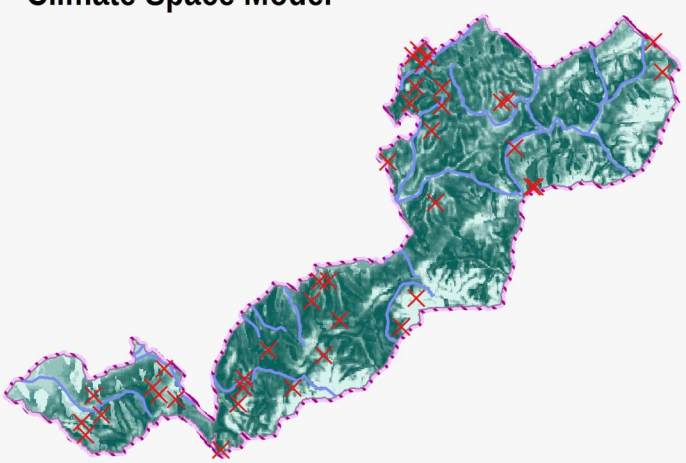
Příloha 12

SROVNÁNÍ METOD PRO PREDIKCI VÝSKYTU BABOČKY JILMOVÉ

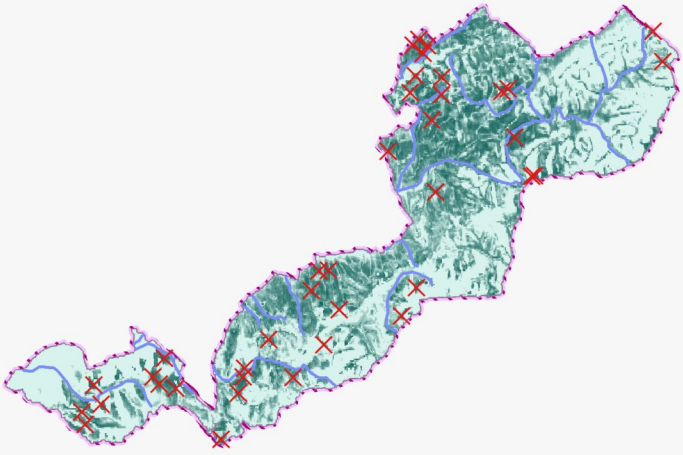
Bioclim



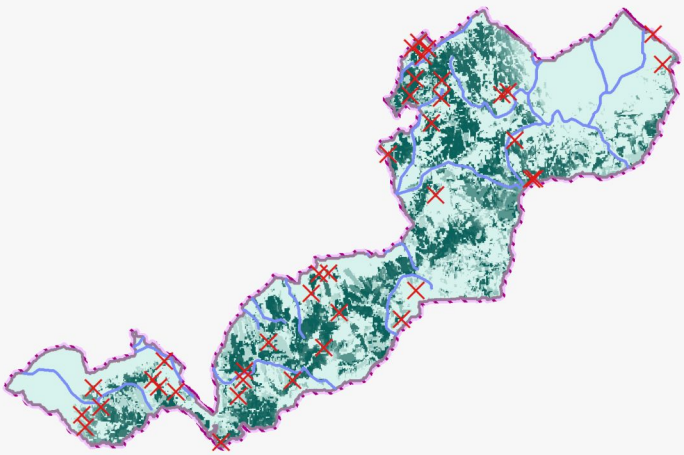
Climate Space Model



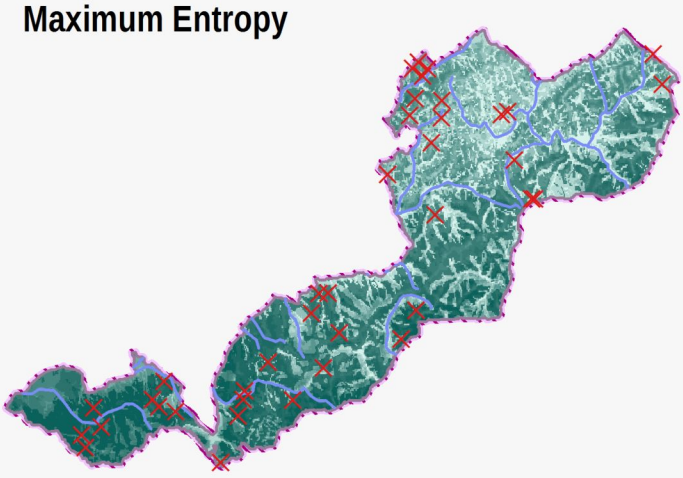
Environmental Distance



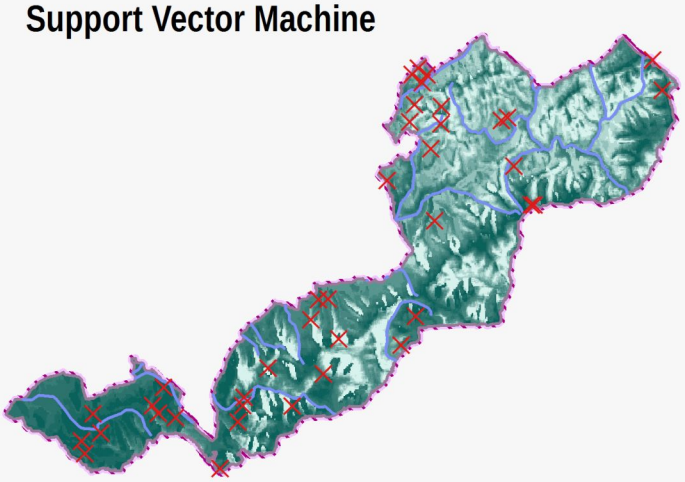
GARP with Best Subsets



Maximum Entropy



Support Vector Machine



Pravděpodobnost výskytu

10 20 30 40 50 60 70 80 90 %

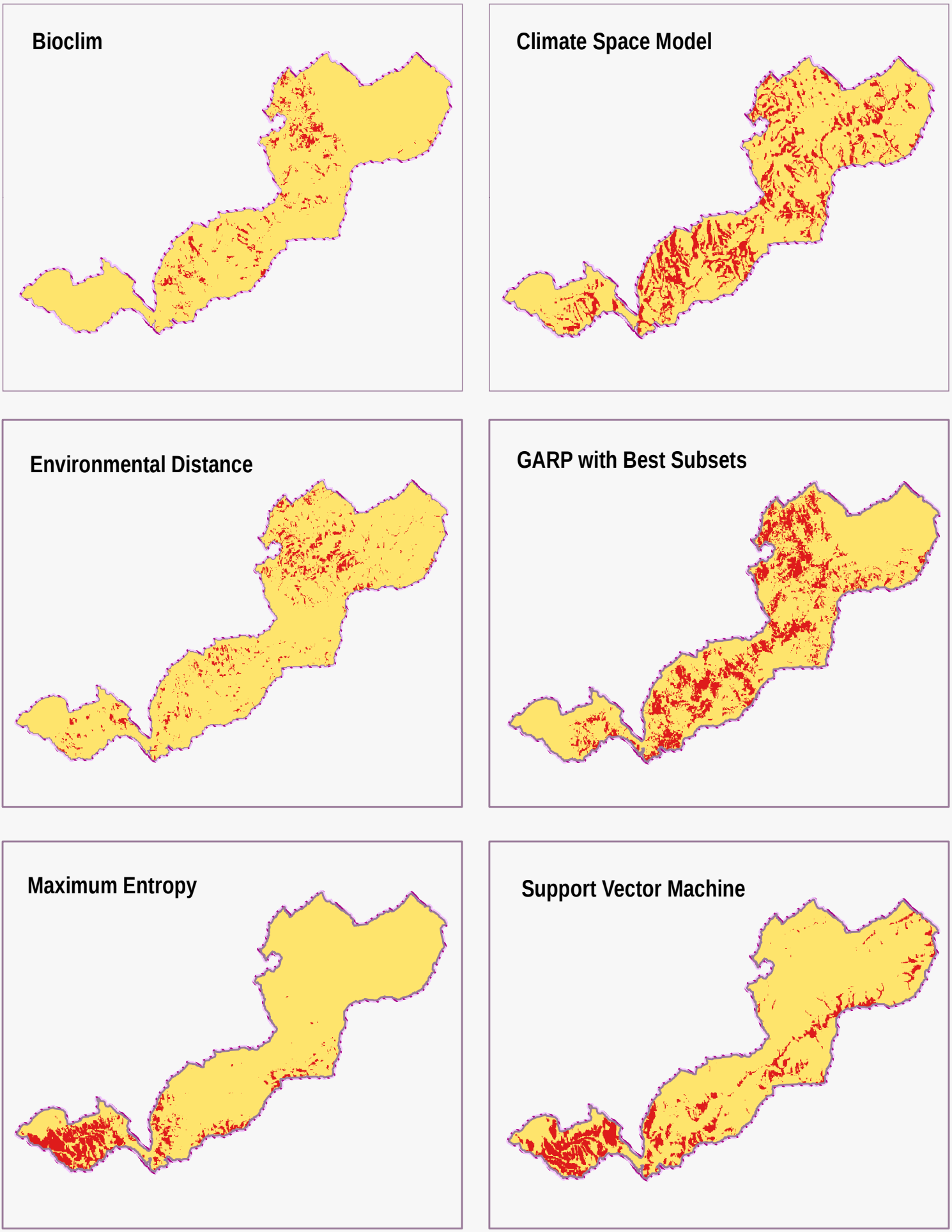
0 10 20 km

území: CHKO Bílé Karpaty

Eliška VLČKOVÁ
Olomouc 2016

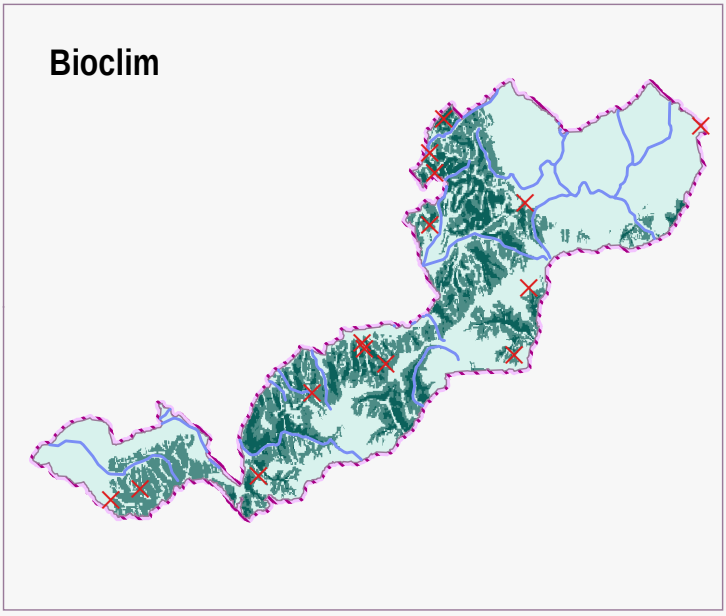
Příloha 13

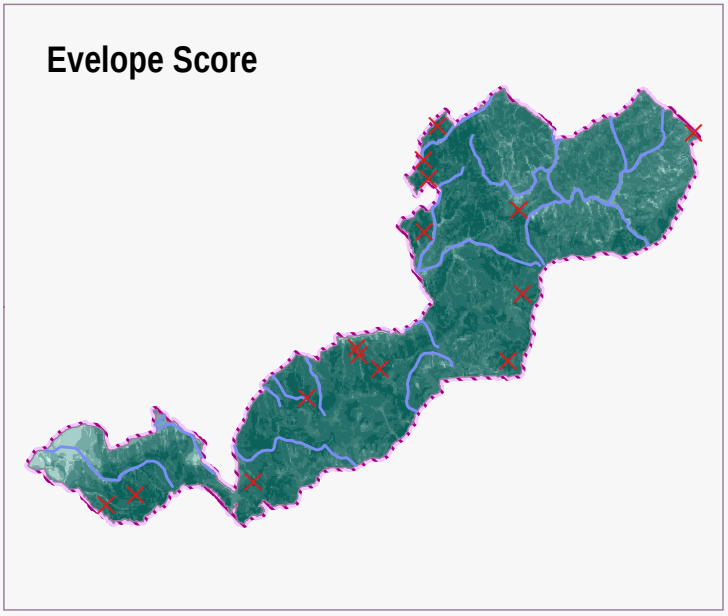
SROVNÁNÍ METOD PRO PREDIKCI VÝSKYTU BABOČKY JILMOVÉ

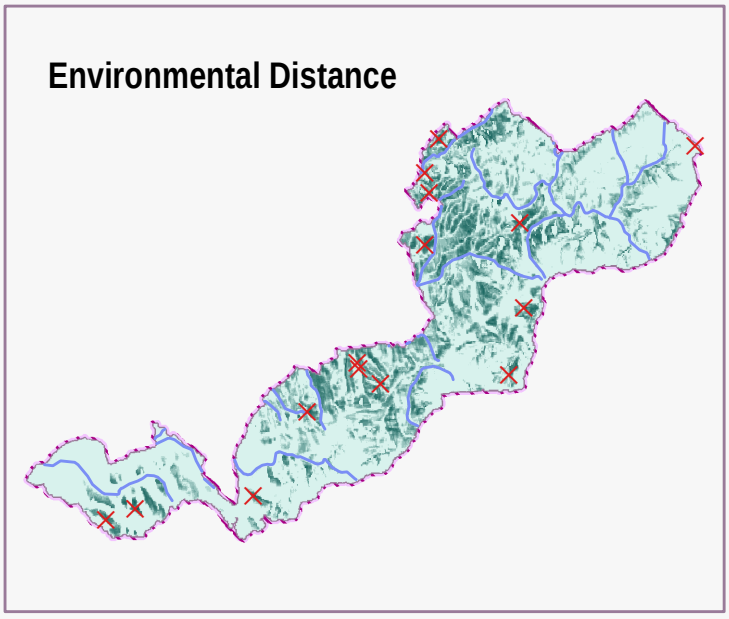


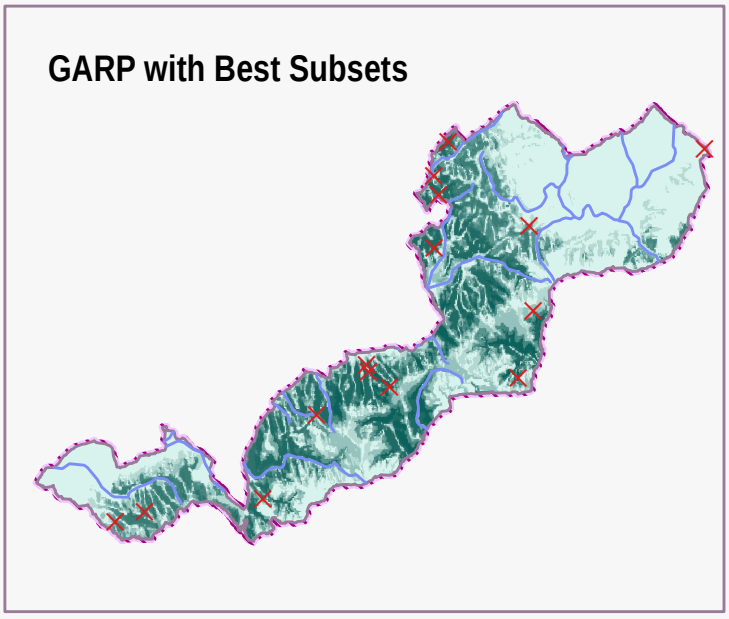
Příloha 14

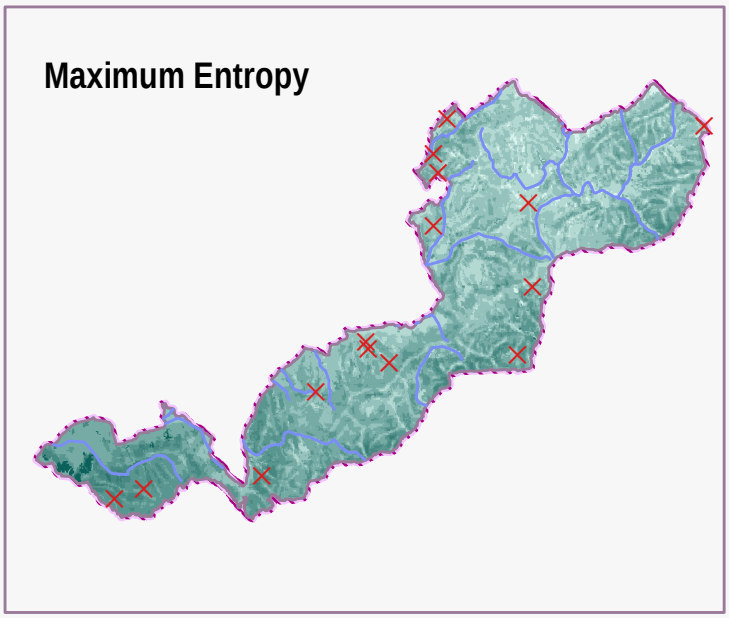
SROVNÁNÍ METOD PRO PREDIKCI VÝSKYTU BĚLOPÁSKA TOPOLOVÉHO

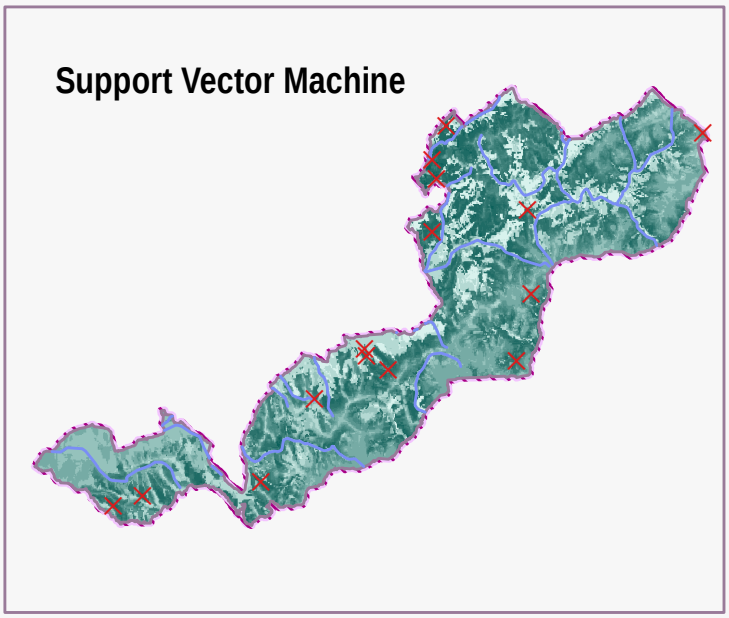












Pravděpodobnost výskytu

10 20 30 40 50 60 70 80 90 %

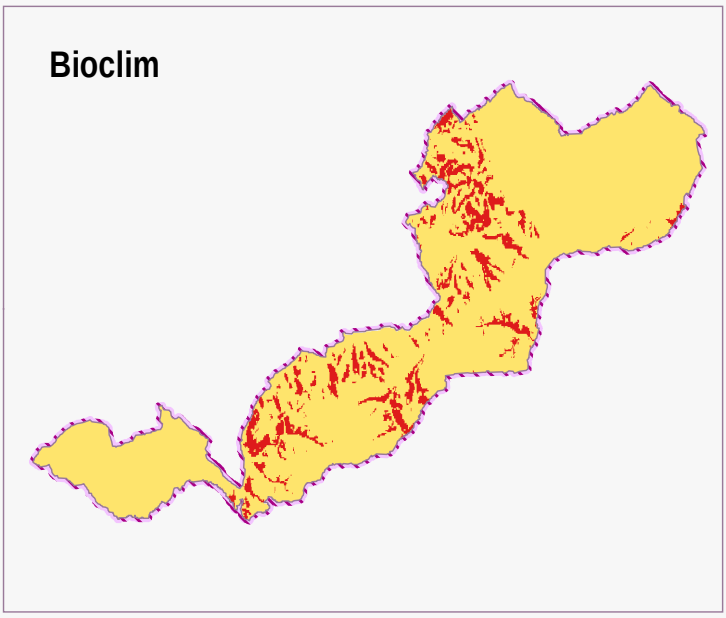
území: CHKO Bílé Karpaty

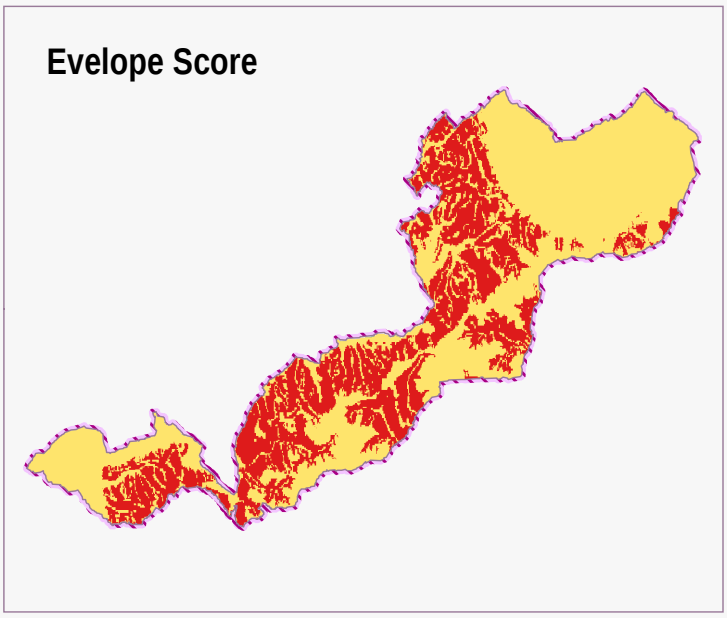
0 10 20 km

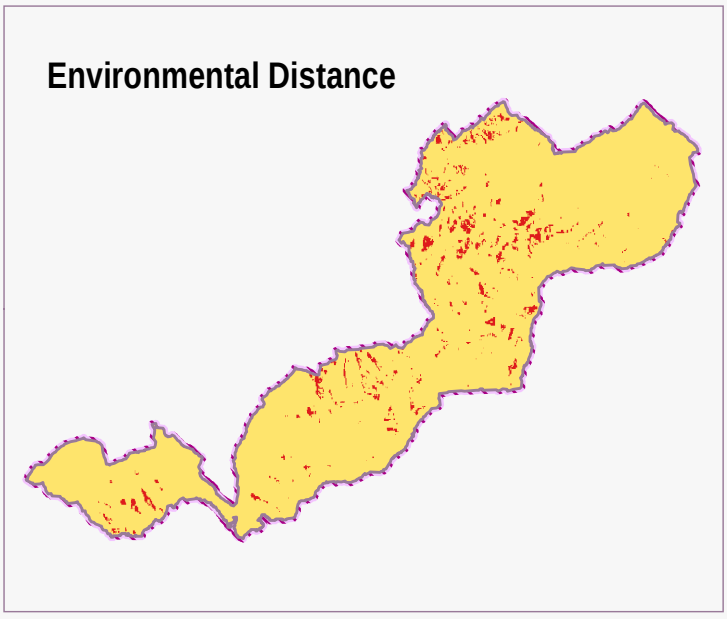
Eliška VLČKOVÁ
Olomouc 2016

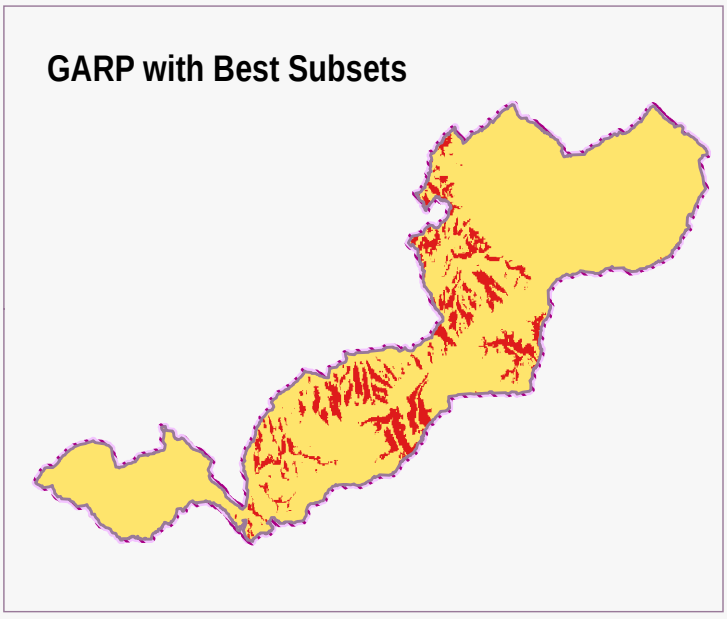
Příloha 15

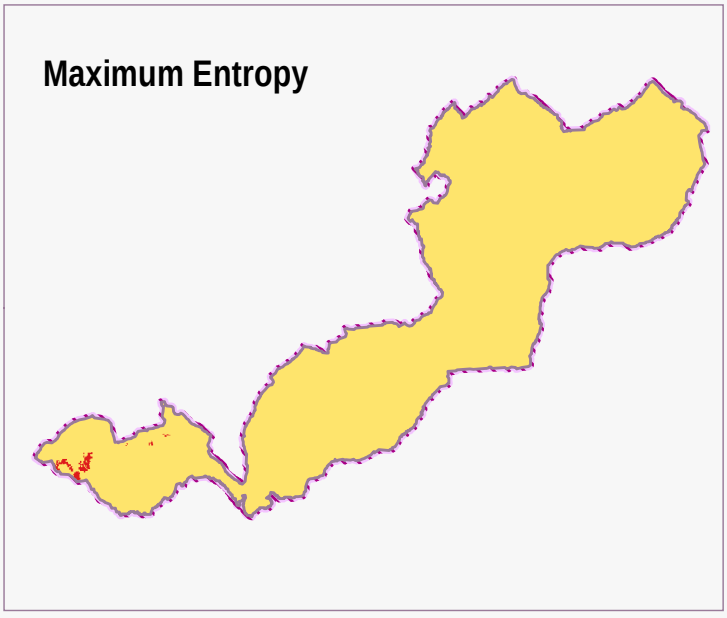
SROVNÁNÍ METOD PRO PREDIKCI VÝSKYTU BĚLOPÁSKA TOPOLOVÉHO

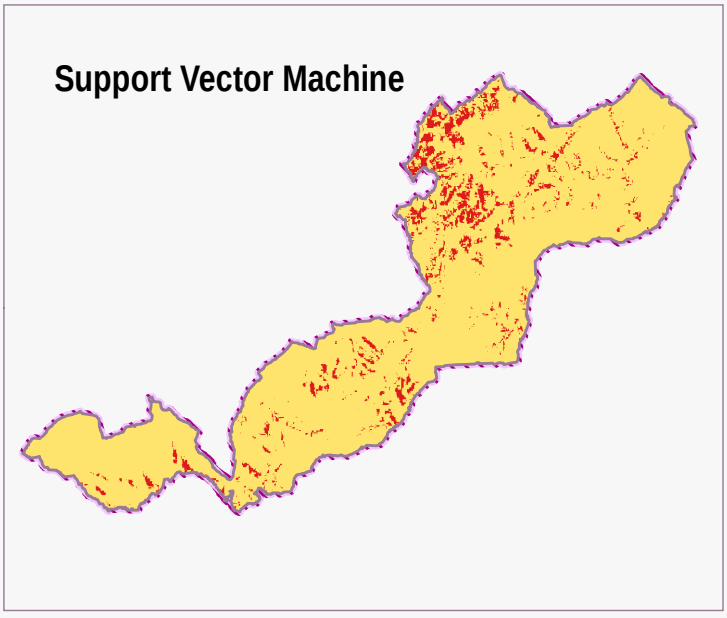












Pravděpodobnost výskytu (%)

0–79

80–100

0 10 20 km

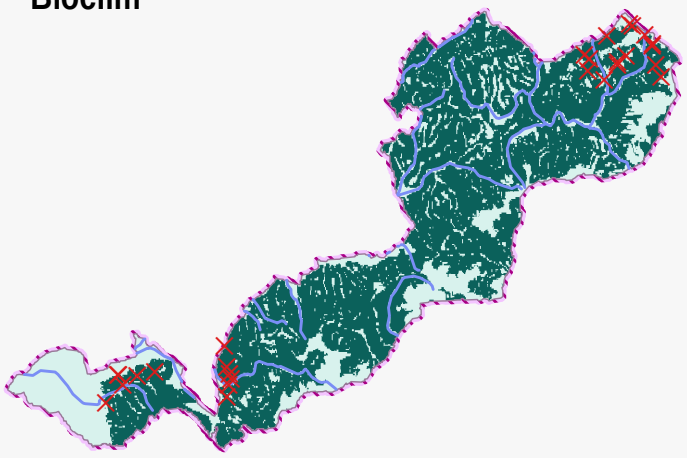
území: CHKO Bílé Karpaty

Eliška VLČKOVÁ
Olomouc 2016

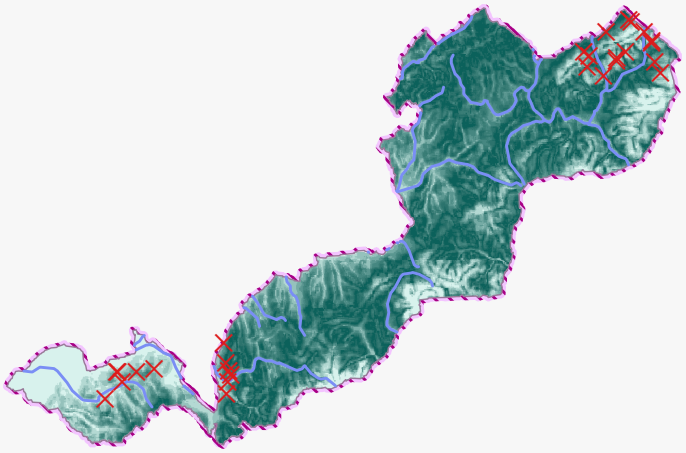
Příloha 16

SROVNÁNÍ METOD PRO PREDIKCI VÝSKYTU MODRÁSKA HNĚDOSKVRNNÉHO

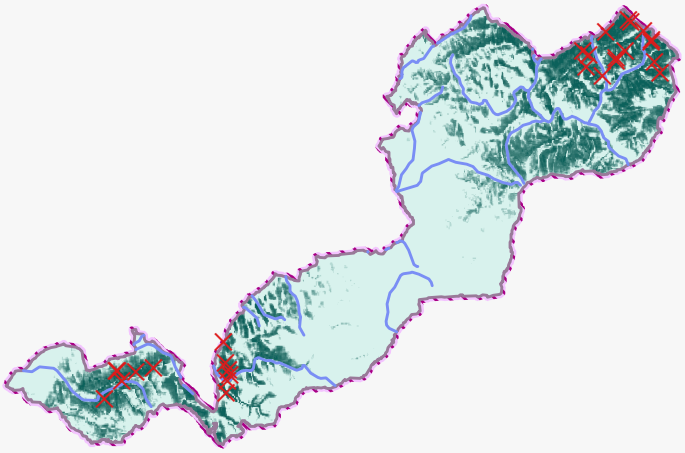
Bioclim



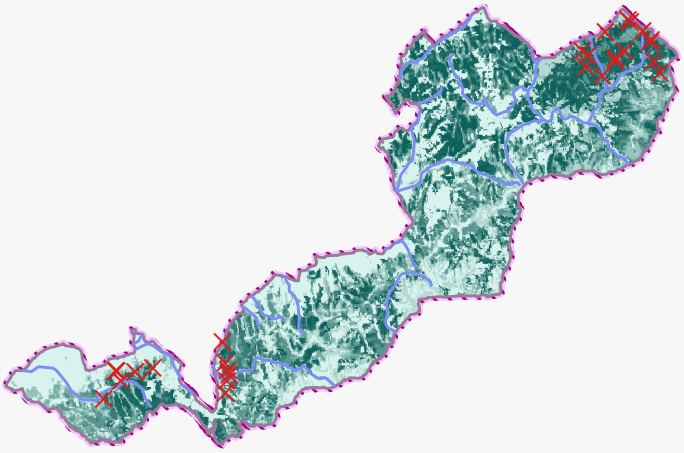
Climate Space Model



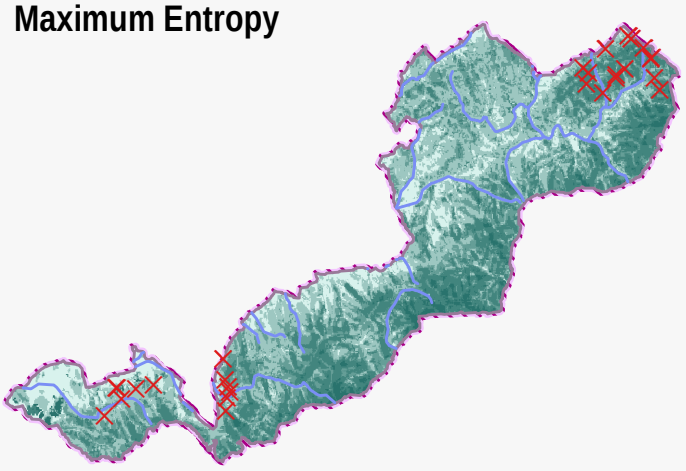
Environmental Distance



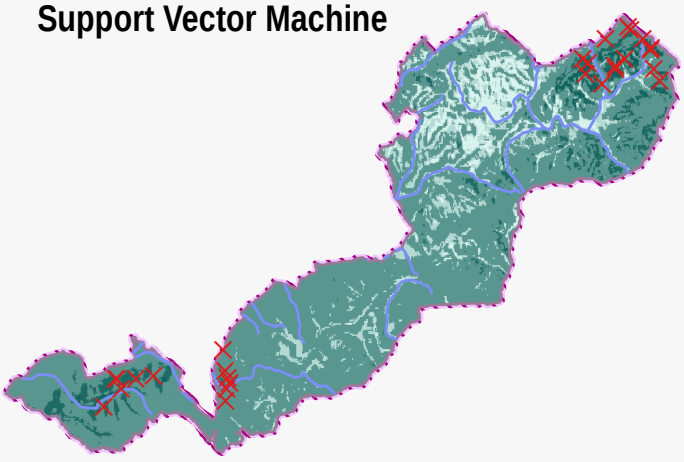
GARP with Best Subsets



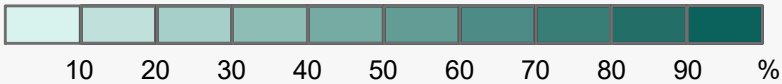
Maximum Entropy

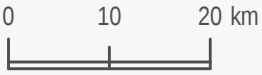


Support Vector Machine



Pravděpodobnost výskytu



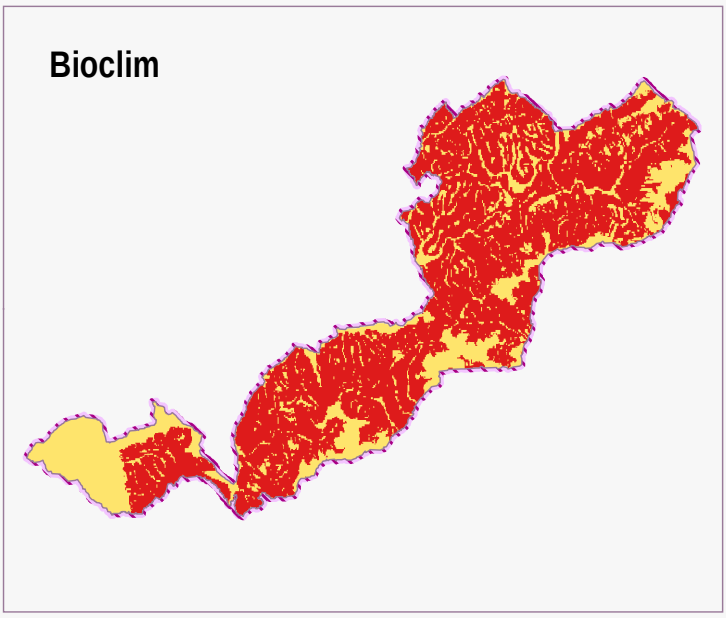


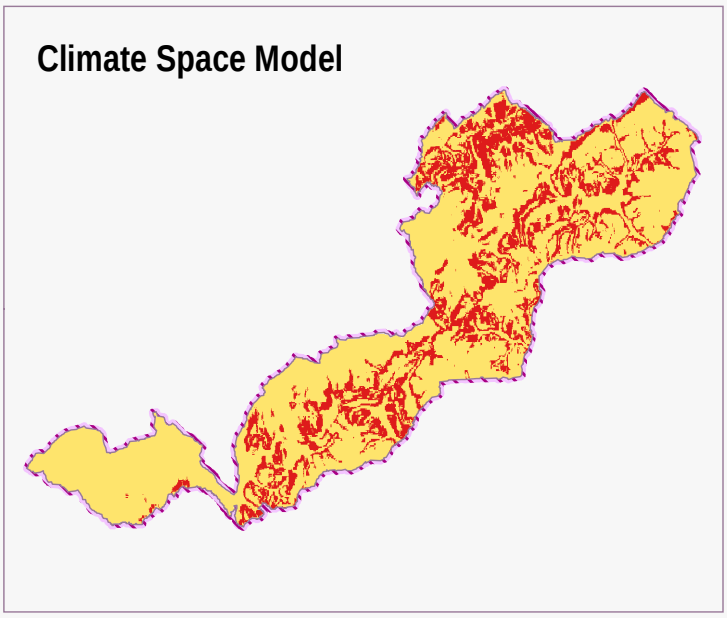
území: CHKO Bílé Karpaty

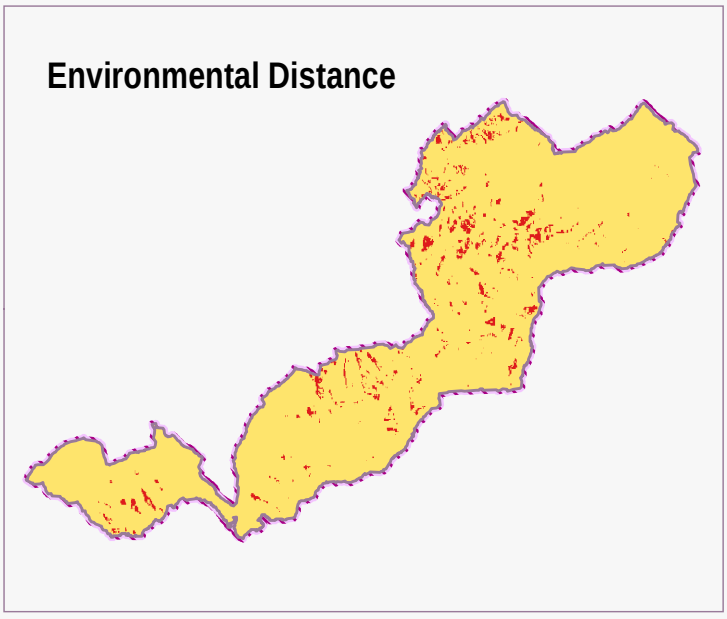
Eliška VLČKOVÁ
Olomouc 2016

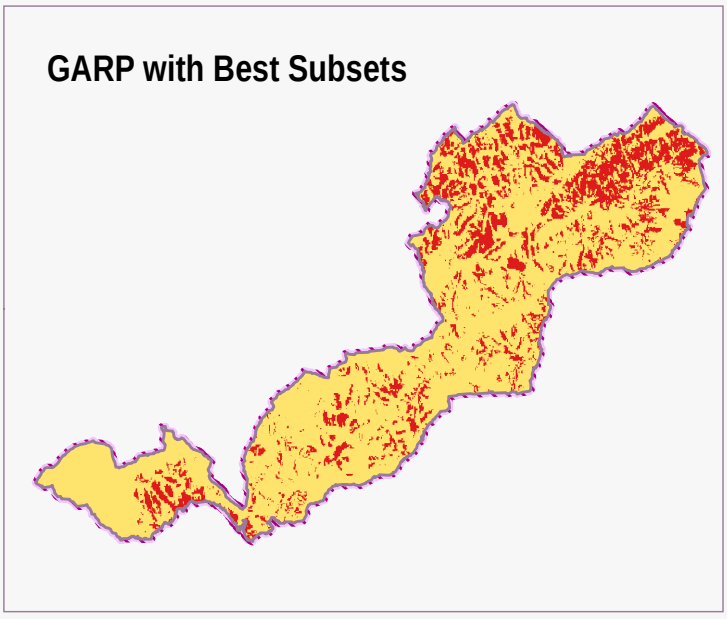
Příloha 17

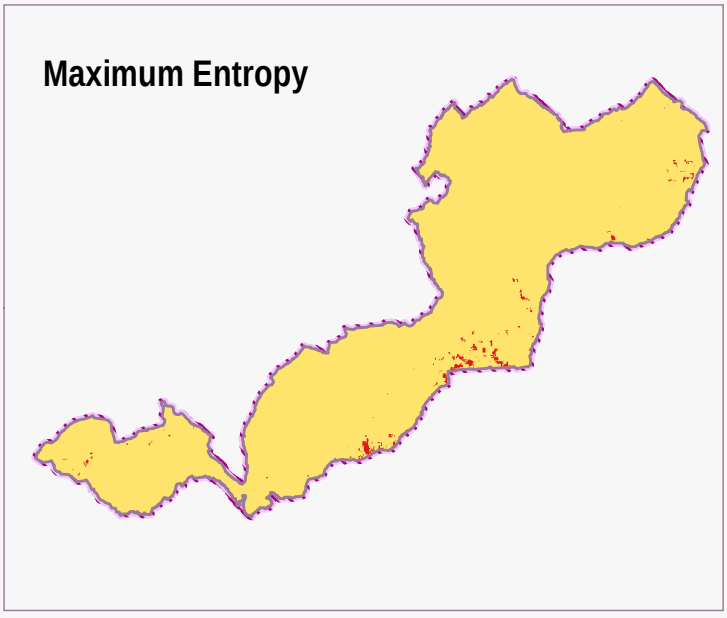
SROVNÁNÍ METOD PRO PREDIKCI VÝSKYTU MODRÁSKA HNĚDOSKVRNNÉHO

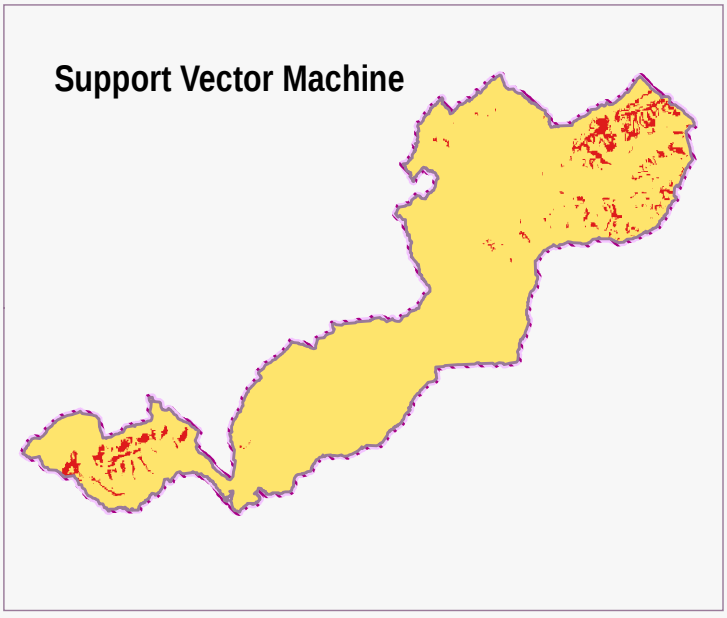












Pravděpodobnost výskytu (%)

0–79

80–100

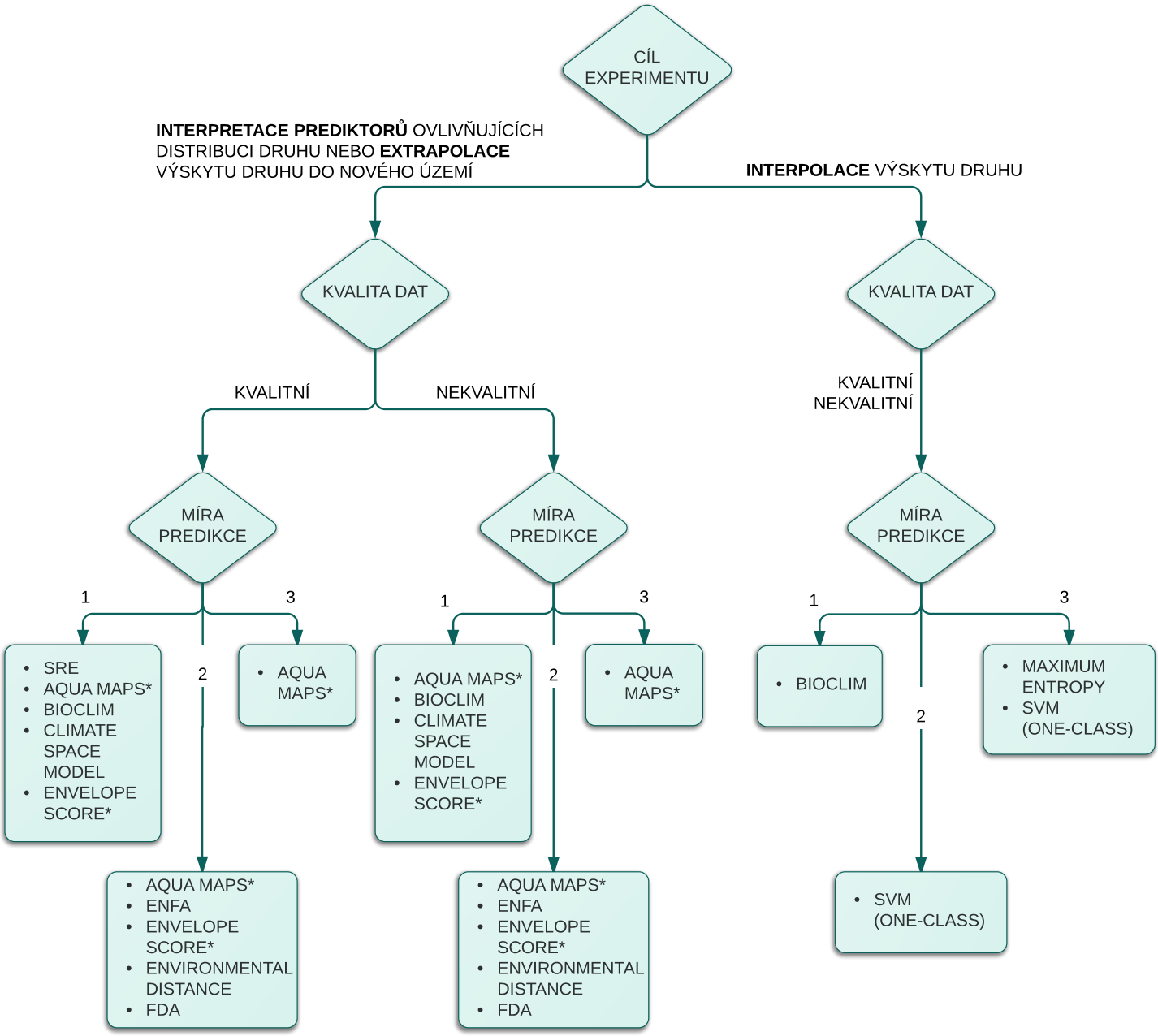
0 10 20 km

území: CHKO Bílé Karpaty

Eliška VLČKOVÁ
Olomouc 2016

ROZHODOVACÍ STROM PRO VÝBĚR ADEKVÁTNÍHO ALGORITMU

PRO *PRESENCE-ONLY* METODY



*NEBYLO ZJIŠTĚNO, ALGORITMUS BYL ZAHRNUT DO VŠECH KATEGORIÍ

ROZHODOVACÍ STROM PRO VÝBĚR ADEKVÁTNÍHO ALGORITMU

PRO *PRESENCE-ABSENCE* METODY