

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA APLIKOVANÉ GEOINFORMATIKY A UZEMNÍHO
PLÁNOVÁNÍ



VLIV ODLIŠNÉ KVALITY ENVIRONMENTÁLNÍCH DAT NA
MODELY DRUHOVÉ DISTRIBUCE

BAKALÁŘSKÁ PRÁCE

Vedoucí bakalářské práce: Ing. Lukáš Gábor

Autor: Kamil Matěj Brabec

2018

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Kamil Matěj Brabec

Územní technická a správní služba

Název práce

Vliv odlišné kvality environmentálních dat na modely druhové distribuce

Název anglicky

The Effect of Different Quality of Environmental Data on Species Distribution Models

Cíle práce

Cílem práce je s využitím virtuálního druhu otestovat, jak ovlivňuje odlišná kvalita environmentálních dat druhové distribuční modely.

Formulace dílčích cílů je úkolem autora.

Metodika

Druhové distribuční modely představují široce používaný nástroj v biogeografii, makroekologii a ochraně přírody, který se s postupným rozvojem stal důležitým prostředkem, využívaným například při určení lokalit potencionálně ohrožených invazními druhy nebo studování vlivu klimatických změn na biodiverzitu. S postupujícím rozvojem začalo být zřejmé, že jedním ze zásadních limitujících faktorů modelování druhové distribuce jsou vstupní data. Nicméně ověření vlivu kvality prostorových dat na modely, je s využitím reálných dat obtížné. Avšak pomocí virtuálního druhu lze snadno určit, jaký vliv mají například rozdílné metody sběru dat nebo jejich kvalita na výsledný model. Úkolem autora je zpracovat literární rešerši na téma druhových distribučních modelů, generování virtuálního druhu a vlivu odlišné kvality prostorových dat na modely druhové distribuce. V praktické části pak pomocí virtuálního druhu otestovat, do jaké míry ovlivňuje odlišná kvalita environmentálních dat výsledné modely.

Doporučený rozsah práce

30 stran

Klíčová slova

Formulace klíčových slov je úkolem autora

Doporučené zdroje informací

Fernández N. et al., Identifying breeding habitat for the Iberian lynx: Inferences from a fine-scale spatial analysis. *Ecol. Appl.* 2013, 1310-1324.

Meynard CH. N. & Kaplan M. D., The effect of a gradual response to the environment on species distribution modelling performance. *Ecography.* 2012, 35(6), 499-509.

Miller J., Species Distribution Modelling. *Geography Compass.* 2010, 4(6), 490-509.

Moudrý V. & Šímová P. Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science.* 2012, 26(11), 2083-2095

Osborne P. E. & Leitao P. J., Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity and Distributions.* 2009, 15(4), 671-6

Zurell D. et al., The virtual ecologist approach: simulating data and observers. *Oikos.* 2010, 119(4), 622-635.

Předběžný termín obhajoby

2017/18 LS – FŽP

Vedoucí práce

Ing. Lukáš Gábor

Garantující pracoviště

Katedra aplikované geoinformatiky a územního plánování

Elektronicky schváleno dne 16. 3. 2018

doc. Ing. Petra Šímová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 18. 3. 2018

prof. RNDr. Vladimír Bejček, CSc.

Děkan

V Praze dne 22. 03. 2018

Prohlášení

Prohlašuji, že jsem bakalářskou práci na téma *Vliv odlišné kvality environmentálních dat na modely druhové distribuce* vypracoval samostatně, pod vedením Ing. Lukáše Gábora a že jsem uvedl veškeré literární prameny a publikace, ze kterých jsem čerpal.

V Praze dne

.....

Poděkování

Díky této možnosti bych chtěl ještě jednou upřímně poděkovat panu Ing. Lukáši Gáborovi, za odborné vedení, ohromnou míru trpělivosti, že mě v mém životě posunul o velký kus dál a rozšířil mé obzory, poskytl mi mnoho cenných rad a ukázal tak správný směr, bez kterého by tato bakalářská práce neměla úspěšný cíl.

Dále bych chtěl poděkovat své rodině za podporu, která mi umožnila věnovat se klidnému studiu. Tereze za velkou morální podporu během kompletování práce a panu Ing. Vítězslavu Moudrému, Ph.D. za seznámení nás s Katedrou aplikované geoinformatiky a územního plánování a tím získat možnost tvořit bakalářskou práci právě pod touto katedrou.

Abstrakt

Modely druhové distribuce (dále jen SDMs) jsou hojně využívaným nástrojem v oborech, jako je biogeografie, makroekologie a ochrana přírody. Pomocí SDMs je možné například vymezit vhodné environmentální podmínky pro výskyt druhu, predikovat potencionální distribuci druhu nebo distribuci druhu na dosud neprozkoumaných lokalitách, určit vliv klimatických změn na rozšíření flóry a fauny či nalézt lokality ohrožené invazivními druhy. Přesnost SDMs zásadně ovlivňuje kvalita vstupních dat. Environmentální proměnné, vstupující do SDMs, musí obsahovat dostatečné informace o prostředí, ve kterém se druh vyskytuje a je v něm schopen reagovat na jeho změny. Proto je při tvorbě modelů vždy důležité volit vhodné rozlišení environmentálních proměnných. Testovat vliv odlišného rozlišení na výkon modelů druhové distribuce pomocí reálných dat však může být obtížné. Řešením je generování virtuálního druhu, které umožňuje snadno vliv environmentálních proměnných studovat. V bakalářské práci byla vytvořena distribuce virtuálního druhu, která byla následně srovnána s modely druhové distribuce, u kterých byly použité environmentální proměnné ve stejném rozlišení (1 km) a horším rozlišení (10 km a 50 km) a různě velké vzorky výskytových dat ($n= 50, 100, 500, 1000, 2000$). K vytvoření modelů byl použit v ekologii hojně využívaný model MaxEnt, který používá pouze presenční data. Výkon modelů, byl posouzen nejen pomocí AUC, ale také pomocí Schoener D indexu, který umožnil srovnat překryv výsledné predikce modelů s „reálným“ rozšířením virtuálního druhu. Z výsledných hodnot je patrné, že environmentální rozlišení nemá tak razantní vliv na výkony modelů.

Klíčová slova: druhové distribuční modely, virtuální druh, rozlišení environmentálních prediktorů, AUC, Schoener D index

Abstract

Species distribution models (hereinafter referred to as “SDMs” only) are a frequently used tool in such sectors as biogeography, macroecology and nature protection. With the help of SDMs it is possible, for example, to outline suitable environmental conditions for species occurrence, to predict potential distribution of a species or species distribution on the sites which have not been explored yet, to determine the impact of climatic changes on flora and fauna distribution or to identify the sites endangered with invasive species. The precision of SDMs is fundamentally affected by the input data quality. The environmental variables entering into SDMs must include sufficient information about the environment in which the species exists and in which it is able to respond to its changes. For this reason it is always important, during creation of the models, to select a suitable resolution of environmental variables. It may, however, be difficult to test the influence of a different resolution on the performance of species distribution models with the help of real data. The solution consists in a virtual species generation, which makes it possible to study the impact of environmental variables easily. The Bachelor’s Thesis provides for creation of a virtual species distribution, which was subsequently compared with the species distribution models, at which the environmental variables were used in the same (1 km) and in worse (10 km and 50 km) resolutions, together with various sizes of the occurrence data samples (n= 50, 100, 500, 1000, 2000). The MaxEnt model which is frequently applied in ecology and uses presence data only was used for creation of the models. The performance of the models was assessed not only with the help of AUC, but also with the help of Schoener’s D index, which made it possible to compare the overlapping of the prediction of models with “real” distribution of a virtual species. From the resulting values it is clear that the environmental resolution does not have so strong influence on the performance of models.

Keywords: species distribution models, virtual species, environmental resolution, AUC, Schoener's D index

Obsah

Seznam použitých zkratk a akronymů:	9
1. Úvod.....	10
2. Literární rešerše	11
2.1 Biotické interakce a ekologická nika.....	11
2.2 Druhové distribuční modely	13
2.3 Metody modelování	14
2.4 Metody vyhodnocování	16
2.5 Vstupní data.....	17
2.5.1 Výskytová data.....	17
2.5.2 Environmentální data	19
2.6 Vliv rozlišení vstupních dat na SDMs	21
2.7 Virtuální druh.....	22
3. Metodika.....	24
3.1 Vymezení a charakteristický popis zájmového území	25
3.2 Použitý software a vstupní data	25
3.2.1 Software	25
3.2.2 Vstupní data	26
3.3 Virtuální druh.....	26
3.4 Druhové distribuční modely	29
4. Výsledky práce.....	30
4.1 AUC.....	30
4.2 Schoener D – index.....	31
5. Diskuze	32
6. Závěr.....	34
Přehled literatury a použitých zdrojů	35
Seznam obrázků.....	39
Seznam tabulek.....	40
Seznam příloh	41

Seznam použitých zkratek a akronymů:

ANN	Uměla neurální síť (Artificial neural network)
AUC	Plocha pod křivkou (Area under cover)
CART	Klasifikační a regresní stromy (Classification and Regression Trees)
ENFA	Analýza faktorů ekologické niky (Ecological niche factor analysis)
GA	Genetické algoritmy (Genetic algorithm)
GAM	Generalizovaný aditivní model (Generalized additive model)
GIS	Geografické informační systém (Geographical Information System)
GLM	Generalizovaný lineární model (Generalized linear model)
MARS	Model mnohorozměrné adaptivní regrese (Multivariate adaptive regression model)
MAXENT	Model maximální entropie (Maximum entropy model)
NPP	Negativní prediktivní síla (Negative predictive power)
PPP	Pozitivní prediktivní síla (Positive predictive power)
ROC	Operační křivka (Receiver operating curve)
SDMs	Druhové distribuční modely (Species Distributions Models)

1. Úvod

Druhové distribuční modely (SDMs) v ekologii stále nabývají na popularitě z důvodu širokého spektra jejich využití. Jsou to modelovací nástroje používané v oblasti biologie a makroekologie. Pomocí těchto modelů lze predikovat druhové šíření v prostoru v rámci aktuálních či budoucích klimatických podmínek, pokles biodiverzity nebo odhalit budoucí biologické invaze (Meynard et al. 2013). Těmito modely je možné simulovat vztahy ovlivňující šíření organismů na zemi a tím lze předpovědět šíření do budoucna (Miller 2010). Technologický pokrok nastal, když se do SDMs zapojila technologie dálkového průzkumu země, tyto změny poskytly lepší a přesnější digitální data (Miller 2010). Znamená to, že modely druhové distribuce mají mnohem širší databázi pro výběr vstupních dat o daleko přesnějších a podrobnějších proměnných, které stále zdokonalují svoji přesnost. Toto přináší výhody detailnějšího zobrazení zájmových oblastí a lze tak zachytit nenápadné změny, které by při nižším rozlišení nebylo možné pozorovat (Valera et al. 2014).

Zkoumání použitého environmentálního rozlišení a určení jeho vlivu na SDMs je obtížné provádět s reálnými daty. Pro určení vlivu odlišných environmentálních proměnných, se začal využívat virtuální druh. Virtuální druh umožňuje oproti práci s reálnými daty porovnávat virtuální realitu a výslednou predikci modelu a přesně tak určit vliv odlišného environmentálního rozlišení na modely (Zurell et al. 2010). Virtuální druh lze vytvořit několika metodami, momentálně je hojně využívaná metoda pravděpodobnostního přístupu, která podporuje dynamické šíření druhů v prostoru a dokáže na změny pozvolna reagovat (Moudrý 2015). Proto byla využita v této práci.

V bakalářské práci byla provedena analýza vlivu environmentálního rozlišení na modely druhové distribuce. Cílem práce bylo ověřit, zda může environmentální rozlišení v různé kvalitě ovlivnit predikční schopnosti SDMs. Pro možnost porovnání bylo třeba vytvořit:

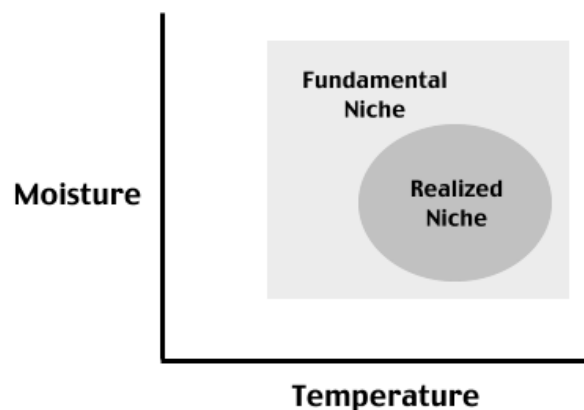
- 1) virtuální druh pravděpodobnostní metodou při environmentálním rozlišení 1 km
- 2) vytvořit modely druhové distribuce algoritmem Maxent.
- 3) vyhodnotit modely pomocí AUC a Schoener D indexu.

2. Literární rešerše

2.1 Biotické interakce a ekologická nika

Podle definice Hutchinsona, (1961) bereme základní niku určitého druhu jako souhrn podmínek daného prostředí, ve které je druh schopen přežít a setrvat (obr. 1). Každé takové místo odpovídá podmínkám životního prostředí, které dovoluje druhu existovat. Biotické interakce jsou vztahy mezi různými skupinami organismů, mohou být kladné (2 organismy spolu kooperují), záporné (kořist a predátor) a neutrální.

Niku podle Hutchinsona můžeme rozdělit na základní niku, kde je druh přirozeně schopný žít a realizovanou niku, která zobrazuje skutečné šíření druhu v přírodě.



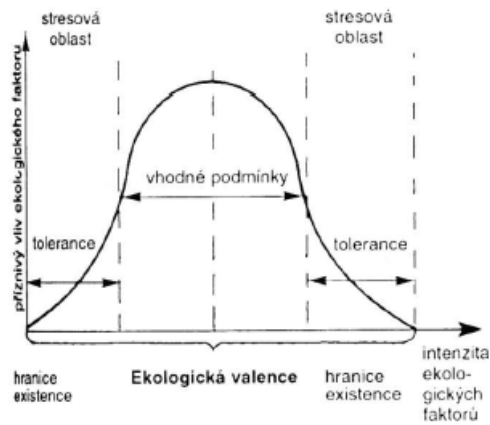
Obr. 1 Znárodnění základní (fundamental niche) a realizované niky (realized niche) jako podmnožiny v environmentálním prostoru, kdy osa x charakterizuje teplotu a y vlhkost.

(Zdroj: Hutchinson 1961)

Tato definice nezahrnuje veškerou složitost na vzájemné působení dvou a více ekologických jevů (ekologická prevalence), což je limitujícím faktorem prostorových modelů a tím omezují konečné výsledky a hodnoty SDMs (Guisan et al. 2007). Tento jev si je potřeba ujasnit před tvořením každého modelu z důvodu finálního výstupu, aby jeho následné prezentování bylo korektní a nevznikl nepoužitelný model (Miller 2010).

Ekologickou prevalenci si můžeme představit pomocí Gaussovy křivky, která zobrazuje reakci druhu na změny v environmentálním gradientu (změna intenzity faktoru přírodního prostředí – např. vlhkost, teplota, světlo které ovlivňují organismy) a lze z ní určit druhovou toleranci (obr. 2). Dále lze vyčíst optimální

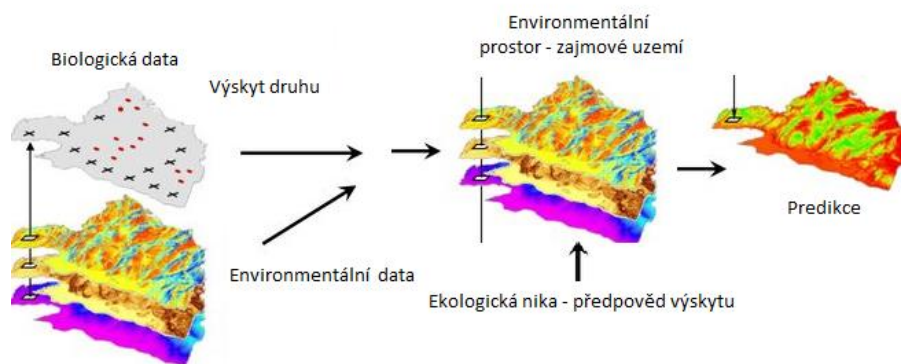
hodnoty ovlivňující výskyt druhu v zájmovém území. Ekologické gradienty prostředí jsou důležitým elementem při tvorbě SDMs (Václavík 2013).



Obr. 2 Znárodnění Gaussovy křivky s vlivy ekologických faktorů

(Zdroj: Václavík 2013)

Miller, (2010) definuje vztahy ekologické niky a jednotlivých gradientů při modelování SDMs takto: nika obecně popisuje environmentální potřeby (teplo, světlo) zkoumaného druhu v životních podmínkách výskytu, podle toho jsou statické metody schopné určit vztah: druh – prostředí. Podle těchto vztahů je možné získat předpověď environmentální vhodnosti prostředí ve zkoumané oblasti. Kterou je možné převést do geografického prostoru, čím získáme budoucí predikci (obr. 3).



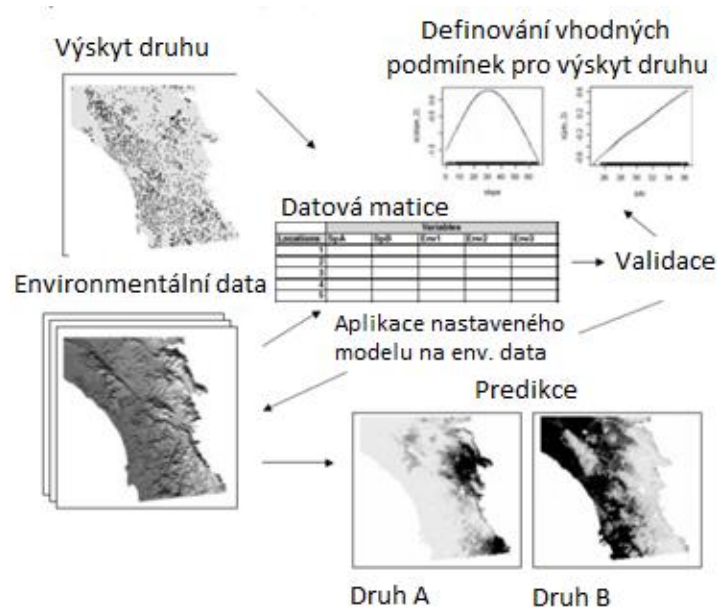
Obr. 3 Proces modelování SDM.

(Zdroj: Guisan & Thuiller 2005, autor)

2.2 Druhové distribuční modely

Historie modelů druhové distribuce začíná v 60. letech 20. století. V tuto dobu vznikaly první modely, které se zabývaly predikcí druhového šíření. Klíčové období rozvoje nastalo v polovině 80. letech 20. století, kdy došlo k rozmachu počítačových technologií, metod dálkového průzkumu země a geografických informačních systémů. Tyto souvislosti razantně přispěly a urychlily přesnost dat, rychlost jejich sběru a přesnost analýz v prostoru (Miller 2010).

SDMs jsou matematicko-statistické nástroje které, kombinují záznamy výskytu druhu s environmentálními gradienty prostředí (obr. 4). Pomocí těchto modelů lze predikovat aktuální či budoucí distribuci druhů. Momentálně jsou široce používaným nástrojem pro sledování ekologických a evolučních procesů odehrávajících se na zemském, vodním a mořském povrchu (Elith & Leathwick, 2009).



Obr. 4. Základní schéma funkce modelů druhové distribuce. Vztahy mezi daty, funkcemi a vyobrazení finální predikce.

(Zdroj: Elith & Leathwick, 2009)

Postup modelování druhové distribuce je složen z šesti navazujících kroků (Guisan & Zimmermann 2000). (I) Rozvržení koncepce, (II) příprava dat, (III) přizpůsobení modelu zkoumaného druhu, (IV) hodnocení modelu, (V) prostorový odhad, (VI) posouzení vhodnosti modelu. Dále je důležité si promyslet následující možnosti, podle kterých budu modelovat: (I) Teorie a data, (II) Metody modelování. Dále jen (I) – definice modelu od vstupních dat. Znamená, že posuzuji vztah, vhodnost a kvalitu dat v budoucím modelování. (II) - Výběr vhodných algoritmických výpočtů, statických metod a stanovení přesnosti v podobě měřítka (Guisan & Thuiller, 2005).

V modelování je důležité brát ohled na klimatické podmínky v podobě ročního období a druhového pohlaví (Guisan & Thuiller, 2005).

2.3 Metody modelování

Od vzniku modelování druhové distribuce bylo vytvořeno mnoho typů modelu. Stále se ale hledá vhodný univerzální model, který by šel jednotně využít pro zpracování všech typů dat, ale to není možné (Pearson 2010). Mezi všemi vzniklými modely existuje hlavní rozdíl. Jde o rozdíl správného využití modelového potenciálu. Ke každé modelovací studii musí existovat rozdílná modelovací metoda, v tom smyslu aby vstupní data byla správně zpracovaná, finální výstup byl korektní a zvolený algoritmus byl aplikovaný správně (Elith & Leathwick 2009).

Výběr vhodného algoritmu není jednoduché. Každý algoritmus přináší jiné výsledky (Elith et al. 2006). Je to způsobeno odlišností každého výzkumu, ten se vždy mění podle metodiky zpracování a požadovaných výsledků. Proto neexistuje jednotný algoritmus a modely nemohou být jednotně zpracovány. Tudíž by se měl použitý algoritmus odvíjet od použitelnosti, charakteristiky výskytových dat, jejich měřitelnosti a přesnosti prediktorů environmentálních dat (Miller & Franklin 2002). Algoritmy používané pro modelování druhové distribuce se dělí podle typu použití výskytových dat. Do algoritmu je možné vložit prezenční nebo prezenčně - absenční data (Miller 2010).

Modely na základech regresní analýzy (odhaduje se hodnota náhodné proměnné) jsou postavené na kombinaci odlišných možností druhového výskytu v závislosti na prostředí výskytu, využívají prezenčně - absenční data (Elith & Leathwick 2009). Výsledkem těchto analýz je pravděpodobnost výskytu druhů od 0 (žádná pravděpodobnost výskytu) do 1 (velmi vysoká pravděpodobnost výskytu). Příklad SDMs, který je založený na regresivní analýze může být model GLM (generalized linear model). Jako další příklady lze uvést model GAM (generalized additive models) a MARS (multivariate adaptive regression models). Tyto dva modely jsou oproti modelu GLM více flexibilní protože umožňují přesněji definovat křivku vyjadřující vztah druhu vůči prostředí (Miller & Franklin 2002). Studie uvádí, že model MARS je proti ostatním modelům lépe použitelný z důvodů úspory času, jeho zpracování dat po výpočetním hledisku je rychlejší získané hodnoty je jednodušší převést do geografického prostoru (Leathwick et al. 2005).

Tento typ modelu je v SDMs hojně využíván z důvodů že reálně simuluje navazování ekologických vztahů, tak jako v reálném světě. Další kladnou vlastností je, že dokáže zpracovávat data s nedokonalým záznamem prevalence druhu (Elith & Leathwick 2009).

Naproti tomu modely založené na neparametrických algoritmech fungují proti výše zmíněné metodě na klasifikaci zpracovaných dat a jsou vhodné k terénnímu průzkumu (Elith & Leathwick 2009). Jsou tak zejména využívány při studiích, kdy se uživatel soustředí na zaznamenání budoucí šíření druhu v zájmové oblasti.

Jako příklady uvádí Miller (2010) modely ANN (artificial neural network), CART (classification and regression trees) nebo GA (genetic algorithm). ANN pravidelně se opakuje v cyklech, umožňující doplnit používané algoritmy, čím mohou dosáhnout přesnějších výsledků. Model CART využívá možnosti nastavení specifického prahu ke každé ekologické proměnné. Data postupně rozděluje do nejlépe vystihujících záznamů prevalence nebo absence druhu. (Elith et al. 2006).

Miller (2010) podotýká, že je k dispozici další možná skupina modelů, která pracuje s prezenčními daty: ENFA (ecological niche factor analysis), MaxEnt (Maximum entropy model). Pomocí modelu ENFA lze sumarizovat proměnné v datech pomocí

nastavením filtru ekologických faktorů (Lobo et al. 2010). Další zmíněný model, který je použitelný při práci s malým objemem prezenčních dat je MaxEnt. Tento model vychází ze strojového učení (zabývá se algoritmy a technikami, které umožňují počítačovému systému 'učit se' a tak dokáže být více efektivní. Ten ze všech svých znalostí o zkoumaném druhu vytvoří co nejpřirozenější rozložení výskytu, čímž se stává často využívaným ohledně jeho přesných výsledků (Phillips et al. 2009).

Přes všechny uvedené příklady stále neexistuje jednotný typ algoritmu, který by šel univerzálně využívat na SDMs stejně. Vždy je nutné vnímat ekologické proměnné, podle kterých lze vybrat příslušný model s vhodným algoritmem, tak aby data správně zpracoval (Pearson 2010).

2.4 Metody vyhodnocování

Pro vyhodnocení predikčních schopností modelu je nutná evaluace (posouzení kvality). Důležité je, aby nebyla testována na stejných datech podle kterých byl model na kalibrovány (Araújo et al. 2005). Těchto podmínek lze docílit následujícími způsoby: (I) Rozdělím datovou sadu na kalibrační část a validační část v poměru např. (80% / 20%). (II) Křížová validace, tato technika rozdělí data do skupin, které se následně mezi sebou opakovaně testují. (III) Bootstrapping, tato metoda z datové sady náhodně převezme část dat, u kterých následně provádí validaci. Společná vlastnost těchto validačních metod je fakt, že pracují pouze s částí výskytových dat (Pearson 2010).

AUC (area under the curve) je evaluační metoda, která je úzce spojená s křivkou ROC (receiver operating characteristic). Křivka ROC je odvozená ze vztahu sensitivity a specificity – vykresluje vztah mezi predikcí prezenze a absence druhu. (Pearson et al. 2007). Allouche et al. (2006) definuje specificitu jako vztah mezi předpokládanou absencí a množstvím zanedbaných chyb a sensitivitu popisuje jako vztah předpokládané prezenze druhu a množstvím zanedbaných chyb. Tyto dva postupy na sobě nejsou vzájemně závislé a neodvíjejí se od prevalence (demografický ukazatel, počtu existujících druhů daného v oblasti výskytu k počtu druhů v dané lokalitě ve sledovaném časovém období). Z křivky ROC se následně

získá hodnota AUC (Area under the ROC curve) AUC zastává hodnotu plochy pod křivkou ROC. (Pearson 2010).

Pomocí AUC měříme do jaké míry je model schopný odlišovat zjištěné prezenze a absence. Měření poskytuje hodnoty od 0 do 1. Kdy hodnota 1 určuje nejlepší možnou přesnost modelu a hodnota 0 ukazuje na méně než náhodnou predikci modelu. Elith et al., (2006) ve své studii zmiňuje, že modely které mají nižší hodnotu AUC jak 0,5 jsou nepoužitelné, protože jejich predikce výskytu není pravdivá.

2.5 Vstupní data

Vstupní data v modelech druhové distribuce představují proměně, které jsou následně těmito modely zpracovány. Vstupní data dělíme: (I) environmentální data, jsou to data, která poskytují informace o prostředí (klimatická, topografická, land use a (II) výskytová data, která představují záznamy prezenze nebo absence druhu v zájmové oblasti. Oba typy dat mají velká kvantitativní i kvalitativní omezení a tak je důležité dbát na jejich správném použití (Pearson 2010).

2.5.1 Výskytová data

Tyto data jsou pro modelování druhové distribuce nejpodstatnější složkou. Protože na jejich kvalitě je postavena kalibrace modelu, tudíž se špatnou kvalitou těchto dat se model stává nepřesným (Elith et al. 2006). A zajistit kvalitní výskytová data pro model není jednoduché. Do různých typů SDMs mohou vstupovat data : (I) prezenční data (datová sada obsahuje pouze záznamy o výskytech), (II) prezenčně – absenční data (porovnávají se podmínky prostředí v oblastech kde se druh vyskytuje s oblastmi kde se nevyskytuje). (III) pseudo – absenční data (v oblasti výskytu je vybrána skupina bodů sloužící jako absenční záznamy). Ve studiích zabývajících se SDMs se však převážně používají hlavně prezenční data (Pearson 2010).

Výskytová data jsou získávána systematickým sběrem v terénu v oblasti výskytu druhu, nebo od třetí strany, které tyto data schraňují. Jsou to nejčastěji muzejní zdroje, databáze univerzit, herbářové sbírky. Tyto výskytové databáze jsou příkladem čistě prezenčních dat (Miller 2010). Výhodou těchto dat je lehká dostupnost a široké spektrum obsahu organismů a oblastí kde se nacházejí. Ty však nelze použít u velké

části modelovacích algoritmů, dokud si k nim neobstaráme absenční nebo pseudo – absenční záznamy. K těmto datům je nutno dodat že je provází chybnost v podobě nerovnoměrného sběru. Často v nich totiž převažují záznamy ze snadno dostupných lokalit, oproti nižšímu počtu záznamů z lokalit hůře dostupných (Edwards et al. 2005).

Výsledné vlivy použití podle typů výskytových dat v modelech druhové distribuce byly sledovány v mnoha studiích. Výsledek odhalil, že každá skupina dat má své specifika pro použití (Miller 2010).

Například při použití pouze prezenčních dat v modelu, vyhodnocovací schopnost nebere v úvah podmínky ovlivňující zbytek zájmové území, která lze vyřešit dosazením pseudo – absenčních záznamů (Edwards et al. 2005). Větší riziko je ignorování druhové prevalence, která udává frekvenci výskytu (Elith et al. 2006) čím vzniká fakt, že nedokážeme rozeznat zkoumaný druh který je vzácný nebo běžně se vyskytující. Tento problém dokáže vyřešit použití prezenčně – absenčních dat, které dokáží tento rozdíl odlišit. Podle Miller (2010) by se čistě prezenční data měla zaznamenávat podle náhodného nebo systematického sběru aby se tak předešlo špatné kalibraci modelu zaviněnou nedostatečným pokrytím vzorků v zájmové lokalitě.

Podle srovnávací studie, která vznikla od Elith et al. 2006 bylo zjištěno, že modely které využívají prezenčně – absenční data dosahovali vyšší predikční schopnosti než modely používající pouze prezenční data. Při použití prezenčně – absenčních dat v modelu potvrzuje lepší kalibraci i Pearson (2010). Při tomto závěru je však vhodné zmínit jednu možnost, která může nastat. Za situace kdy se absenční záznamy nacházejí na místech vyhovující pro výskyt druhu, ale v daný čas sběru výskytových dat se na místě druh neobjevuje. Tak následná správná kalibrace modelu je chybná. A výsledná predikce bude nepřesná. S tím souvisí další možná situace falešných absencí, tato situace znamená, že druh na místě výskytu nebyl zaznamenán, ale reálně se zde nachází. Při tomto jevu se přímo poškozují kalibrace modelu (Pearson 2010). Modely obsahující tyto druhy chyb následně mají chybnou predikci.

Z toho plyne, že každý typ výskytových dat zastává své klady a zápory. Při modelování druhové distribuce jsou k dispozici převážně prezenční data (Pearson 2010).

Modely druhové distribuce s využitím výskytových dat se dnes dají tvořit snadno podle internetových databází, které jsou volně přístupné (příklad zdrojů Tab. 1). Tyto zdroje obsahují hlavně prezenční data.

Tab. 1 Příklad zdrojů výskytových dat používaných při tvorbě SDMs

Organizace	URL	Dostupná data
Agentura ochrany přírody a krajiny ČR	http://portal.nature.cz/	nálezová databáze druhů na území ČR
Global Biodiversity Information Facility	http://data.gbif.org/welcome.htm	rostliny, živočichové, houby, mikroorganismy
NatureServe	http://www.natureserve.org/	rostliny, živočichové a ekosystémy v Jižní a Severní Americe

(Miller 2010, autor)

Další vliv, který dokáže zásadně ovlivnit přesnost modelu je velikost počtu vzorku. Čím je počet vzorků výskytových dat vyšší, tím se výsledná predikční schopnost modelu zvyšuje (Miller 2010).

2.5.2 Environmentální data

Druhý typ dat využívaných pro modely druhové distribuce jsou environmentální data. Tyto data využíváme u SDMs hlavně jako zdrojové prediktory. Jednou z metod, jak získat environmentální data, je dálkový průzkum země. Tato metoda poskytuje možnost aplikace velice detailních digitálních modelů terénu do SDMs (Elith et al., 2009). Navíc zde vzniká nová možnost jak vnímat povrch země v jiném spektru než viditelném, z čeho vyplívá rozšíření využití environmentálních proměnných. Z těchto primárních dat si můžeme pomoci software GIS zajistit další odvozené proměnné, která nám mohou zvýšit přesnost predikce (např. vzdálenosti od intravilánu). Pro modely druhové distribuce jsou nejvíce využívány klimatické vrstvy, které poskytují

informace o teplotě a množství srážek. Topografické vrstvy, které udávají nadmořskou výšku a sklon. A poslední dosti používané pro SDMs, vrstvy poskytují informace o půdním typu – land use (Miller 2010).

2.5.2.1 Klimatická data

Klima se pokládá za zdrojový faktor druhové distribuce a hraje největší roly. Klimatické vrstvy pro modelování vznikají z datového sběru z meteorologických stanic. Hijmans et al. v roce 2005 vytvořil portál WorldClim s volně dostupnými globálními klimatickými daty. Tento data-set obsahuje 19 klimatických vrstev zahrnujících sezonní a extrémní limitující faktory. Miller (2010) popisuje klimatická data jako souhrn přímých (teplota, světlo, záření) a nepřímých zdrojových (vlhkost) proměnných jako použitelné vrstvy při modelování.

Při využívání klimatických dat je potřeba brát zřetel ohledně jejich původu vzniku. Tyto data vznikají sběrem z meteorologických stanic, které nejsou systematicky rozestavěné, a proto nezajistí dokonalé pokrytí oblasti sběru. Z toho vyplývá, že data mohou v určitých oblastech disponovat nespolehlivostí v podobě mírné nepřesnosti (Hijmans et al., 2005). Pro dosažení přesných výsledků predikce je důležité vybrat takové prediktory, které nabývají stejné hodnoty jako zkoumaný druh (Guisan et al., 2007).

2.5.2.2 Topografická data

Topografické vrstvy stanovují hodnoty jako nadmořská výška a sklon. Využívají se v kombinaci s přímými a zdrojovými proměnnými v nižší prediktivní přesnosti modelu, když se použijí se zkušenostmi z předešlé praxe se zkoumaným druhem, mají na model vliv pozitivní (Guisan & Zimmermann 2000). Van niel et al. (2004) dokázal v jeho studiích, že při správném použití topografických proměnných modely SDMs dosahují menší chybnosti oproti modelu bez topografických proměnných, či proměnných s jednoduchým topografickými hodnotami.

Tab. 2 Příklad zdrojů environmentálních dat používaných při tvorbě SDMs

Organizace	URL	Dostupná data
WorldClim	http://www.worldclim.org/	interpolované klimatické rastry pro celý svět v rozlišení 1 km
UNEP	http://www.grid.unep.ch	typy půd pro celý svět
ČGS	http://mapy.geology.cz/geocr_50	geologická mapa ČR včetně typů půd

(Miller 2010, autor)

Do modelu SDMs se dají využít kromě klimatických a topografických dat další proměnné hodnoty, které mají vliv na výskyt druhu nebo jejich distribuci (Elith et al., 2009). Podle Miller (2010) jsou to informace o půdním typu, geologii území a druhu.

2.6 Vliv rozlišení vstupních dat na SDMs

U modelování druhové distribuce je dalším důležitým faktorem rozlišení vstupních dat, které může ovlivnit přesnost predikce (Guisan et al. 2007). Rozlišení lze definovat jako velikost pixelového čtverce - zrna u rastrových sítí environmentálních dat. Zhoršováním rozlišení environmentálních dat se postupně klesá hodnota proměnných v datech a vytrácejí se tak informace o povrchu. Čím více informací o povrchu zaznamenáváme tím více je krajina heterogenní. Tato souvislost je důležitá u SDMs z ohledu navazování vazeb (druh – ekologické potřeby pro výskyt). Zkoumané druhy, které mají vysoké ekologické nároky na habitat tak při modelování v nízkém rozlišení ztrácí vazby na prostředí, které jsou důležité pro kalibraci modelu (Goodchild 2011). Další problém ohledně rozlišení nastává s výskytovými daty, které mají jiné rozlišení než data environmentální. Environmentální data je následně potřeba zhoršit na rozlišení výskytových dat (Guisan et al. 2007).

Goodchild, (2011) publikuje názor, že problematika spojená s environmentálním rozlišením není důkladně analyzována. Hlavním důvodem je dostupnost a finanční náročnost přesných environmentálních dat. S technickým pokrokem dálkového průzkumu země postupně tyto důvody mizí.

Kromě rozlišení se u SDMs řeší problém s použitým měřítkem. Každé modelování druhové distribuce pro odlišný druh sebou nese použití jiného měřítka. Například pro modelování druhové distribuce mikroorganismů bude použito jiné měřítko, než u modelování druhové distribuce ptáků. Je to dané tím že všechny organismy se v prostředí šíří podle odlišných pravidel (Varela et al. 2014). Rozdělení rozlišení je založené na třech zásadních principech, které by měly být respektovány: (I) pozornost při tvorbě SDMs by se vždy měla soustředit na měřítko, ve kterém se modelovaný proces odehrává, (II) univerzální měřítko pro SDMs neexistuje, (III) pokud se změní měřítko, mohou se podstatně změnit i studované procesy (Guisan et al. 2007).

2.7 Virtuální druh

V předešlých kapitolách bylo zmiňováno, že kvalita výskytových a environmentálních dat zásadně ovlivňuje přesnost modelů druhové distribuce (Duputié et al. 2014). Zjistit jaký podíl mají chyby ve výskytových a environmentálních datech na výslednou přesnost při tvoření SDMs je obtížné. Řešením je využití virtuálního druhu (Moudrý 2015).

Základní myšlenka využití virtuálního druhu v SDMs souvisí s vygenerováním rastru environmentální vhodnosti. U kterého se nastaví uměle nadefinovaná odpověď virtuálního druhu na environmentální proměnné v prostředí za pomoci odezvové funkce (např. Gaussova křivka) (Meynard & Kaplan 2012). Vzniká rastr environmentální vhodnosti, ten lze konvertovat na rastr prezence a absence, který obsahuje datovou sadu o prezenci a absenci druhu v binární podobě. Tento rastr je zobrazením virtuální reality výskytu nadefinovaným virtuálním druhem. Finálním krokem je získání dat z virtuální reality odběrem vzorků prezence, který simuluje sběr výskytových dat v terénu. Pomocí virtuálního druhu byl zajištěn vzorek výskytových dat, který dále může být v SDMs zpracován jako výskytová data reálného druhu.

Velkou výhodou u virtuálního druhu je možnost definice uměle vytvořené odpověď i na environmentální proměnné, tímto krokem lze kontrolovat vzorec druhového výskytu a dává nám možnost modelovat SDMs ve zvoleném měřítku, nebo zkoumat odlišné environmentální rozlišení dat na modely druhové distribuce (Zurell et al. 2010).

Virtuální druh lze vygenerovat pomocí *SDMvspecies* a *Virtualspecies* (package) Moudrý (2015).

SDMvspecies počítá s prahovými hodnotami. Nastavením prahové hodnoty dosáhne uživatel výskytu druhu nad touto hranicí, nikdy však pod ní (Xiaoquan 2015). Při použití této metody generace nastává kritika ze strany (Meynard & Kaplan 2012, Moudry 2015) kteří dokládají, že vede k nadhodnocení modelu a jeho výsledek může být mylný.

Virtualspecies funguje proti výše zmiňovanému rozdílně v tom, že nepočítá s prahovou hodnotou, ale užívá pravděpodobností přístup. Program má lepší použitelnost (odstraňuje předešlé nedostatky s nadhodnocováním modelu) s možností přizpůsobení definovat jakýkoliv vztah k druhu – vzrůstá prostředí při modelování. Výhodou je i možnost nastavení obecného rozšíření a jeho aplikace při převodu rastru vyjadřující vhodnost prostředí na prezenci – absenci druhu (Leroy et al. 2015, Moudry 2015).

3. Metodika

Jako zájmové území pro mou praktickou část práce byla zvolena oblast Skandinávie, hlavním důvodem volby byla rozloha území. V následujícím metodickém postupu byl vytvořen virtuální druh z klimatických vrstev z globální databáze WorldClim, kde byli použity klimatické vrstvy Bio5 - poskytují informace o nejvyšší teplotě nejteplejšího měsíce, Bio6 - nejnižší teplotě nejchladnějšího měsíce a Bio12 – úhrn ročník srážek. Podle kterých jsem nadefinoval vztah virtuálního druhu ke každé environmentální proměnné. Tímto krokem jsem získal rastr environmentální vhodnosti prostředí, který jsem následně převedl na prezenčně – absenční rastr a získal tak virtuální realitu s výskytovými daty. Jako další krok jsem stanovil počet odběru vzorků pro prezenční data (n= 50, 100, 500, 1000 a 2000). Modely druhové distribuce byly následně vytvořeny podle metody MaxEnt z důvodu využití pouze prezenční data, druhý důvod pro zvolení metody MaxEnt je způsob jakým data zpracovává. Jako environmentální prediktory pro modelování jsem použil vrstvy Bio5, Bio6 a Bio12 podobně jako u vytváření virtuálního druhu. Tento postup byl 3x zopakován pro 3 odlišené environmentálního rozlišení (1 km, 10 km, 50 km) v 50 po sobě jdoucích cyklech ze kterých bylo dohromady získáno 250 informačních-výsledků pro hodnocení vlivu environmentálního rozlišení na modely druhové distribuce. Pro výsledné vyhodnocení modelů jsem použil metody AUC a Schoener D indexu, jejich hodnoty jsem porovnal zvlášť pro každé rozlišení a učinil tak závěr podle odborných studií zaměřených na tuto problematiku.

3.1 Vymezení a charakteristický popis zájmového území

Jako zájmové území byla zvolena část Skandinávie ležící v severní části Evropského kontinentu, která zahrnuje severské země jako Norsko, Švédsko a Finsko. Vybrané území, má plochu 1 111 214 km² a leží v oblasti od mírného pásma až po oblast severního polárního kruhu. Díky Golfskému proudu má oblast na většině území relativně mírné podnebí. Velká část skandinávského pohoří má podnebí horské tundry. V oblasti vybraného území se nachází velké množství jezer a morén, které představují pozůstatky doby ledové (Rönnqvist 2010).

3.2 Použitý software a vstupní data

3.2.1 Software

Pro účely bakalářské práce byl použit software R a ArcMap.

R je programovací jazyk a softwarové prostředí určené pro statickou analýzu a její následné vyobrazení. Softwarové prostředí lze rozšířit pomocí knihoven označované jako balíčky (package). Při práci s R se využívá příkazového řádku na jeho obsluhu. Software je distribuován bezplatně na www.R-project.org. Pro vyhotovení práce bylo nutno o několik rozšíření (Package). Jedná se o rozšíření *raster*, *virtualspecies*, *dismo*, *maptools*, *sp*, *reshape2*, *rgdal*, *sqldf*, *maps*, *rJava*, *rock*, *vcd* a *boot*.

ArcMap je software který dokáže pracovat v prostředí geografického informačního systému (GIS). Tento software umožňuje pracovat s prostorovými daty. Data lze přes rozhraní ArcMapu vytvářet, spárovat a především je dokáže analyzovat. Tím v nich dokáže najít nové vztahy a následně je přehledně zobrazit. Výsledky lze následovně používat ve formátu tradičních map.

Během prováděné studie za pomoci tohoto softwaru byly uskutečněny první pracovní kroky, kdy byly vrstvy severních států Finska, Norska, Švédska (dostupné na www.gadm.org) sjednoceny pomocí funkce *Merge* aby vznikla 1 souvislá vrstva z 3. Následně byla použita funkce *Extract by Mask* podle které byly oříznuty klimatické vrstvy souvislou vrstvou severských států. Klimatické vrstvy jsou dostupné z databáze WorldClim. Posledním krokem z příprav byla funkce *Resample* která dokáže změnit rozlišení pixelových buněk v rastru na (1 km, 10 km, 50 km).

Tímto způsobem byla vytvořena datová sada v podobě klimatických vrstev z databáze WorldClim která bude vstupovat do software R kde, bude následně zpracována pro studování vlivu environmentálního rozlišení na modely druhové distribuce.

3.2.2 Vstupní data

WorldClim je internetová databáze která poskytuje soubor globálních klimatických dat v rozlišení od 1 km do 340 km v souřadnicovém systému WGS 1984. Tyto data se dají využít v prostorových analýzách GIS, tedy i pro modelování druhové distribuce. Vrstvy vznikly jako množiny dat z měření meteorologických stanic, které se pomocí interpolace aplikovali na zemský povrch. Obsahují informace typu: minimální, průměrná, maximální teplota, množství srážek, rychlosti větrů a další klimatické proměnné. Nabízí i průměrné roční hodnoty a extrémní jevy počasí (např. srážky v nejvlhčím měsíci, srážky v nejsušším měsíci a nejvyšší teplota nejteplejšího měsíce a nejnižší teplota nejchladnějšího měsíce (Hijmans et al. 2005).

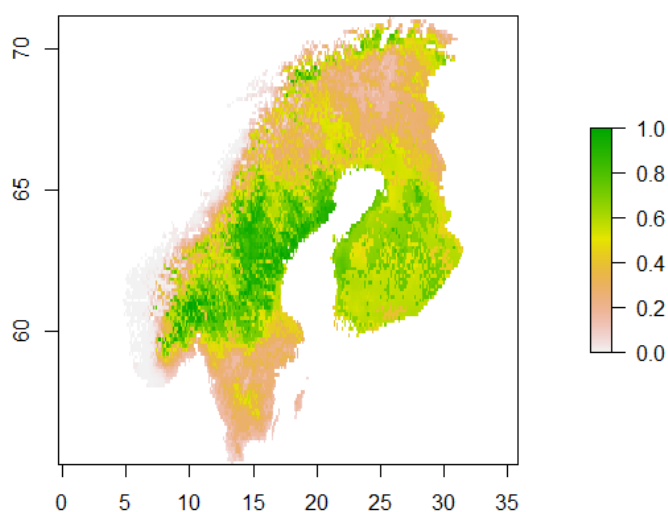
V této práci bylo z databáze WorldClim použito vrstev Bio5, Bio6 a Bio12 které poskytují informace o nejvyšší teplotě nejteplejšího měsíce, nejnižší teplotě nechlazenějšího měsíce a ročních srážkách.

3.3 Virtuální druh

Virtuální druh byl vytvořen v software R prostřednictvím volně dostupného balíčku (package) *Virtualspecies*. Prvním krokem v tomto programu bylo nahrát a sjednotit klimatická data (environmentální vrstvy) bio5, bio6, bio12 předpřipravené z ArcMap. Tato funkce má název *raster* pro nahrání, následuje funkce *stack* pro sjednocení.

Pomocí odezvové funkce *formatFunction*, kterou nadefinuje uživatel, lze stanovit umělá odpověď virtuálního druhu na environmentální proměnné. Následuje funkce *generateSpFromFun*, čím vytvoříme rastr o environmentální vhodnosti prostředí pro daný druh (obr. 5).

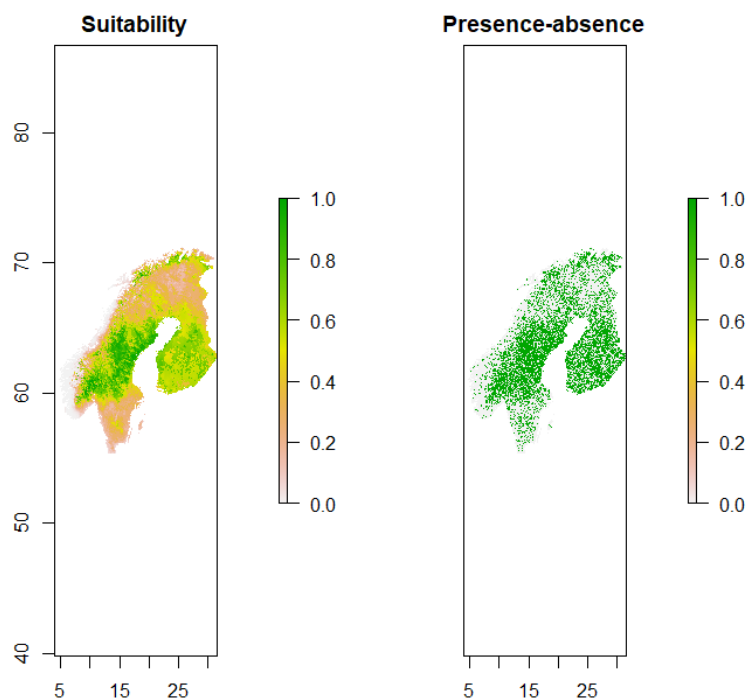
Environmental suitability of the virtual species



Obr. 5 Ukázka environmentální vhodnosti – obrázek je výstup z 3 klimatických vrstev, kde jsme z každé z nich nadefinovali umělou odpověď na reálný svět. Zelené barva nám určuje vhodné prostředí pro výskyt zkoumaného druhu, bílá barva nejméně vhodné prostředí pro výskyt druhu.

(Zdroj: Autor)

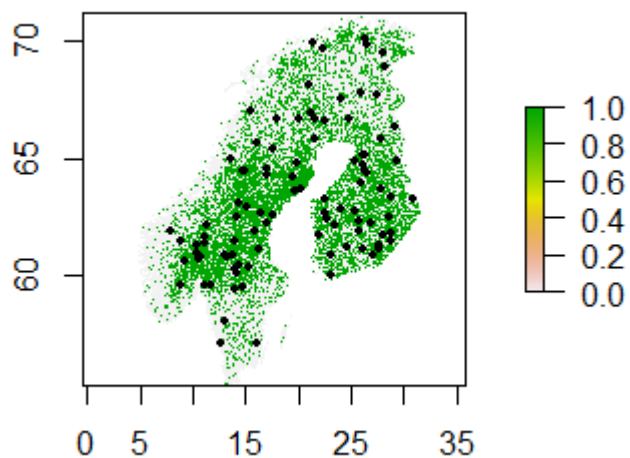
Abychom získali informace o výskytech druhu, tak se rastr environmentální vhodnosti konvertuje na prezenčně-absenční rastr (obr. 6). Po této akci je vhodné aplikovat metodu pravděpodobnostního přístupu, hodnoty vhodnosti prostředí jsou pomocí logistické a lineární funkce převedeny na pravděpodobnost výskytu a tvar funkce kontroluje uživatel. Odůvodnění je takové že druh se následovně může objevit v celém environmentálním gradientu jako v reálném světě (Moudrý 2015). Slouží na to v R funkce *convertToPA* kde se definují hodnoty alfa (sklon křivky) a beta (poloha inflexního bodu, hodnoty jsem navolil 0.5 a -0.2, tyto hodnoty jsou v základním nastavení a není potřeba je měnit.



Obr. 6 Převod rastru environmentální vhodnosti na prezenčně absenční rastr.

(Zdroj: Autor)

U modelů druhové distribuce konstruované pomocí prezenčně-absenčních dat se následně provádí *sampling*, tento nástroj simuluje sběr výskytových dat v zájmové oblasti (obr. 7). R nabízí pro tuto situaci funkci *sampleOccurrences*. Kde stanovíme počet vzorků (*sample-size*) – $n = 50, 100, 500, 1000, 2000$. A následně získáme informace v podobě bodů, které zachycující data výskytu.



Obr. 7 Ukázka odběru vzorků výskytu druhu pro (*sample = 100*).

(Zdroj: Autor)

3.4 Druhové distribuční modely

Modely druhové distribuce jsem tvořil podle metody MaxEnt (Maximum entropy model), ten ze všech svých znalostí o zkoumaném druhu vytvoří co nejpřirozenější rozložení výskytu druhu Phillips et al., (2009). Z tohoto důvodu je často využíván ve studiích, která využívají jako datovou sadu prezenční data. MaxEnt jsme ponechali v základním nastavení (default). Pozměnili jsme pouze nastavení v testování. Kde byla testovaná data rozdělena na validační 20% a kalibrační 80% z celkového počtu prezenčních dat. Klimatické vrstvy Bio5, Bio6 a Bio12 které poskytují informace o nevyšší teplotě nejteplejšího měsíce, nejnižší teplotě nechladnějšího měsíce a ročních srážkách byly použity jako prediktory.

Pro vyhodnocení modelů byly použity metody AUC (diskriminační ukazatel přesnosti distribučních modelů) a Schoener D index (shodnost překryvu niky). Tyto vyhodnocovací metody nám zajistí hodnoty, dle kterých jsme schopni ohodnotit přesnost a výkon daného modelu.

AUC nabývá hodnot od 0 do 1, kde 1 udává parametry nejlepší možné přesnosti, 0,8 – 0,9 dobrá přesnost, 0,7 – 0,8 průměrná přesnost, 0,6 – 0,7 podprůměrná přesnost a hodnota nižší než 0,5 poukazuje na model, který není schopný správné predikce, tudíž je nedostačující a nejde použít (Kienast et al. 2012).

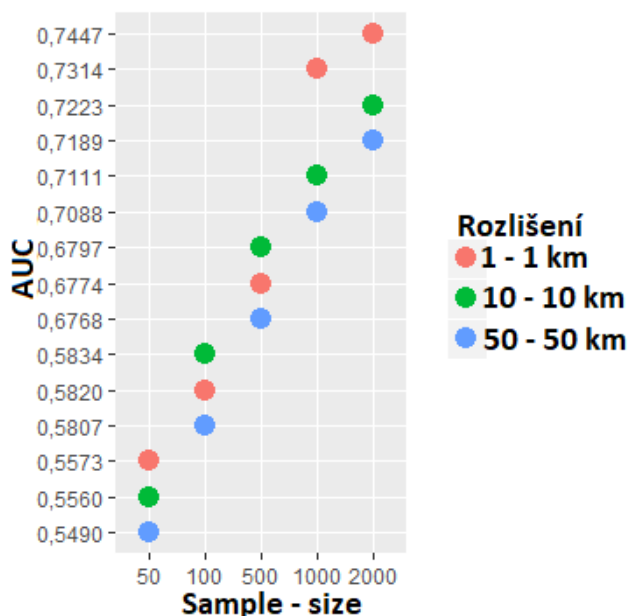
Schoener D index (niche overlap) udává míru překryvu – podobnosti niky mezi dvěma oblastmi (environmentální data). Představuje tak poměr shody mezi nimi. Nabývá hodnot od 0 do 1. Kdy hodnota 1 znamená absolutní shodu neboli totožnost, hodnoty blíže k 0 udávají nízkou míru překryvu. V mém modelu do Schoener D indexu vstupují vrstvy environmentální vhodnosti prostředí a distribuční mapa vytvořená algoritmem MaxEnt (Broennimann et al. 2012).

4. Výsledky práce

4.1 AUC

Z obrázku 8 lze vyčíst, že environmentální rozlišení nemá razantní vliv výkon modelu. Při posuzování modelu pomocí AUC, bylo zjištěno, že výsledná odchylka při porovnání modelů s odlišným rozlišením environmentálních proměnných (konkrétně u vzorků 100 a 500) je zanedbatelná (0,01 – 0,03). Razantnější odchylka byla zaznamenána při vzorku 1 000 a 2 000 (0,1). Lze podotknout, že nejpřesnější environmentální rozlišení (1 km) přináší nejvyšší hodnoty AUC při vzorku 50, 1000 a 2000. Při vzorku 100 a 500 dosáhlo nejvyšších hodnot AUC rozlišení 10 km.

S navyšujícím se počtem vzorku, dosahoval model vyšších hodnot AUC při rozlišování mezi absencemi a prezencemi. Z teoretických studií vyplívá, že čím větší počet vzorku do modelu vstupuje, tím vyšší přesnosti model nabývá. Při počtu vzorku 50 dosáhly modely hodnot AUC (1 km - 0,5573, 10 km - 0,5560 a 50 km - 0,5490) zatímco při počtu vzorku 2 000 dosáhly modely hodnot AUC (1 km - 0,7447, 10 km - 0,7223 a 50 km - 0,7189).



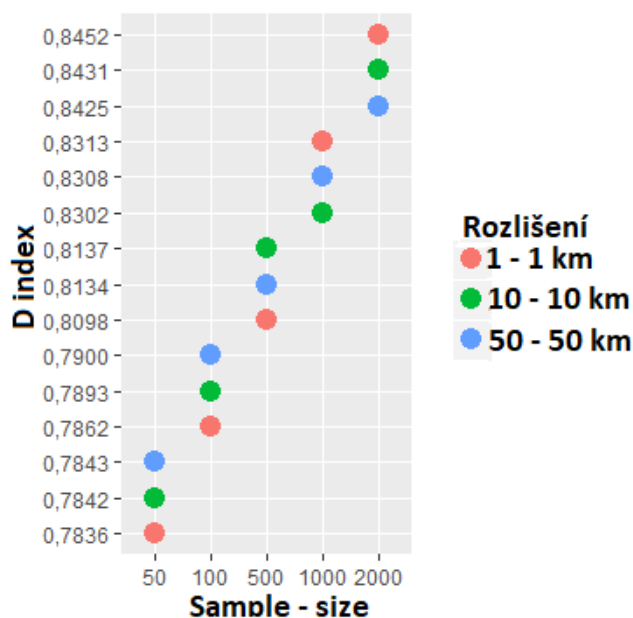
Obr. 8 Graf výsledných hodnot získaných vyhodnocovací metodou AUC. Kde na ose x jsou vyobrazeny počty vzorků (sample-size) a na ose y výše hodnot AUC.

(Zdroj: Autor)

4.2 Schoener D – index

Modely byly validovány i druhou metrikou, a to pomocí Schoenerova D indexu (Obr. 9). Při vzorku 50 a 100 byl největší překryv niky zaznamenán při environmentálním rozlišení 50 km (0,7843, 0,7900). Při vzorku 500, nejvyšší hodnoty D indexu bylo dosaženo při rozlišení 10 km (0,8137). Při vzorku 1 000 a 2 000 byla nejvyšší hodnota D indexu zaznamenána při rozlišení 1 km (0,8313, 0,8452).

S rostoucím počtem vzorků dochází k lineárně se navyšujícím hodnotám překryvu niky podobně jako metody AUC, například při environmentálním rozlišení 1 km byla hodnota D indexu při vzorku 50 (0,7836) při vzorku 2 000 dosáhl model hodnoty D indexu (0,8452).



Obr. 9 Graf výsledných hodnot získaných vyhodnocovací metodou D indexu. Kde na ose x jsou vyobrazený počty vzorků (sample-size) a na ose y výše hodnot D indexu.

(Zdroj: Autor)

5. Diskuze

Výsledné hodnoty se stejným vzorkem, dosáhli velice podobných hodnot u AUC i D indexu. Změny rozlišení z 1 km na 10 km a na 50 km nezpůsobily žádné extrémní výkyvy v hodnotách AUC a D indexu. Hodnoty se mění pouze v řádech setin a tisícín, z toho lze učinit závěr, že environmentálního rozlišení nemá tak razantní vliv na výkon modelu.

U modelů s využitím virtuálního druhu, byla při validaci Schoener D indexem zaznamenána dobrá predikční schopnost i při nízkém počtu vzorku (0.78), při validaci AUC byla dosažena nedostačující kvalita modelů při nízkém počtu vzorků (0.54 – 0.55) Nezer et al., (2017).

Když porovnáme výsledné hodnoty získané metodou AUC. Odchylka mezi jednotlivými environmentálními rozlišeními činila (0,02 – 0,03). Tato odchylka odpovídá hodnotám získaných při podobné studii od autorů Nezer et al., (2017), kde byla porovnávána predikční schopnosti modelu v rozlišení 10 m, 100 m, a 1 km. Z jejich studie byl zaznamenán pokles hodnot AUC mezi rozlišením 10 m a 1 km, podobně jako v mé studii kdy odchylka činila hodnotu cca 0,03.

Obdobná studie zaměřená na environmentální rozlišení, se kterou lze porovnat hodnoty výzkumu z mé práce je od autorů Guisan et al. (2007), kde bylo znovu doloženo obdobných výsledků jako při mé studii. Pokles hodnot AUC ve studii od Guisan et al. (2007) byl mezi rozlišením 200 m a 5 km. U mé studie jsem zaznamenal výraznější pokles hodnot mezi rozlišením 1 km a 50 km a to v řádech (0,02 – 0,03) při vyšší vzorku > 1000.

Za zmínění stojí podrobná studie environmentálního rozlišení zpracovaná taktéž autory Guisan et al. (2007). Kde se průzkum zabýval vlivem rozlišením environmentálních dat, rozlišení dat bylo až 10 násobně zhoršené (ze 100 m až do 1 km). Výzkum doložil obdobné výsledky jako vyšlé z mé studie. Bylo dosaženo mírného lineárního poklesu hodnot AUC na úrovni setin (cca – 0,03).

Ohledně získaných výsledků AUC které činí průměrné odchylky (0,02 – 0,03), lze z výše uvedených odborných studií vyvodit, že má studie byla provedena správně, protože dokládá podobné hodnoty. Studie rovněž potvrdili, že takto mále odchylky nemohou hrát velký význam na výkon modelu (Guisan et al., 2007).

Zároveň je vhodné zmínit, že existují studie, kde zkoumaný vliv rozlišení nehrál žádnou roli. Xu et al., (2016) ve své studii o slonech nezaznamenal žádnou vazbu mezi zvoleným rozlišením a predikční schopnosti modelu. Guisan et al. (2007) podotýká, že lze dohledat i takový případ, kdy horší rozlišení zvýšilo predikční schopnost modelu.

Další zjištění se týká ohledně výše počtu použitých vzorků, se kterými model pracuje. Modely, které využívají vyššího počtu vzorku, dosahují vyšších hodnot AUC a Schoener D indexu než modely, které využívají menší počet vzorků. Zároveň modely, které používaly vyšší počet vzorku, byly více citlivě na zhoršení rozlišení, než modely které používali nižší počet vzorků (Pearce & Ferrier 2000).

6. Závěr

Bakalářská práce byla zaměřena na problematiku vlivu environmentálního rozlišení na SDMs. Byla vytvořena distribuce virtuálního druhu, která se následně porovnávala s modely druhové distribuce, u kterých byly použity environmentální proměnné s rozlišením 1km a zhoršeným rozlišením 10 km a 50 km. Dále se v práci zabývala jaký vliv má na SDMs výše použitého vzorku výskytových dat ($n= 50, 100, 500, 1000, 2000$). Cílem práce poté bylo prokázat, zda má environmentální rozlišení vliv na výsledky predikce modelu. Modely byly validovány pomocí AUC a D indexu. Nejvyšších hodnot AUC bylo dosaženo při využití nepřesnějšího rozlišením 1 km a počtu vzorků 2000 (0,7447). Nejvyšší hodnota D indexu byla dosažena při využití nepřesnějšího rozlišení 1 km a počtu vzorků 2000 (0,8452). Z výsledku práce bylo prokázáno, že environmentální rozlišení nemá tak razantní vliv na modely druhové distribuce. Hodnoty odchylek AUC a D indexu se pohybují v řádech tisícín a setin. Takto mále odchylky dokáží jen nepatrně zhoršit výkon modelů a proto jsou zanedbatelné. Podle studie od Kienast et al., (2012) na klasifikaci přesnosti AUC, lze výsledky z mé práce rozčlenit do několika kategorií. Nejvyšší získaná přesnost byla 0,7447, která má status průměrných hodnot. Výsledná hodnota AUC je přímo závislá na výši použitých vzorků. Nejnižší získaná přesnost činila 0,5490 je označována za nedostačující. Výsledky mé bakalářské práce, se shodují s výsledky s podobně zaměřenými studii zkoumající vliv environmentálního rozlišení na modely druhové distribuce a jejími autory jsou Nezer et al., (2017), Xu et al., (2016) a Guisan et al., (2007). Pro rozšíření této práce by bylo vhodné ji doplnit o více kategorií rozlišení environmentálních dat, použít extrémní počty vzorků nebo se zaměřit na různě rozšířené druhy. Tato úprava by mohla docílit lepšího zjištění schopnosti modelů pomocí validačních technik AUC a D indexu.

Přehled literatury a použitých zdrojů

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*. 43, 1223-1232.

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

Broennimann, O., Fitzpatrick, M., Pearman, P., Petitpierre, B., Pellissier, L., Yoccoz, N., Graham, C. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global ecology and biogeography*, 21, 481-497.

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

Edwards, T., Cutler, D., Zimmermann, N., Geiser, L., Alegria, J., 2005. Model-based Stratifications for enhancing the detection of rare ecological events. *Ecology*, 86, 1081-1090.

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

Elith, J., Catherine, H., Graham, R., Dudík, M., Ferrier, M., Guisan, S., Hijmans, R., Huetman, F., Leathwick, J., Lehmann, A., Li, J., Lohman, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberón, J., Williams, S., Wisz, M., Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.

Goodchild, M., 2011. Scale in GIS: An overview. *Geomorphology*. 130, 5-9.

Guisan, A., Graham, C., Elith, J., Huettmann, F., NCEAS Species Distribution Modelling Group, 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13, 332–340.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: Offering more than simple habitat models. *Ecology Letters* 8, 993–1009.

Guisan, A., Zimmermann, E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.

Hijmans, J., Cameron, E., Parra, J., Jones, P., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978.

Hutchinson, G., 1961: *The Paradox of the Plankton.* The American Naturalist. Chicago: University of Chicago Press.

Kienast, F., Bolliger, J., Zimmermann, N., 2012. *Advanced Landscape Ecology: Lecture*

Notes. Dostupné z: [ftp.wsl.ch/pub/kienast/mast_old/class_no1/script_ds1_11.pdf](ftp:wsl.ch/pub/kienast/mast_old/class_no1/script_ds1_11.pdf)

Leroy, B., Meynard, C., Bellard, C., Courchamp, F., 2016. Virtualspecies, an R package to generate virtual species distributions. *Ecography*, 39, 599-607.

Lobo, J., Jiménez-Valverde A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography*. 33, 103-114.

Leathwick, J., Rowe, D., Richardson, J., 2005 Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, 50, 2034-2052.

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

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

Miller, J., Franklin, J., 2002. Modelling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*. 157, 227-247.

Moudrý, V., 2015. Modelling species distributions with simulated virtual species, *Journal of Biogeography*. 42, 1365-1366.

Moudrý, V., Šimová, P., 2012. Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science*. 26, 2083-2095.

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

Pearce, J., Ferrier, S., 2000: Evaluating the predictive performance of habitat models developer using logistic regression. *Ecological Modelling* 133, 225–245.

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

Pearson, R., Raxworthy, J., Nakamura, M., Peterson, T., 2007. Predicting species distributions from small numbers of occurrence records. *Journal of Biogeography*. 34, 102-117.

Phillips, S., Dudík, M., Elith, J., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*. 19, 181-197.

Rönnqvist, H. 2010. Predicting surfacing internal erosion in moraine core dams (Doctoral dissertation, KTH).

Václavík, T., Modelování habitatu a distribuce druhů. 2013 Dostupné z: <https://tomasvaclavik.files.wordpress.com> .

Van Niel, K., Laffan, S., Lees G., 2004. Effect of error in the DEM on environmental variables for predictive vegetation modelling. *Journal of Vegetation Science*. 15, 747-756.

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

Xiaoquan, K., 2015. Create virtual species with sdmvspecies. Dostupné z: <http://cran.revolutionanalytics.com/web/packages/sdmvspecies/vignettes/sdmvspecies.pdf>.

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

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

Seznam obrázků

Obr. 1 Znázornění základní a realizované niky.....	11
Obr. 2 Znázornění Gaussovy křivky s vlivy ekologických faktorů.....	12
Obr. 3 Proces modelování SDM.....	12
Obr. 4 Základní schéma funkce modelů druhové distribuce.....	13
Obr. 5 Ukázka environmentální vhodnosti.....	27
Obr. 6 Převod rastru environmentální vhodnosti na prezenčně absenční rastr.....	28
Obr. 7 Ukázka odběru vzorku výskytu druhu pro (sample = 100).....	28
Obr. 8 Graf výsledných hodnot získaných vyhodnocovací metodou AUC.....	30
Obr. 9 Graf výsledných hodnot získaných vyhodnocovací metodou D indexu.....	31

Seznam tabulek

Tab. 1 Příklad zdrojů výskytových dat používaných při tvorbě SDMs..... 19

Tab. 2 Příklad zdrojů environmentálních dat používaných při tvorbě SDMs..... 21

Seznam příloh

Příloha A: Mapa prezentující zájmového území Skandinávie v Evropě

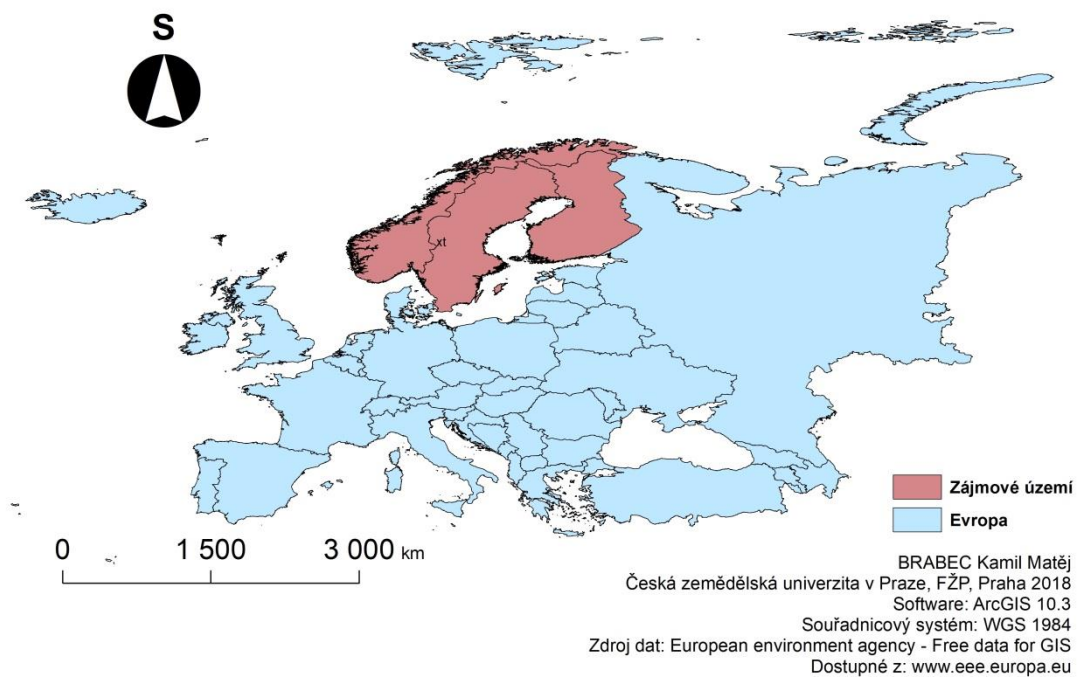
Příloha B: Mapa klimatických vrstev na území Skandinávie v rozlišeních 1 km, 10 km a 50 km

Příloha C: Tabulka výsledků vyhodnocací metody AUC

Příloha D: Tabulka výsledků vyhodnocovací metody Schoener's D indexu

Příloha A: Mapa prezentující zájmového území v Evropě

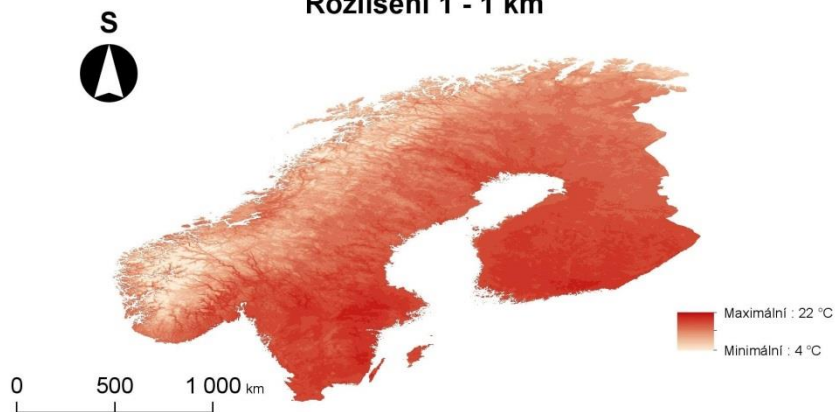
Zájmové území



**Příloha B: Mapy všech klimatických vrstev na území Skandinávie
v různých rozlišeních.**

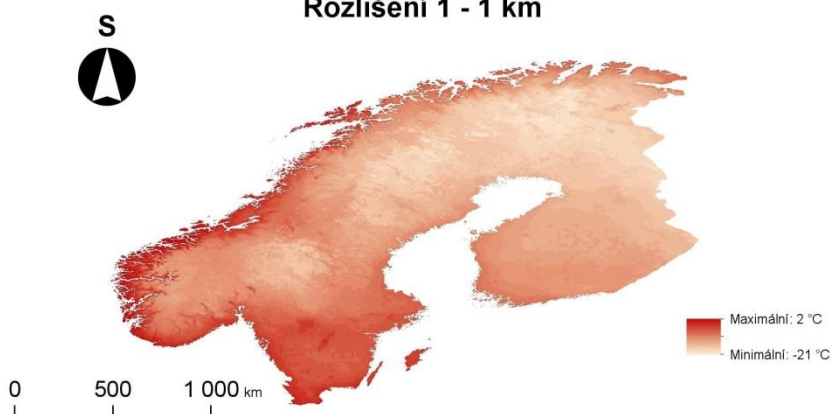
Maximální teplota nejteplejšího měsíce

Rozlišení 1 - 1 km



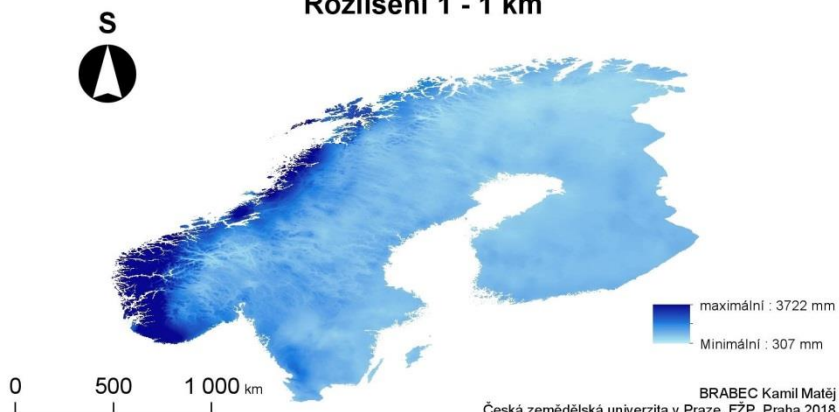
Minimální Teplota nejchladnějšího měsíce

Rozlišení 1 - 1 km



Roční srážky

Rozlišení 1 - 1 km



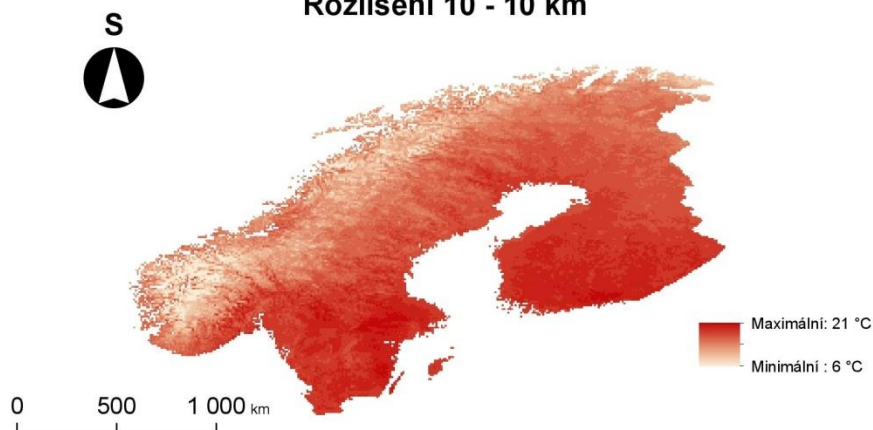
BRABEC Kamil Matěj
Česká zemědělská univerzita v Praze, FZP, Praha 2018
Software: ArcGIS 10.3

Souřadnicový systém: WGS 1984

Zdroj dat: WorldClim - Global Climate Data: Free climate data for ecological modeling and GIS
Dostupné z: www.worldclim.org

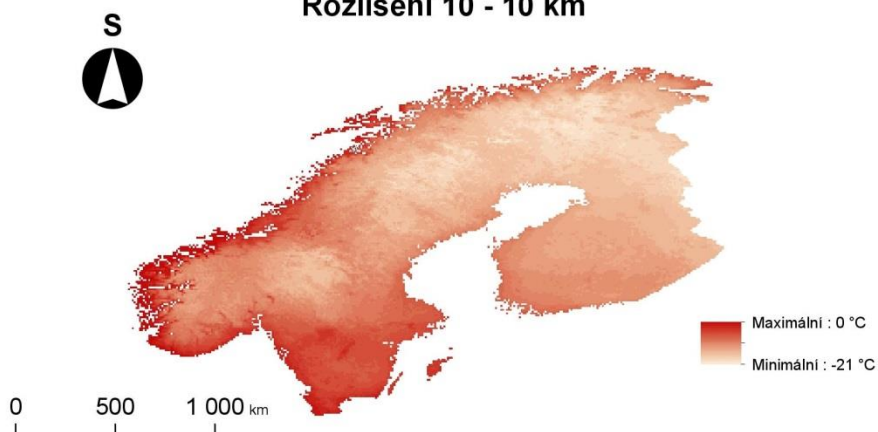
Maximální teplota nejteplejšího měsíce

Rozlišení 10 - 10 km



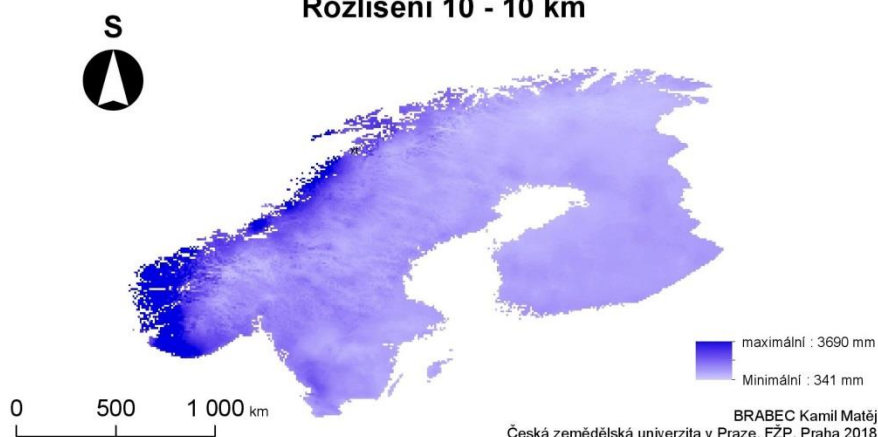
Minimální Teplota nejchladnějšího měsíce

Rozlišení 10 - 10 km



Maximální teplota nejteplejšího měsíce

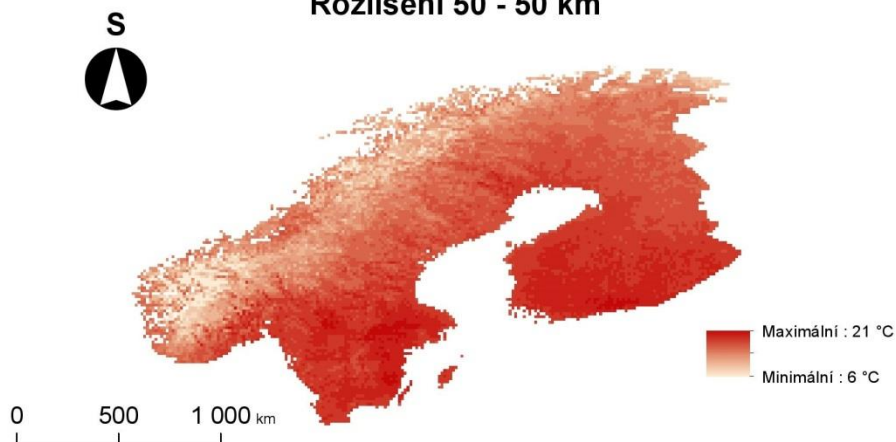
Rozlišení 10 - 10 km



BRABEC Kamil Matěj
Česká zemědělská univerzita v Praze, FZP, Praha 2018
Software: ArcGIS 10.3
Souřadnicový systém: WGS 1984
Zdroj dat: WorldClim - Global Climate Data: Free climate data for ecological modeling and GIS
Dostupné z: www.worldclim.org

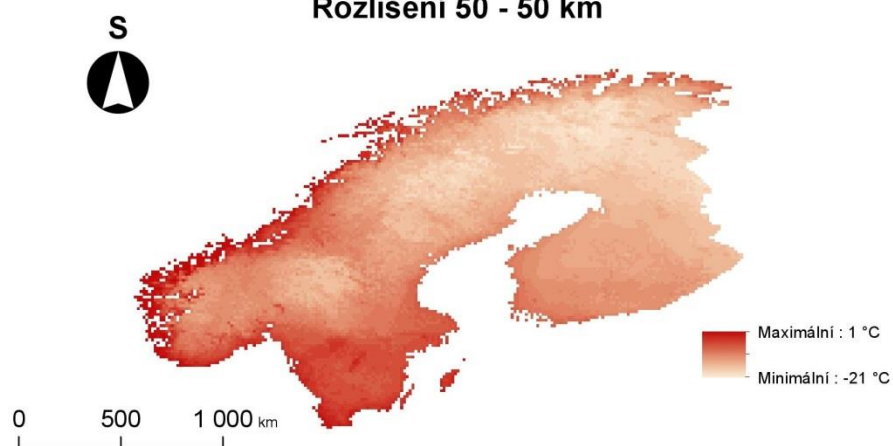
Maximální teplota nejteplejšího měsíce

Rozlišení 50 - 50 km



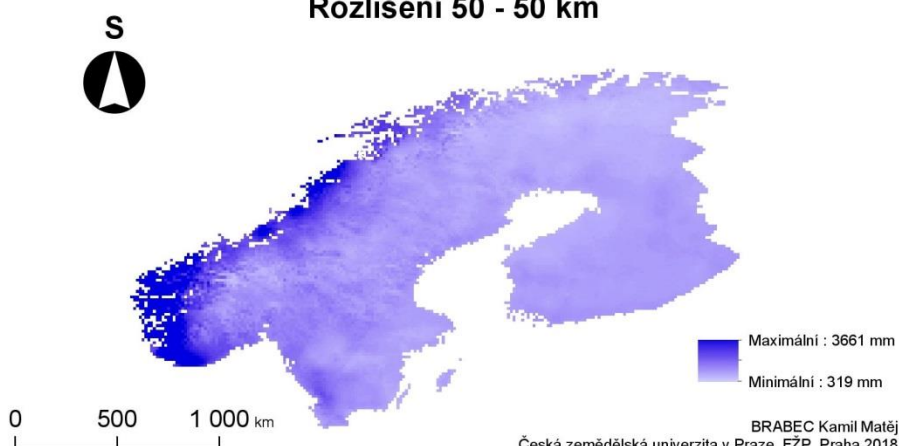
Minimální Teplota nejchladnějšího měsíce

Rozlišení 50 - 50 km



Roční srážky

Rozlišení 50 - 50 km



BRABEC Kamil Matěj
Česká zemědělská univerzita v Praze, FŽP, Praha 2018
Software: ArcGIS 10.3
Souřadnicový systém: WGS 1984
Zdroj dat: WorldClim - Global Climate Data: Free climate data for ecological modeling and GIS
Dostupné z: www.worldclim.org

Příloha C: Tabulka výsledků metody AUC

AUC

sample - size *	1 km *	10 km *	50 km *
50	0,5573	0,5560	0,5490
100	0,5820	0,5834	0,5807
500	0,6774	0,6797	0,6768
1000	0,7314	0,7111	0,7088
2000	0,7447	0,7223	0,7189

km * - environmentální rozlišení klimatických vrstev

sample-size * - počet prvku sample-size opakující se v 50. cyklech

Příloha D: Tabulka výsledků metody Schoener D indexu

D - index

sample - size *	1 km *	10 km *	50 km *
50	0,7836	0,7842	0,7843
100	0,7862	0,7893	0,7900
500	0,8098	0,8137	0,8134
1000	0,8313	0,8302	0,8308
2000	0,8452	0,8431	0,8425

km * - environmentální rozlišení klimatických vrstev

sample-size * - počet prvku sample-size opakující se v 50. cyklu