

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA APLIKOVANÉ GEOINFORMATIKY
A ÚZEMNÍHO PLÁNOVÁNÍ



ZKVALITŇUJÍ ENVIRONMENTÁLNÍ FILTRY MODELY
DRUHOVÉ DISTRIBUCE?

DIPLOMOVÁ PRÁCE

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

Diplomant: Bc. Lukáš Gábor

2016

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Lukáš Gábor

Aplikovaná ekologie

Název práce

Zkvalitňují environmentální filtry modely druhové distribuce ?

Název anglicky

Do environmental filters improve predictions of species distribution models ?

Cíle práce

Cílem práce je otestovat, zda environmentální filtrování presenčních dat vede ke spolehlivějším modelům druhové distribuce.

- Generování virtuálního druhu
- Vytvoření modelů druhové distribuce bez a s využitím environmentálního filtru
- Porovnání vytvořených modelů z hlediska predikované presence/absence a pravděpodobnosti výskytu

Metodika

Modelování druhové distribuce (Species Distribution Modelling, SDM) patří mezi oblíbené nástroje v ekologii a biogeografii. Jejich cílem je obvykle najít environmentální proměnné, které mohou mít významný vliv na rozšíření druhu nebo rozšíření druhu predikovat. Nicméně data použitá pro kalibraci těchto modelů mohou mít zásadní vliv na jejich přesnost. Zatímco s využitím reálně sbíraných dat je obtížné tyto vlivy kvantifikovat, generováním tzv. virtuálního druhu lze snadno určit, jaký vliv mají rozdílné metody sběru dat na výsledný model. Úkolem autora je zpracovat literární rešerši na téma generování virtuálních druhů a vytvořit virtuální druh, na jehož příkladu otestuje, zda environmentální filtrování vede k lepším modelům.

Doporučený rozsah práce

40 – 50 stran textu

Klíčová slova

AUC, druhové distribuční modely, filtrování, pravděpodobnostní přístup, virtuální druh

Doporučené zdroje informací

- BORIA, Robert A., Link E. OLSON, Steven M. GOODMAN a Robert P. ANDERSON. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*. 2014, 275, 73-77. DOI: 10.1016/j.ecolmodel.2013.12.012. ISSN 03043800.
- LEROY, Boris, Christine N. MEYNARD, Céline BELLARD a Franck COURCHAMP. Virtualspecies, an R package to generate virtual species distributions. *Ecography*. 2015, , 001-009. DOI: 10.1111/ecog.01388. ISSN 09067590.
- MEYNARD, Christine N. a David M. KAPLAN. The effect of a gradual response to the environment on species distribution modeling performance. *Ecography*. 2012, 35(6), 499-509. DOI: 10.1111/j.1600-0587.2011.07157.x. ISSN 09067590.
- MILLER, Jennifer. Species Distribution Modeling. *Geography Compass*. 2010, 4(6), 490-509. DOI: 10.1111/j.1749-8198.2010.00351.x. ISSN 17498198.
- MOUDRÝ, Vítězslav. Modelling species distributions with simulated virtual species. *Journal of Biogeography*. 2015,42(8). DOI: 10.1111/jbi.12552. ISBN 10.1111/jbi.12552.
- VARELA, Sara, Robert P. ANDERSON, Raúl GARCÍA-VALDÉS a Federico FERNÁNDEZ-GONZÁLEZ. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*. 2014, , no-no. DOI: 10.1111/j.1600-0587.2013.00441.x. ISSN 09067590.

Předběžný termín obhajoby

2015/16 LS – FŽP

Vedoucí práce

Ing. Vítězslav Moudrý, Ph.D.

Garantující pracoviště

Katedra aplikované geoinformatiky a územního plánování

Elektronicky schváleno dne 29. 2. 2016

doc. Ing. Petra Šímová, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 7. 3. 2016

prof. RNDr. Vladimír Bejček, CSc.

Děkan

V Praze dne 12. 03. 2016

Prohlášení

Prohlašuji, že jsem diplomovou práci na téma *Zkvalitňují environmentální filtry modely druhové distribuce?* vypracoval samostatně, pod vedením Ing. Vítězslava Moudrého, Ph.D., a že jsem uvedl veškeré literární prameny a publikace, ze kterých jsem čerpal.

V Praze dne 10. 4. 2016

.....

Poděkování

Chtěl bych poděkovat Ing. Vítězslavu Moudrému, Ph.D. za velkou míru trpělivosti, odborné vedení a cenné rady, bez kterých by nebylo možné tuto práci zkompletovat. Mé díky patří také celé rodině za důvěru, trpělivost a podporu během celého studia a všem ostatním, kteří přispěli ke vzniku této práce.

Abstrakt

Druhové distribuční modely (SDM) představují široce používaný nástroj v biogeografii, makroekologii a ochraně přírody, který se s postupným rozvojem stal důležitým prostředkem, využívaným například při určení lokalit potenciálně ohrožených invazními druhy nebo studování vlivu klimatických změn na biodiverzitu. S postupujícím rozvojem začíná být zřejmé, že jedním ze zásadních limitujících faktorů modelování druhové distribuce jsou vstupní data. Nejsnáze dostupná jsou presenční data, která však trpí nerovnoměrným sběrem – například s převahou záznamů na snadno dostupných lokalitách. Cílem této práce je ukázat, že popularitu získávající klimatické filtrování vstupních prezenčních dat za účelem odstranění nerovnoměrného samplování ovlivňuje negativním způsobem výsledný model. Pro tento účel byly na území Iberského poloostrova vygenerovány virtuální druhy s různou druhovou prevalencí a rozdílným počtem zaznamenaných výskytů. Následně byly vytvořeny pomocí nástroje Maxent druhové distribuční modely s a bez klimatického filtru, které byly vyhodnoceny pomocí AUC. Rozdíl mezi virtuální realitou, která je prezentována vhodností prostředí virtuálního druhu a výsledným modelem byl testován párovým T testem. Komparace AUC potvrdila, že druhové distribuční modely založené na klimatickém filtrování mají lepší diskriminační schopnost. Nicméně to jen ukazuje na dovednější práci s vybraným vzorkem dat, který již nereflektuje realitu. Oproti tomu srovnání rozdílů mezi realitou a modelem s a bez klimatického filtrování pomocí párového T testu ukazuje vyšší kongruenci mezi nefiltrovaným modelem a realitou. Tím bylo dokázáno, že klimatické filtrování nevede k vyšší validitě druhových distribučních modelů.

Klíčová slova: AUC, druhové distribuční modely, filtrování, pravděpodobnostní přístup, virtuální druh

Abstract

Species distribution models (SDM) are widely used tool in biogeography, macroecology and nature conservation. With gradual development, it has become an important means used by, for example, in determining the potentially threatened locations by invasive species, or studying the impact of climate change on biodiversity. With the progressive development it becomes obvious that one of the major factors limiting the species distribution modelling are input data. The presence data are most readily available, but they suffer from an uneven collection - for example, with a predominance of records in easily accessible locations. The aim of this work is to show, that popular climate filtering of presence data input, in order to eliminate uneven sampling, affects the final model in a negative way. For this purpose, there were virtual sorts of different species and different prevalence of recorded occurrences on the territory of the Iberian Peninsula generated. Subsequently, species distribution models with and without climate filters were created by using Maxent. They were evaluated by AUC. The difference between virtual reality, which is presented to the suitability of the virtual species, and the resulting model was tested by paired T test. Comparison of the AUC confirmed that the species distribution models based on climate filtering have better discriminative ability. However, it only points to the skilful work with the selected sample bias that already does not reflect reality. In contrast, comparison of the differences between virtual reality and the models with and without climate filtering using a paired T test shows greater congruence between unfiltered models and virtual reality. Thus it was proved that the climate filtering does not lead to higher validity species distribution models.

Keywords: AUC, species distribution models, filtering, probabilistic approach, virtual species

Obsah

Seznam použitých zkratk a akronymů	9
1 Úvod	10
2 Literární rešerše	12
2.1 Druhov \acute{e} distribuční modely (SDMs)	12
2.2 Ekologická nika a biotické interakce.....	14
2.3 Vstupní data.....	17
2.3.1 Biologická data.....	18
2.3.2 Environmentální data	19
2.3.3 Klimatická data	19
2.3.4 Topografická data.....	20
2.3.5 Chyby v datech.....	21
2.4 Metody modelování.....	22
2.5 Validace SDMs.....	24
2.6 Virtuální druhy	27
2.6.1 Využití virtuálních druhů při modelování druhové distribuce.....	27
2.6.2 Komparace prahové metody a pravděpodobnostního přístupu.....	28
2.6.3 Vliv graduální reakce druh – prostředí na progres modelu.....	31
3 Metodika	33
3.1 Softwarové požadavky a vstupní data	33
3.1.1 Použitý software.....	33
3.1.2 WorldClim	33
3.2 Definování virtuálního druhu	33
3.3 Data výskytu.....	36
3.4 Klimatické filtrování dat	38
3.5 Druhov \acute{e} distribuční modely	39
4 Výsledky práce	40

5	Diskuze	43
6	Závěr	45
	Přehled literatury a použitých zdrojů	46
	Seznam obrázků	52
	Seznam tabulek	53
	Seznam příloh	54

Seznam použitých zkratk a akronymů

ANN	Umělá neurální síť (Artificial neural network)
AUC	Plocha pod křivkou (Area under cover)
CART	Klasifikační a regresní stromy (Classification and Regression Trees)
CCR	Správná klasifikační míra (Correct classification rate)
ČGS	Česká geologická služba
ENFA	Analýza faktorů ekologické niky (Ecological niche factor analysis)
ENMs	Modely ekologické niky (Ecological niche models)
GA	Genetické algoritmy (Genetic algorithm)
GAM	Generalizovaný aditivní model (Generalized additive model)
GIS	Geografický informační systém (Geographic 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)
PELCOM	Celoevropský monitoring využívání půdy a složení krajiny (Pan-European Land Use and Land Cover Monitoring)
PPP	Pozitivní prediktivní síla (Positive predictive power)
ROC	Operační křivka (Receiver operating curve)
SDMs	Druhové distribuční modely (Species distribution models)
UNEP	Environmentální program spojených národů (United Nations Environment Programme)

1 Úvod

Druhové distribuční modely reprezentují klíčový nástroj modelování v biogeografii a makroekologii. Sofistikované statistické metody využívané při tvorbě modelů umožňují predikovat druhové rozšíření v prostoru vlivem současných, ale i budoucích klimatických podmínek (Meynard et al. 2013). Jak ukázaly některé studie (například Duputié et al. 2014), přesná data představují základní předpoklad validních modelů. Typ dostupných dat, jejich nedostatečná kvalita (Guillera-Arroita et al. 2015) a obtížný vývoj adekvátních kontrolních metod určujících přesnost modelů představují nejvíce limitující faktory v aplikaci druhových distribučních modelů (Meynard et al. 2013). Řešením je využití virtuálního druhu s uměle stanovenou odpovědí druhu na environmentální gradienty. Díky tomu lze kontrolovat vzorec druhového výskytu, což umožňuje provádět modelování ve větším měřítku (Zurell et al. 2010) nebo zkoumat vliv filtrovacích metod redukujících počet míst s druhovým výskytem (Valera et al. 2014).

Diskutovaným tématem při generování virtuálního druhu je použití prahové metody či pravděpodobnostního přístupu. I přes častou aplikaci prahové metody (například Hirzel et al. 2001; Real et al. 2006; Bombi & D'Amen 2012; Valera et al. 2014) dokazují nedávné studie (Meynard & Kaplan 2012; Moudrý 2015 etc.), že je pro účely modelování druhové distribuce výhodnější užívat metodu pravděpodobnostního přístupu. Prahová metoda jak ukazuje Meynard & Kaplan (2012) totiž představuje extrémní případ, kdy je druh bez výjimky reprezentován pouze na jedné straně environmentálního gradientu. V této práci proto bude druhová prezence vygenerovaná pomocí pravděpodobnostní metody a to (i) se strmější křivkou simulující použití prahové metody a (ii) s křivkou s pozvolným nárůstem. To umožní otestovat efekt graduální křivky korigující reakci druh – environmentální prostředí na výkon modelu.

K ověření validity modelů jsou používány diskriminační a kalibrační metody. Preferovanější diskriminační metody (sensitivita, specificita, kappa, AUC etc.) ukazují schopnost modelu rozlišit prezence absence. Oproti tomu kalibrační metody ilustrují, do jaké míry model predikuje pravděpodobnost výskytu v porovnání s použitými daty druhového výskytu (Giancristofaro & Salmaso 2003). V nedávné studii, Valera et al. (2014) použila k validaci modelu založeného na práci s virtuálním druhem diskriminační metodu AUC a v kombinaci s aplikací klimatického filtrování dosáhla velmi vysokých hodnot AUC ($> 0,9$). Moudrý (2015) nicméně poukazuje na fakt, že

filtrovaná data jsou bez ohledu na jejich charakteristiku vždy nahodilá a jejich použití není vhodné například při studiích zaměřených na ochranu přírody. Proto je podstatné při vyhodnocování modelů nespolehat pouze na vysoké AUC, jehož progresi pozitivně ovlivňuje již zmíněné filtrování dat (klimatické, geografické etc.), ale je vhodné zároveň použít také kalibrační metody, založené na komparaci výsledných modelů se vstupními daty (Lawson et al. 2014).

Cílem práce je navázat na teoretický článek *Modelling species distributions with simulated virtual species* (Moudrý 2015) a na praktické studii potvrdit, že aplikace klimatického filtrování na prezenční data nevede k validnějším modelům druhové distribuce. K prokázání teorie je nezbytné: (i) vygenerovat virtuální druh, (ii) vytvořit modely druhové distribuce s a bez aplikace klimatického filtrování a (iii) porovnat vytvořené modely z hlediska predikované prezenze – absence a pravděpodobnosti výskytu.

2 Literární rešerše

2.1 Druhové distribuční modely (SDMs)

Predikční modely mají dlouhou historii. První pokusy o vytvoření modelů předpovídajících prostorové rozšíření druhů můžeme datovat již od počátku 60. let 20. století. Klíčové období jejich rozvoje nastalo přibližně v polovině 80. let 20. století, kdy došlo k rozmachu nejen počítačových technologií, ale zejména metod dálkového průzkumu země a geografických informačních systémů. Ty přinesly rozvoj zásadních nástrojů pro prostorovou analýzu, ale i nové, přesnější a podrobnější prostorové podklady, usnadňující tvorbu map a distribučních modelů. Miller (2010) popisuje distinkci mezi slovem mapa a prostorový model jako jemný rozdíl od produktu (mapa) k procesu (prostorový model).

Druhové distribuční modely (z anglického originálu Species distribution models, dále jen SDMs) také nazývané modely ekologické niky (Stockwell 1999), modely klimatické niky (Araújo & Peterson 2012) nebo prediktivní modely habitatové distribuce (Guisan a Zimmermann 2000) představují fundamentální nástroj používaný ekology, biogeografy a institucemi ochrany přírody na celém světě. Druhové distribuční modely umožňují kvantifikovat vztah mezi environmentálními faktory a distribucí druhů, který může být použit k popisu a určení důležitosti specifických faktorů jednotlivých druhů a predikci druhové distribuce (Miller 2010). Toho lze využít v oblastech, kde nedošlo k mapování výskytu zájmových druhů nebo k predikci změn v rozšíření druhů vlivem klimatických změn (Elith & Leathwick 2009). Dále lze modelování druhové distribuce aplikovat při plánování ochrany přírody (Araújo & Guisan 2006), k potvrzení biogeografických hypotéz (Leathwick et al. 2005) a dle Edwads (2005) i k detekci nových populací vzácných druhů (další využití viz Tab. 1).

Proces modelování druhové distribuce se skládá ze šesti vzájemně navazujících kroků (Guisan & Zimmermann 2000): (i) vytvoření koncepce, (ii) příprava dat, (iii) přizpůsobení modelu zkoumanému druhu, (iv) evaluace modelu, (v) prostorová predikce a (vi) posouzení aplikovatelnosti modelu. Mnoho rozhodnutí musí být učiněno již v počáteční fázi přípravy modelu. Guisan & Thuiller (2005) rozlišují dvě zásadní oblasti přípravy modelu: (i) teorie a data: definování modelu od vstupních dat jeho interpretaci na základě definovaných cílů a empirických zkušeností, vytvoření pracovních hypotéz, posouzení dostupných i nedostupných dat a relevance

environmentálních prediktorů v závislosti na zkoumaném druhu a měřítku, definice postupů doplňkového sběru dat nebo doplnění dat z jiných zdrojů, určení geografického měřítka. (ii) Metody modelování: výběr vhodných algoritmů, statistických metod umožňujících evaluaci modelu a stanovení jeho přesnosti. Například výběr vhodného měřítka závisí na velikosti původního rozšíření druhu a jeho využití zdrojů v krajině, ale také na znalosti environmentálních gradientů ve studované oblasti (Austin 2002). Při modelování je důležité, zohlednění odlišnosti v ročním období nebo rozdílnosti pohlaví (Guisan & Thuiller 2005 ex Kearney & Porter 2004). V současné době, se však na začátku výzkumu rozhoduje jen malá část z výše uvedeného vlivem nedostatečných znalostí o některých cílových druzích, což vede k modelům s nízkou prediktivní schopností (Guisan & Thuiller 2005).

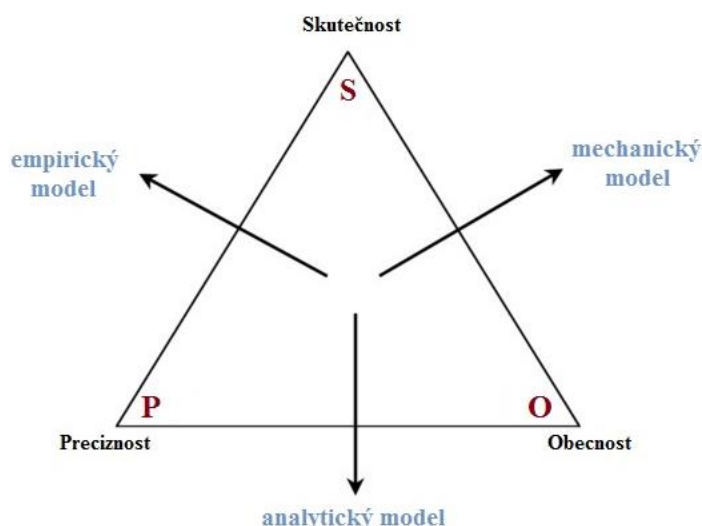
Tab. 1 Příklady možného využití SDMs v ekologii, biogeografii a ochraně přírody

Aplikace SDMs	Zdroj
kvantifikace environmentální niky	Austin et al. (1990), Vetaas (2002)
testování biogeografických, ekologických a evolučních hypotéz	Leathwick (1998), Anderson et al. (2002), Graham et al. (2004)
posouzení druhových invazí a jejich rozšíření	Beerling et al. (1995), Peterson (2003)
posouzení impaktu změn klimatu, využití krajiny a jiných environmentálních proměnných na druhovou distribuci	Thomas et al. (2004), Thuiller (2004)
podpora tvorby vhodných plánů na obnovu druhového rozšíření a mapování vhodných lokalit k reintrodukci druhů	Pearce & Lindenmayer (1998)
podpora ochrany přírody, příprava vhodného managementu a výběr vhodných lokalit k územní ochraně	Ferrier (2002), Araújo et al. (2004)
budování biogeografických nebo ekogeografických regionů	Guisan & Theurillat (2005)
modelování druhového složení na základě individuálních druhových prediktorů	Leathwick et al. (1996), Ferrier et al. (2002), Guisan a Theurillat (2000)
zdokonalení výpočtu vhodných vzdáleností mezi ekologickými oblastmi, větší dynamika modelů řešících genetický tok	Guisan & Thuiller (2005)

(Zdroj: Guisan & Thuiller 2005)

Kontrola metodiky použitých statistických metod a teorie by měla být součástí každého kroku tvorby SDMs. Pomocí ní lze včas odhalit nedostatky vedoucí k nepřesným výsledkům (například nadhodnocení nebo podhodnocení modelu ...) a využít je k vylepšení modelu (Guisan & Zimmermann 2000).

Levins (1966) klasifikuje modely do tří kategorií: mechanické, empirické a analytické (matematické) modely (Obr. 1). Levinsovo (1966) dělení, rozvedli Guisan & Zimmermann (2000), kteří tvrdí, že je možné dosáhnout pouze dvou ze tří žádoucích vlastností modelu (skutečnost, preciznost, obecnost). Tento fakt je respektován dodnes, protože jak upozorňuje Kienast et al. (2012) žádný model nemůže simultánně dosahovat vysokého hodnocení (preciznosti), být založen na fyziologických procesech či mechanismech (skutečnost) a být všeobecně aplikovatelný u modelů, které se od sebe liší podmínkami a okolnostmi (obecnost).



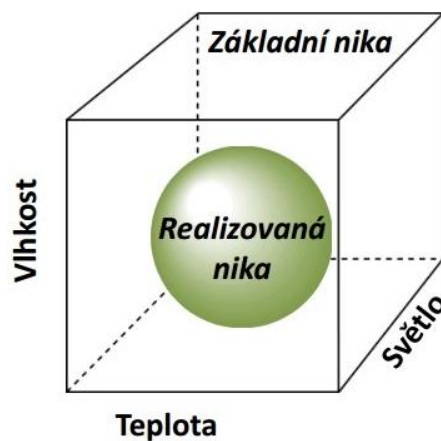
Obr. 1 Odlišnost analytického, empirického a mechanického modelu

(Zdroj: Guisan & Zimmermann 2000)

2.2 Ekologická nika a biotické interakce

Ekologická nika tvoří teoretický základ pro popis vztahu mezi druhem a environmentálními veličinami. Vztah mezi druhem a jeho prostředím je fundamentální otázkou v ekologii a v posledních letech se ke zjištění tohoto vztahu využívá téměř výhradně SDM. Navzdory faktu, že se jedná o základní ekologický pojem, existuje mnoho různých definic, které se pohybují na obecné úrovni. Příkladem může být definice, která charakterizuje niku jako „výsek“ prostředí obývaného jedním druhem v rozmezí hodnot ekologických faktorů. Pro tvorbu SDMs je zásadní definice G. E. Hutchinsona (1961), který definoval niku jako konkrétní n-rozměrný nadprostor,

v environmentálním prostředí, kde může konkrétní druh existovat. Svou definici Hutchinson dále rozvedl a rozčlenil ji na fundamentální (základní) niku popisující rozsah přírodních podmínek, kde je druh přirozeně schopný žít a skutečnou niku (realizovanou), která ilustruje reálné rozšíření druhu v přírodě. Realizovaná nika (rozdíl mezi základní a realizovanou nikou je znázorněn na Obr. 2) je pak podle Hutchinsona (1961) tvořena podmnožinami základní niky jako výsledek biotických interakcí (predace, symbióza etc.). Ackerly (2003) definuje ještě niku potenciaální jako podmnožinu základní niky, která není omezená prostředím. Domnívá se totiž, že ne všechny možné kombinace environmentálních proměnných existují ve zkoumané oblasti nebo dokonce na Zemi. Zjednodušeně lze říci, že SDMs kvantifikují fundamentální niku definovanou Hutchinsonem (Guisan & Thuiller 2005).

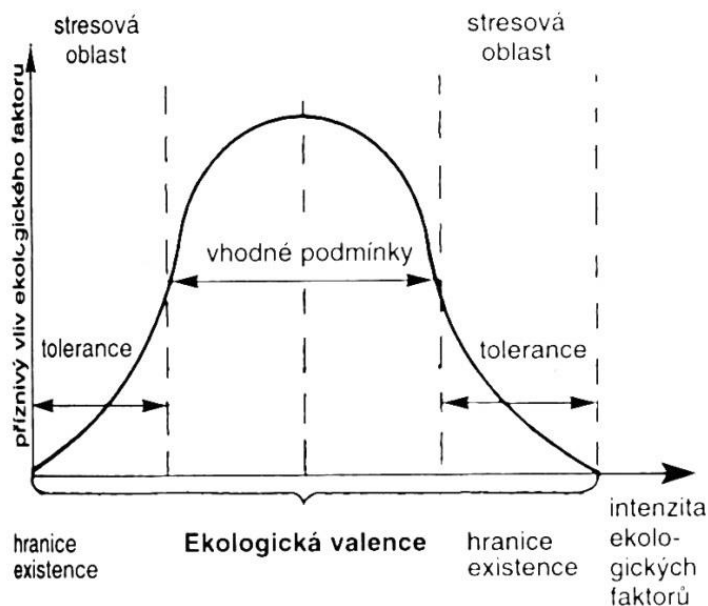


Obr. 2 Komparace základní a realizované niky v n -rozměrném prostoru

(Zdroj: Václavík 2013)

Další autoři (Guisan et al. 2006; Soberón 2007) poukazují na fakt, že Hutchinsonova (1961) definice nezahrnuje složitost některých pozitivních interakcí což je limitujícím faktorem prostorových modelů, který nebyl vyřešen a striktně omezuje výsledky studií používajících SDMs. Tyto teoretické limity prostorových modelů je potřeba si ujasnit před začátkem modelování a při interpretaci výsledků s nimi kalkulovat (Miller 2010). Každá environmentální proměnná, se kterou při tvorbě SDMs pracujeme, tvoří samostatnou dimenzi v n -rozměrném prostoru definovaném dle Hutchinsona (1961) a společně s odezvou druhu na tyto proměnné tvoří jeho distribuci v prostoru (Miller 2010). Vliv těchto proměnných na výskyt druhu lze znázornit například pomocí Gaussovy křivky (dále také pomocí kvadratické, lineární, logistické funkce etc.), jež umožňuje vystihnout ekologickou prevalenci druhu (viz Obr. 3). Tvar křivky nám ukazuje reakci druhu na změnu gradientu a lze z ní vyčíst fyziologickou toleranci,

prahové hodnoty výskytu nebo optimální hodnotu proměnné, která ovlivňuje výskyt druhu ve zkoumaném území.



Obr. 3 Ukázka Gaussovy křivky a od ní se odvíjející druhové prevalence

(Zdroj: Václavík 2013)

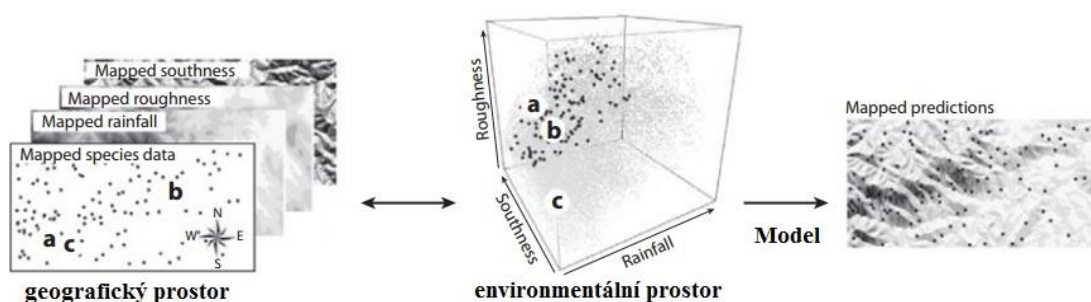
Použité gradienty prostředí tvoří podstatný aspekt při tvorbě SDMs. Austin (1980) kategorizoval všechny environmentální proměnné těchto prediktorů v závislosti na jejich biologické kauzalitě do tří různých kategorií:

1. Zdrojové proměnné – důležité pro vývoj organismů (světlo, voda, živiny, ...)
2. Přímé proměnné – mají přímý vliv na vývoj organismů, ale nespádají do první kategorie (teplota, vlhkost vzduchu, pH ...)
3. Nepřímé proměnné – nemají přímý vliv na vývoj organismů, avšak často korelují se zdrojovými a přímými proměnnými (nadmořská výška, sklon terénu, geologie ...)

Soberón (2007) jejich použití odlišuje dle použitého geografického měřítka. Zdrojové proměnné důležité pro vývoj organismů, navrhuje použít při práci menšími územními celky (respektive takovými územními celky, pro které máme data v podrobném rozlišení). Přímé proměnné by se měly využívat pro modely pracující s většími územními celky (územní celky, pro která jsou k dispozici pouze data v hrubém rozlišení) a měly by popisovat druhový habitat. Nepřímé proměnné pak logicky vycházejí z korelačního vztahu organismu a lokality výskytu druhu.

Později byly tyto typy proměnných také klasifikovány jako proximální (zdrojové) a distální (nepřímé) s gradienty ležícími mezi těmito třídami (přímé), které se v závislosti na vlivu gradientu na organismus přiklání buď k distálním či proximálním proměnným (Kienast et al. 2012).

Miller (2010) objasňuje, jak se ekologické niky a s nimi i výše popsané gradienty, přetvářejí v průběhu modelování SDMs: ekologická nika popisuje fitness druhu v environmentálním prostoru, kvantitativní statistické metody umožňují popsání vztahu druh – prostředí a výsledná predikční mapa převádí vztah druh – prostředí do geografického prostoru (viz Obr. 4).



Obr. 4 Vztah mezi geografickým prostorem, environmentálním prostorem a mapovou predikcí využívající pouze environmentální prediktory. Vzdálenosti v geografickém prostoru mohou být odlišné od těch v environmentálním prostoru, protože při transformaci environmentálních prediktorů dochází k prostorové autokorelaci.

(Zdroj: Elith & Leathwick 2009)

Při tvorbě SDMs zejména při výběru prediktivních proměnných je potřeba si ujasnit, co bude účelem studie. Chceme dosáhnout precizně vytvořeného modelu (přesné predikce) nebo biologické hodnověrnosti (vysvětlení vztahu druh – prostředí)? Model sloužící k návrhu ochranných opatření bude vyžadovat jiná vstupní data a typ modelu než model, který bude sloužit k odhadu vlivu klimatických změn na habitat druhů (Kienast et al. 2012).

2.3 Vstupní data

Druhovému distribučnímu modelování vyžaduje využití prostorových dat reprezentujících environmentální proměnné a záznamy o pozorování druhů (Elith & Leathwick 2009). Ferrier & Guisan (2006) dělí vstupní data na biologická (biologické průzkumy) a environmentální proměnné odvozené z dálkového průzkumu. Miller (2010) environmentální data rozlišuje na klimatická, topografická a ostatní.

V současnosti existují dvě metody konceptualizace vlastností reálného světa (rozložení zájmových dat v prostoru a čase): (i) koncepce diskrétních objektů umožňuje zobrazit počítatelné jevy v prostoru nebo pohybu, je vhodná zejména při práci s organismy a k jejich zobrazení slouží body, linie a plochy, (ii) koncepce kontinuálních jevů umožňuje zobrazit prvky, které mají nejasně stanovené hranice (eroze půdy, teplota etc.). Obě koncepce mohou být reprezentovány vektorovými nebo rastrovými daty (Goodchild 2011). V případě rastru je rozlišení modelů závislé na velikosti buněk. Při použití vektorových dat se rozlišení odvíjí od vzdálenosti mezi jednotlivými body.

Využití rastrových nebo vektorových dat je v praxi závislé na existujících datech, schopnosti softwaru pracovat s různými datovými formáty, typu analýzy a specifikaci modelu. Goodchild (2011) upozorňuje na fakt, že většina dostupných vektorových dat má nízkou kvalitu rozlišení a proto v případě vědeckých výzkumů pracovat s rastrovými daty.

2.3.1 Biologická data

Systematické terénní průzkumy výskytu druhů, ideálně založené na systematickém sběru dat (například stratifikovaný sběr) jsou dle Austina (1998) nejcennějším zdrojem biologických dat sbíraných napříč celým světem. Z globálního hlediska jsou ale nedostatečná. V minulosti byla pozornost vědců zaměřena na muzejní sbírky, které jsou obsáhlým a cenným zdrojem biologických dat. Engler et al. (2004), ale i Zaniwski et al. (2002) poukazují na jejich omezené využití při tvorbě SDMs a to s ohledem na fakt, že jsou takto získaná data založená pouze na presenci, kdežto většina algoritmů je založena na prezenčně – absenčních datech.

Dnes jsou modely druhové distribuce založeny na globálních databázích (příklad zdrojů viz Tab. 2), které nicméně obsahují pouze prezenční data. Řešením je generování „pseudo-absence“ vycházející z prezenze druhu. Tím lze získat nejlepší základ pro tvorbu modelů, při kterých se využívají biologická data založená na prezenci (Ferrier & Guisan 2006). Miller (2010) nicméně upozorňuje, že techniky používající „pseudo-absenci“ jsou předmětem výzkumů až v poslední době a jejich efekt na výsledný model není ještě plně popsán.

Při práci s biologickými daty je potřeba brát v úvahu potencionální chybovost. Ta může vyplývat z nejasného původu dat nebo s detekovatelností živočichů spojenou

s jeho vysokou mobilitou nebo krypticitou. V případě rostlin sezonalitou či vývojovou fází (Edwards et al. 2005). V důsledku abiotických interakcí (geografické bariéry, disturbance etc.), budou navíc, správně použitá data absence druhu, vyskytující se ve vhodném habitatu problematická při kalibraci modelu (Miller 2010).

Tab. 2 Příklad zdrojů biologických dat používaných při tvorbě SDMs

Organizace	URL	Dostupná data
Global Biodiversity Information Facility	http://data.gbif.org/welcome.htm	rostliny, živočichové, houby, mikroorganismy
National Invasive Species Information Center	http://www.invasivespeciesinfo.gov/	databáze invazivních druhů na území Spojených států
NatureServe	http://www.natureserve.org/	rostliny, živočichové a ekosystémy v Jižní a Severní Americe
North American Breeding Bird Survey	http://www.pwrc.usgs.gov/BBS/	data sbíraná od roku 1966 s více jak 400 druhy ptactva
Agentura ochrany přírody a krajiny ČR	http://portal.nature.cz/	nálezová databáze druhů na území ČR

(Zdroj: Miller 2010, autor)

2.3.2 Environmentální data

Environmentální data (příklad zdrojů viz Tab. 3) jsou složena z vrstev vzniklých využitím dat získaných dálkovým průzkumem Země a aplikací nástrojů GIS, jež musí podle Miller (2010) představovat vhodnou kombinaci přímých, nepřímých a zdrojových proměnných. Jako opětovně používané prediktory při tvorbě SDMs jsou napříč odbornou veřejností aplikovány klimatické a topografické proměnné, jejichž kombinace dokáže vystihnout fyziologickou závislost druhů na vodu, teplotu a ve větších měřítkách i druhovou variaci, závislost na vlhkosti prostředí nebo tok energií v rámci ekosystému (Kienast et al. 2012).

2.3.3 Klimatická data

Klimatická data mohou reprezentovat souhrn, jak přímých (teplota), tak zdrojových (vlhkost) proměnných, jako souhrn těchto proměnných (průměrné srážky) nebo z ekologického pohledu více komplexní a relevantnější proměnné, jako například sezónní maxima (Miller 2010). Parra et al. (2004) považuje klima za určující faktor druhové distribuce.

Dostupné klimatické vrstvy vznikají interpolací dat získaných z pozemních meteorologických stanic. I přes pokročilé metody dálkového průzkumu jsou ovšem pro některé oblasti data dostupná pouze v omezené míře a kvalitě. V roce 2005 Hijmans et al. vytvořil volně dostupný soubor globálních klimatických dat. Ten je k dispozici v různém rozlišení a obsahuje celkem 19 klimatických vrstev zahrnujících každoroční trendy, sezónní či extrémní limitující faktory (Miller 2010).

2.3.4 Topografická data

Jednoduché topografické faktory reprezentující především nepřímé proměnné prostředí (nadmořská výška, teplota, sklon ...), se mohou projevit v kombinaci s přímými a zdrojovými proměnnými v nižší prediktivní přesnosti modelu. Pokud se, ale použijí v závislosti na zkušenostech se zkoumaným druhem, mohou mít na výsledný model vliv pozitivní (Guisan & Zimmermann 2000). Miller (2010) vidí přínos v možnosti odstranění anomálií v druhovém rozšíření, vzniklých působením mikroklimatu, jenž má vliv na dostupnost vody, teplotu nebo sluneční záření.

Z toho lze odvodit, že je lepší používat jako vstupní data pouze biologické a klimatické proměnné, které tvoří základ SDMs. Van Niel et al. (2004) však upozorňuje, že to není vždy pravidlem. V jeho studiích dokázal, že v případě použití komplexních topografických proměnných dosahují modely menší chybovosti ve srovnání s modely bez topografických proměnných nebo těmi s použitím jednoduchých topografických proměnných.

Tab. 3 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 x 1 km
PRISM	http://www.prism.oregonstate.edu/index.phtml	klima Spojených států (měsíční, roční, historické) v rozlišení 800 x 800 m
UNEP	http://www.grid.unep.ch/data/data.php?category=lithosphere	typy půd pro celý svět
ČGS	http://mapy.geology.cz/geocr_50/	geologická mapa ČR včetně typů půd
PELCOM	http://www.geo-informatie.nl/projects/pelcom/	land cover pro území Evropy v rozlišení 1 x 1 km

(Zdroj: Miller 2010; autor)

Kromě výše uvedených klimatických a topografických dat lze využívat mnoho jiných vrstev reprezentujících další proměnné mající vliv na výskyt nebo distribuci druhu. Miller (2010) uvádí příkladem geologii území, půdní typ a druh (reprezentace dostupnosti potravy). Aplikací ArcGIS nebo jiného dostupného softwaru (MaxEnt, R) lze do modelů aplikovat specifické proměnné (vzdálenost k vodním plochám, urbánním plochám apod.). Osborne et al. (2001) upozorňuje na možnost počítání vzdálenosti k nějaké disturbanci nebo zdroji na kterém je druh závislý. Možné je také využití distribučních map jiných druhů, doplňujících dat získaných dálkovým průzkumem a map s využitím krajiny nebo vegetačními indexy.

2.3.5 Chyby v datech

Při práci se všemi druhy vstupních dat je vždy potřeba uvažovat nad jejich potenciální chybovostí, která by mohla mít negativní vliv na výsledný model. Například environmentální data nasbíraná v terénu mohou zachycovat jen část druhové niky (Welk 2004). Kromě výše uvedených faktorů, které mají vliv na biologická data (krypticita, vysoká mobilita etc.) se na jejich kvalitě negativně podepisuje také chybné sbírání dat v terénu. Důsledkem toho mohou vznikat místa zachycená mnoha vzorky (atraktivní území) a lokality bez jediného záznamu (Loiselle et al. 2007). Nevhodně použité geografické aspekty, absence nebo nevhodně aplikované environmentální prediktory povedou vždy ke špatné předpovědní schopnosti modelu.

Předem odhadnout, jaké důsledky budou mít použité environmentální a prostorové proměnné na vytvořený model, není vzhledem k množství existujících metod jednoduché, a tak většina modelů bude z nějakého důvodu chyby obsahovat. Schopnost tvorby precizních modelů, vytvořených aplikací správných matematicko-statistických modelů a zároveň užitečných i z biologického hlediska je otázkou dalšího výzkumu. Nicméně i model, který bude obsahovat chyby, může být užitečný v rámci studie, pro kterou je vytvářen, pokud bude jeho tvůrce schopen správně interpretovat jeho výsledky. To je ten nejdůležitější aspekt, který může přinést vzhled do problematiky i lidem, kteří se v dané oblasti neorientují.

Některé studie (Stockwell & Peterson 2002; Seo et al. 2009; Austin a Niel. 2011) ukazují důležitost přizpůsobení kvality respektive rozlišení vstupních dat v závislosti na studované oblasti. Proto je bez ohledu na počet vstupních dat či jejich chybovost důležité vybírat data ve vhodném rozlišení tak, aby nedocházelo ke zkreslení výsledků.

2.4 Metody modelování

V posledních 3 dekáдах vznikalo mnoho SDMs, a i přesto, stále výzkumníci obtížně hledají vhodný model pro zpracování svých dat a naplnění vytyčených cílů (Li & Wang 2013). Přes podobné výstupy odlišných algoritmů mezi nimi existuje rozdíl ve smyslu jejich využití. Výběr správného algoritmu není vzhledem k hojnému počtu modelů a možnostem aplikace mnoha různými způsoby jednoduchým procesem. Odlišné využití SDMs zmíněné v předchozích kapitolách vyžaduje aplikaci rozdílných modelů (Gutiérrez et al. 2012). Ty se logicky musí lišit výzkum od výzkumu a nemohou být jednotně evaluovány. Použitý algoritmus by se měl odvíjet od měřitelnosti a charakteristiky biologických dat, měřitelnosti a objemu environmentálních dat a očekávaného mapového výstupu (Miller 2010). Jiménez-Valverde et al. (2008) ilustruje odlišné použití algoritmů podle typu SDMs: (i) predikce skutečné oproti potenciální distribuci druhu, (ii) predikce pravděpodobnosti výskytu versus vhodnosti výskytu nebo (iii) použití dat prezenčních versus prezenčně – absenčních. Používané algoritmy dělíme na regresivní (parametrické) a neparametrické. (Aguirre-Gutiérrez et al. 2013). Modely pak na pracující pouze s prezenčními nebo prezenčně – absenčními daty (Miller 2010).

Modely založené na bázi regresní analýzy fungují na principu kombinace různých variací druhového výskytu či abundance v závislosti na prostředí jeho výskytu a využívají prezenčně – absenční data (Elith & Leathwick 2009). Jejich výstupem je pravděpodobnost výskytu či abundance druhu. Příkladem modelů založených na regresní analýze je GLM (generalized linear model) včetně jeho neparametrické extenze (v mnoha studiích kombinováno s kvadratickou nebo kubickou transformací), GAM (generalized additive models) a MARS (multivariate adaptive regression models). GAM a MARS jsou více flexibilní oproti GLM a umožňují přesněji určit křivku vyjadřující vztah druh – prostředí (Miller & Franklin 2002). Leathwick et al. (2005) uvádí, že MARS je v porovnání s GAM a GLM rychlejší po výpočetní stránce a výsledky je možné snadněji konvertovat do geografického prostoru. V porovnání s ostatními metodami, je nevýhodou regresivních analýz, vyšší ovlivnění prostorovou autokorelací, jež má negativní vliv na výslednou kvalitu modelu (Miller et al. 2007). Na druhou stranu je tento typ modelů využíván díky jejich schopnosti realistického popisu ekologických vztahů v SDMs, schopnosti zpracování dat s nedokonale zaznamenanou prezencí (abundance zahrnující data s mnoha nulami vyjadřujícími, že

druh nebyl zaznamenán) a možnosti přidávat kombinace různých prediktorů a jejich interakcí (Elith & Leathwick 2009).

Metody s využitím neparametrických algoritmů se využívají při studiích zaměřených na predikci budoucího rozšíření druhu (Elith & Leathwick 2009). Jsou založeny na řízené klasifikaci zpracovávaných dat a jsou vhodné k explorativním účelům. Většina z nich pracuje na principu dělení dat do několika podmnožin, z nichž průměrem počítá výsledky (Miller 2015 ex Franklin & Miller 2010). Miller (2010) uvádí jako příklady ANN (artificial neural network), CART (classification and regression trees) nebo GA (genetic algorithm). ANN je iterativní metoda, umožňující implementovat ostatní používané algoritmy (nejčastěji se používají klasifikační regresivní stromy) čímž lze dosáhnout přesnějších výsledků. CART využívá stejně jako ANN prezenčně – absenční data a je založen na identifikaci specifického prahu každé ekologické proměnné. Data opakovaně rozdělují do homogenních podmnožin, nejlépe vystihující výskyt nebo absenci druhu. Kontrola výběru proměnných a předběžný odhad koeficientu vedou při používání neparametrických metod k automatické detekci a přizpůsobení prediktorů. Díky tomu se tyto metody používají ve stále větší míře (Elith et al. 2006).

Miller (2010) uvádí ještě jako samostatnou skupinu modely pracující pouze s prezenčními daty: Gowerova metrika, Mahalanobisova vzdálenost, nebo ENFA (ecological niche factor analysis). ENFA umožňuje sumarizaci variability v datech do několika faktorů, který vysvětlují: (i) marginalitu druhu (odchylka ekologického optima od nejvíce frekventovaných podmínek v území), (ii) toleranci druhu k suboptimálním podmínkám (Lobo et al. 2010). Velmi efektivní při práci s malým objemem prezenčních dat je MaxEnt (Maximum entropy model). MaxEnt je model vycházející ze strojového učení a využívá prezenčních dat. Pomocí všech dostupných znalostí o zkoumaném druhu vytváří co nejpřirozenější rozložení výskytu. Jak ale upozorňuje Phillips et al. (2009) jeho přesnost významným způsobem klesá, pokud se používá při extrapolaci.

I přes mnohé studie řešící používání algoritmů při tvorbě SDMs neexistuje stále ještě žádný jednotný a konzistentní algoritmus, který by se mohl používat pro všechny typy modelů bez ohledu na zkoumaný druh, geografickou distribuci, specifčnost habitatu a jiných ekologických proměnných. Volba vhodného modelu a algoritmu tedy stále

spočívá na výzkumnících a jejich empirických zkušenostech (příklady nejčastěji používaných softwarů a algoritmů viz Tab. 4).

Tab. 4 Nejčastěji používané softwary a algoritmy

Software	Algoritmy	URL	Zdroj
BIOMAPPER	ENFA	http://www.unil.ch/biomapper	Hirzel et al. (2001)
BIOMOD	GLM, GAM, CART, ANN	http://r-forge.r-project.org/projects/biomod/	Thuiller et al. (2009)
GARP	GA, GLM	http://lifemapper.org/desktopgarp	Stockwell and Peters (1999)
GRASP	GAM	http://www.unige.ch/ia/climate/grasp	Lehmann et al. (2002)
MaxEnt	Maximum Entropy	http://www.cs.princeton.edu/~schapire/maxent/	Phillips et al. (2009)

(Zdroj: Miller 2010; autor)

2.5 Validace SDMs

Pokročilé statistické metody (Meynard & Kaplan 2012) používané k validaci modelů založených na prezenčně – absenčních datech (lze použít i na prezenčních datech pokud zahrneme uměle vytvořená data pseudoabsenční) jsou odvozeny od chybové matice (confusion matrix). Ta k evaluaci používá nezávisle pořízená data z jiné lokality nebo pořízené odlišnou vzorkovací strategií. Protože ve většině případů nejsou taková data dostupná, lze rozdělit podle Pearsona et al. (2007) původní datovou sadu na dvě části: (i) data určená ke kalibraci (75 % z původního množství dat) a (ii) data použitá k validaci modelu (25 % původních dat). Z dat se následně vytvoří chybová matice (znázornění viz Tab. 5) skládající se z počtu záznamů rozdělených do 4 kategorií: (a) skutečně pozitivní, (b) falešně negativní, (c) falešně pozitivní a (d) skutečně negativní (Allouche et al. 2006).

Tab. 5 Obecná chybová matice tvořená pozorovanou prezencí (P) a absencí (A), která se využívá k hodnocení přesnosti u binárních kategorických dat. Šedé buňky indikují nesprávné výsledky (data falešné prezence nebo absence druhu), bílé buňky indikují výsledky správné (data skutečné prezence nebo absence druhu)

		Data získaná pozorováním	
		Prezence	Absence
Simulovaná data	P	a skutečně pozitivní	b falešně pozitivní
	A	c falešně negativní	d skutečně negativní

(Zdroj: Kienast et al. 2012)

Na základě uvedené matice lze formulovat další testy zaměřené na posouzení přesnosti modelu (nejčastější metriky viz Tab. 7). CCR (correct classification rate) a Kappa jsou metody vycházející z chybové matice ve formátu 2 x 2, ale obě lze v zásadě aplikovat na libovolnou n x n matici (Kienast et al. 2012). Kappa nabývá hodnot od 1 (perfektní model) do 0 (náhodná shoda modelu a reality). Pokud kappa (K) dosahuje negativních hodnot, předpovědní schopnosti modelu jsou nulové (Allouche et al. 2006).

Specificita (Specificity) a sensitivita (Sensitivity) jsou další často používané validační charakteristiky odvozené od chybové matice. Specificita vyjadřuje poměr mezi předpokládanou pozorovanou absencí a kvantifikací opomenutých chyb. Sensitivita vyjadřuje poměr předpokládaného pozorování prezence druhu a kvantifikací opomenutých chyb (Allouche et al. 2006). Oba postupy jsou navzájem nezávislé a neodvíjejí se ani od prevalence (poměr lokalit, ve kterých se druh vyskytuje).

Shora uvedené validační metody jsou závislé na stanovení prahové metody. Existují nicméně i metody, které více ovlivňuje volba prevalence. Miller (2010) jako příklad uvádí: NPP (negative predictive power) a PPP (positive predictive power).

Jednou z dalších metod využívanou k hodnocení modelů je ROC (receiver operating curve) vyjadřující vztah mezi specificitou a senzitivitou při různých diskriminačních hladinách. K interpretaci ROC se využívá AUC (area under cover). Od dříve uvedených metod se odlišuje nezávislostí na prahové hodnotě (Manel et al. 2001). Díky tomu není nutné sestavovat chybovou matici a realizovat s tím spojené počítání

opomenutých chyb. Jiménez-Valverde (2012) upozorňuje, že lze AUC využívat také u prezenčních dat, bez nutnosti aplikovat data absenční nebo pseudoabsenční. Je, ale nutné, modifikovat jeho znázornění jako poměr predikované oblasti výskytu ke všem možným prahovým hodnotám. AUC nabývá hodnot od 0 do 1, přičemž 1 udává parametry excelentního modelu a nižší hodnoty než 0,5 ukazují na model neschopný predikce (Kienast et al. 2012). Srovnání hodnot AUC a Kappa viz Tab. 6.

Tab. 6 Indexy sloužící ke klasifikaci přesnosti prediktivní schopnosti modelu

Přesnost	AUC	Kappa
výborná	0,9 - 1	0,8 - 1
dobrá	0,8 - 0,9	0,6 - 0,8
průměrná	0,7 - 0,8	0,4 - 0,6
podprůměrná	0,6 - 0,7	0,2 - 0,4
nedostačující	0,5 - 0,6	0 - 0,2

(Zdroj: Kienast et al. 2012)

Nevýhodou aplikace AUC je fakt, že znevýhodňuje modely potencionální distribuce (lokality, kde by druh mohl žít a reprodukovat se v závislosti na environmentálních podmínkách) a favorizuje modely realizované distribuce (lokality kde druh ve skutečnosti žije). Proto považuje Jiménez-Valverde (2012) AUC za nevhodnou evaluační metodu hodnotící SDMs.

Tab. 7 Běžně používané diskriminační metriky k určení přesnosti modelů druhové distribuce založených na binárních datech

Typ metriky	Popis	Výpočet
Sensitivita	poměr správně předvídané pozorované prezenze	$a/(a+c)$
Specificita	poměr správně předvídané pozorované absence	$d/(b+d)$
PPP	procento skutečně predikované prezenze	$a/(a+b)$
NPP	procento skutečně predikované absence	$d/(c+d)$
Kappa	procento shody	$\frac{[(a+d) \pm (((a+c)(a+b)+(b+d)(c+d))/n)]}{[n \pm (((a+c)(a+b)+(b+d)(c+d))/n)]}$

(Zdroj: Miller 2010; Kienast et al. 2012)

2.6 Virtuální druhy

2.6.1 Využití virtuálních druhů při modelování druhové distribuce

V současné době výzkumníci v oblasti ekologie, biogeografie a ochrany přírody vzrůstajícím tempem spoléhají na SDMs při výzkumech zaměřených na budoucí rozšíření druhů. Důsledkem toho je modelování druhové distribuce konstantně předmětem metodologických výzkumů týkajících se nových technik či evaluačních postupů (Barbet-Massin et al. 2012). Metodologické zdokonalení modelů je nutné podrobit validaci a testování, jež bylo donedávna založeno na využití reálných nebo empirických dat. Využití takových dat je však při validaci modelů problematické, protože data obsahují mnoho faktorů, které mohou model ovlivnit, a tím značně omezují možnost generalizace modelu (Leroy et al. 2015). Řešením je využití uměle vytvořených druhů (Meynard et al. 2013) s uměle stanovenou odpovědí druhů na gradienty prostředí. Výhodou je také možnost korekce druhové prevalence, ovlivnění reakce druhu na environmentální gradienty nebo úprava četnosti druhu (Hirzel et al. 2001). Další způsoby využití virtuálního druhu jsou uvedeny v Tab. 8.

Tab. 8 Příklady využití uměle vytvořených druhů

Využití virtuálních druhů	Zdroj
identifikace vlivu použití vstupních dat	Ward et al. 2009
modelování pouze s prezenčními daty nebo s využitím "pseudo-absenčních" dat	Barbet-Massin et al. 2012
vliv použití vzorkování s odlišnou prevalencí než je prevalence zkoumaného druhu	Real et al. 2006
použití různých korekcí u různě zkreslených dat	Albert & Thuiller 2008
efekt použitého měřítka na distribuční mapy	Bombi & D'Amen 2012

(Zdroj: Valera et al. 2014, autor)

Meynard et al. (2013) dělí tvorbu virtuálního druhu do tří kroků: (i) modelace pravděpodobnosti výskytu nebo indexu vyjadřujícího vhodnost habitatu založené na jedné nebo více environmentálních proměnných, (ii) transformace pravděpodobnosti výskytu na mapu prezenze/absence, (iii) testování získaného modelu druhové distribuce.

V současnosti je možné využívat několik softwarů, které lze využít k vytvoření virtuálního druhu. Příkladem může být RangeShifter nebo HexSim. Leroy et al. (2015)

nicméně upozorňuje na fakt, že tyto softwary nelze použít k vyjádření vztahu druh – prostředí což je jeden z hlavních důvodů tvorby SDMs. Takto vytvořené virtuální druhy nelze použít při simulaci těchto vztahů na výsledný model. Například Moudrý (2015) uvádí dva odlišné, statistické softwary, které eliminují výše uvedená negativa. Jedná se o SDMspecies a virtualspecies.

SDMspecies je software pracující s prahovou hodnotou (Xiaoquan et al. 2015). Nastavením prahové hodnoty dosáhne uživatel výskytu druhu vždy nad touto hranicí, nikdy však pod ní (Bombi & D'Amen 2012). Tento postup, ale například kritizuje Meynard & Kaplan (2012) či Moudrý (2015), neboť vede k nadhodnocení modelu a jeho výsledek je tím pádem zavádějící. Leroy et al. (2015) softwaru vytýká omezenou možnost nastavení vztahu druh – prostředí a poukazuje na chybějící eventualitu nastavení pravděpodobnosti výskytu či absenci možnosti vzorkování druhového výskytu.

Software virtualspecies výše uvedené nedostatky odstraňuje. Navíc, možností uživatele definovat jakýkoliv vztah druh – prostředí, vzrůstá implikace komplexního prostředí při modelování. Výhodou je i možnost nastavení prevalence a její aplikace při převodu rastru vyjadřujícího vhodnost prostředí na prezenci – absenci druhu (Meynard et al. 2013). Vysoká flexibilita, integrace výše uvedených eventualit a možnost plného přizpůsobení je důvodem proč používat tento software (Leroy et al. 2015; Moudrý 2015). Důležité je, že oproti prvnímu uvedenému softwaru nepracuje s prahovou hodnotou, ale využívá pravděpodobnostního přístupu. Rozdílům mezi těmito dvěma přístupy je popsán v další kapitole.

2.6.2 Komparace prahové metody a pravděpodobnostního přístupu

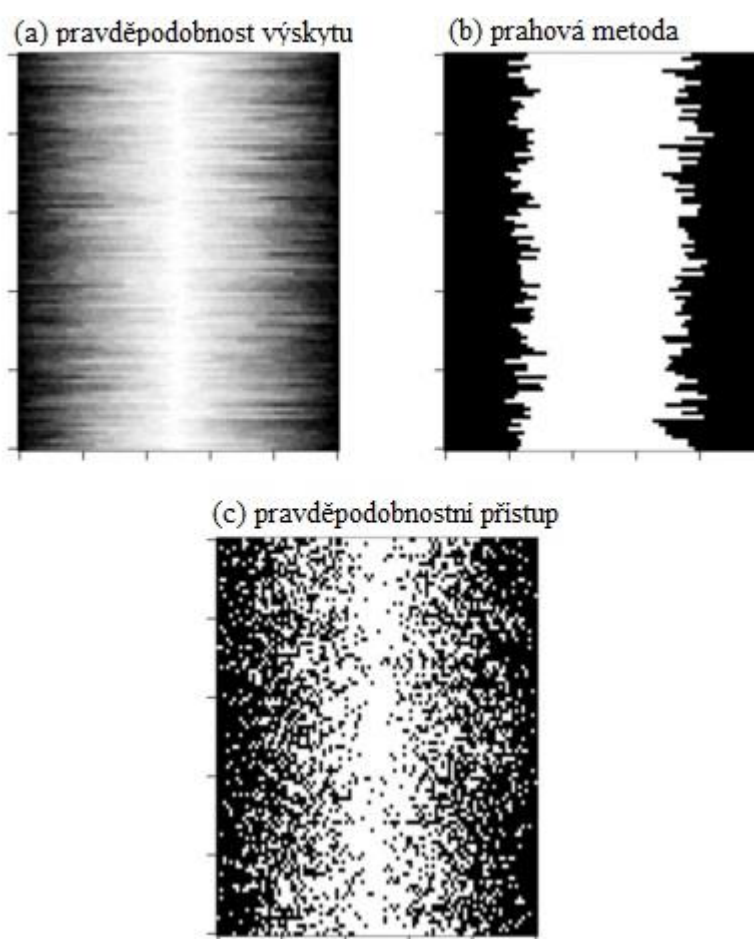
V mnoha studiích se používá fixní prahová metoda k transformaci simulovaného pravděpodobného výskytu virtuálního druhu na prezenčně – absenční mapu (Bombi & D'Amen, 2012; Valera et al. 2014). Princip prahové metody spočívá ve stanovení specifické pravděpodobnosti výskytu, kdy se druhy vyskytují vždy tam, kde se hodnoty pohybují nad stanovenou hranicí nikdy však pod ní. To umožňuje efektivní translaci jakékoliv křivky simulující reakci druhu na změnu prostředí do jedné hodnoty a tím je možno eliminovat náhodné chyby vzniklé vzorcem druhového rozšíření (Meynard & Kaplan 2012). I přes tuto výhodu, existuje několik důvodů proč nepoužívat tuto metodu při modelování druhové distribuce.

Za prvé, ekologické teorie jsou založeny na dynamickém vzoru výskytu druhů v závislosti na prostoru a čase (Hanski 1998). Navzdory tomu prahová metoda používá staticky stanovenou odpověď druhu na environmentální gradienty. Výjimku v tomto ohledu tvoří globální prezenčně – absenční data, která jsou shromažďována v delším časovém horizontu. Taková data v kombinaci s celosvětovým měříkem popsaný problém eliminují. Aplikace prahové metody však není vhodné obecně přenášet na všechny datové sady. Konkrétně u dat sbíraných na malém území v krátkém časovém horizontu reprezentujících dynamické rozšíření druhu jsou vzorce výskytu důležité a není žádoucí eliminovat je použitím prahové metody (Meynard & Kaplan 2012). Prahová metoda se může objevit, i pokud aplikujeme pravděpodobnostní přístup v případě, že velmi omezíme rozsah environmentálních proměnných, čímž dojde k rapidní změně ve výskytu druhu a vznikne tím neúmyslně stanovená prahová hodnota.

Druhým důvodem je fakt, že mnoho statistických modelů fungujících na principu logistické křivky (například GLM nebo GAM) je nevhodných pro modelování s pomocí prahové metody. Sklon logistické křivky v bodě inflexe (bod, ve kterém křivka dosahuje svého maxima) je v případě prahové hodnoty nekonečný, čímž ovlivňuje správnou konvergenci modelu. Řešením problému jsou alternativní algoritmy, které jsou ale v případě některých modelů chybové (Meynard & Kaplan 2012).

Prahová metoda také v některých případech zkresluje konečná data. Výsledkem jsou nesprávné nebo nekompletní odpovědi na výzkumnou hypotézu. Ilustrovat to můžeme na výzkumu Hirzela et al. (2001), který zkoumal vliv zvolené funkce znázorňující odezvu mezi druhem a environmentálním prostředím na výsledný SDM. Při použití prahové metody došel k závěru, že volba odezvové funkce nemá žádný signifikantní vliv na výsledek. Oproti tomu Elith & Graham (2009) aplikací pravděpodobnostní metody prokázali důležité rozdíly související se zvolenou odezvoovou funkcí a druhovou odezvou.

Dalším důvodem je nemožnost opakování simulace. Jak bylo řečeno dříve, prahová metoda eliminuje variabilitu reakce druhu na změnu podmínek prostředí. Výsledky opakované simulace prezenze – absence budou z tohoto důvodu mít vždy stejný výsledek a iterace nemá žádný smysl. V případě použití pravděpodobnostního přístupu však každé nové opakování přinese nový výsledek s odlišným vzorem druhového výskytu. Eventualita neustále opakovat simulaci a proces tvorby predikčních modelů ve stejných podmínkách otvírá prostor simultánně vyhodnocovat důsledky různých chyb plynoucích z prezenčně – absenčních dat. Odstranění takových chyb povede v budoucnu k vylepšení predikčních možností druhových modelů (Meynard et al. 2013).



Obr. 5 Srovnání druhové distribuce aplikací prahové metody a pravděpodobnostního přístupu. Každý panel ilustruje využití virtuální krajiny s osami x a y reprezentujícími geografické souřadnice. Graduující stínování znázorňuje (a) pravděpodobnost výskytu vytvořenou z prezenčně – absenčních dat, ve zbývajících panelech pak černá znázorňuje vysokou pravděpodobnost výskytu, oproti tomu bílá znázorňuje nízkou pravděpodobnost výskytu: (b) modelace pravděpodobnosti výskytu pomocí prahové metody, (c) modelace pravděpodobnosti výskytu aplikací pravděpodobnostní metody.

Zdroj: (Meynard & Kaplan 2012)

Vlivem výše uvedených negativ spojených s prahovou metodou, poukazuje stále více studií (Elith & Graham 2009; Santika 2011; Meynard & Kaplan 2012; Moudrý 2015) na fakt, že by se měl místo prahové metody preferovat pravděpodobnostní přístup (rozdíl mezi prahovou metodou a pravděpodobnostním přístupem viz Obr. 5). Jeho aplikací je prezenze – absence nahodilým procesem převedena na pravděpodobnost výskytu, jež graduálně odpovídá na environmentální proměnné. Vzniklá mapa výskytu druhů nabývá hodnot od 0 (vždy nepřítomný) do 1 (vždy přítomný). Například, pokud bude pravděpodobnost výskytu 0,1 na straně jedné a 0,6 na straně druhé, budou taková místa se stejnými environmentálními podmínkami obsazena v jednom a šesti případech z 10 (Moudrý 2015).

2.6.3 Vliv graduální reakce druh – prostředí na progres modelu

Meynard & Kaplan (2012) uvádějí, že ke kontrole graduální odpovědi druhu na změnu prostředí se nejčastěji používá logistická křivka, umožňující propojení environmentálních gradientů s pravděpodobností výskytu:

$$P_i = \frac{1}{1 + e^{-Y_i}}$$

kde p_i vyjadřuje pravděpodobnost výskytu druhu v lokalitě i a Y_i funkci environmentálního gradientu případně jejich kombinaci na lokalitě i .

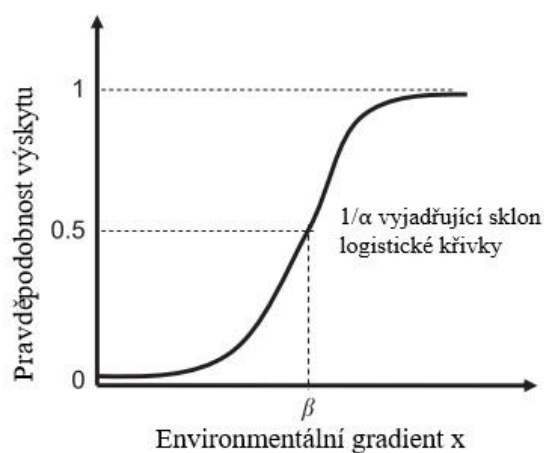
Pro lepší vyjádření vztahu sklonu a inflexního bodu logistické křivky s rozsahem environmentálních gradientů stejně jako biologických proměnných druhů lze Y_i vyjádřit:

$$Y_i = \frac{x_i - \beta}{\alpha}$$

kde x_i vyjadřuje environmentální gradient nebo jejich kombinaci v lokalitě i , β je bod inflexe logistické křivky (například hodnota environmentálního gradientu ve kterém je pravděpodobnost výskytu 50%) a $1/\alpha$ vyjadřuje sklon logistické křivky v bodě inflexe (umožňuje kontrolovat graduální odpověď druhu na environmentální gradient (viz Obr. č. 6). Nízké hodnoty α simulují reakci druhu v případě použití prahové hodnoty. Vysoké hodnoty α simulují lineární reakci druhu. Výsledkem aplikace středních hodnot je klasická logistická křivka ve tvaru S. V případě kontroly druhové prevalence je třeba nejdříve prozkoumat vliv parametrů na progres modelu. Obecně

lze tvrdit, že při zafixované hodnotě α a nárůstu hodnoty β , se bude snižovat druhová prevalence (Meynard & Kaplan 2012).

Logistická křivka (příklad viz Obr. 6) je užitečná při modelování za pomoci virtuálního druhu, protože umožňuje kontrolovat, jak rychle bude druh reagovat na změnu environmentálního gradientu pomocí výše definované funkce Y_i . Zároveň definuje kontinuum mezi prahovým a lineárním tvarem křivky. Díky tomu lze zpracovat uvedené extrémy jako dva konkrétní příklady v rámci jedné více obecné logistické formulace.



Obr. 6 Schéma logistické křivky s výše uvedenými parametry. $1/\alpha$ vyjadřuje sklon křivky a β (nízká hodnota β zvyšuje šanci na výskyt druhu, vysoká hodnota β jí naopak snižuje) je inflexním bodem environmentálního gradientu s pravděpodobností výskytu 0,5. Sklon logistické křivky reprezentuje, jak rychle se bude měnit výskyt druhu na určitém území v závislosti na změnách environmentálního gradientu v blízkost inflexního bodu.

(Zdroj: Meynard & Kaplan 2012)

3 Metodika

3.1 Softwarové požadavky a vstupní data

3.1.1 Použitý software

K vytvoření modelu druhové predikce pomocí virtuálního druhu je potřeba standardní verze matematického softwaru R (použitá verze softwaru: R 3.2.2), volně dostupného na www.r-project.org a několika dalších extenzí, které lze implementovat v programu R (dismo, maps, raster, rgal, rJava, roxygen2, sp, sqldf, testthat, virtualspecies). Možností je také použít verzi Maxentu v rozhraní java, jež umožňuje jednodušší modelování druhové predikce. Pokud nemá uživatel předem připravená data území, bude potřeba instalace extenze rworldmap.

Pro účely diplomové práce nicméně posledně uvedené rozšíření nebude potřeba. Vstupní klimatické vrstvy byly předem upraveny v programu ArcMap ve verzi 10.3, kde pomocí funkce *Clip* došlo k jejich oříznutí tak, aby pokrývaly pouze území Iberského poloostrova (hranice území dostupné na <http://www.eurogeographics.org/>).

3.1.2 WorldClim

WorldClim představuje soubor globálních klimatických vrstev vytvořených interpolací průměrných měsíčních dat získaných z meteorologických stanic v letech 1960 – 1990 v rozlišení 1 x 1 km² (Hijmans et al. 2005). Vrstvy obsahují informace o průměrných měsíčních srážkách, celkovém měsíčním srážkovém úhrnu, minimální a maximální teplotě a dalších 19 bioklimatických proměnných (například srážky v nejsušším měsíci, srážky v nejvlhčím měsíci nebo nejvyšší teplota nejteplejšího měsíce). Data jsou volně dostupná na <http://www.worldclim.org/> pro akademické či nekomerční účely a lze je aplikovat při tvorbě mapových výstupů nebo při prostorovém modelování v GIS.

3.2 Definování virtuálního druhu

Prvním krokem při tvorbě virtuálního druhu je definování odezvové funkce (Gaussova křivka, lineární, kvadratická funkce etc.) pro každou environmentální proměnnou vstupující do modelu druhové distribuce včetně jejich následné kombinace (násobení, sčítání etc.) pomocí funkce *generateSpFromFun*. Tím vznikne mapa ilustrující vhodnost environmentálního prostředí pro výskyt druhu.

V rámci diplomové práce bylo při generování umělého druhu použito tři vrstev dostupných z výše popsaného zdroje obsahující informace o maximální teplotě v nejteplejším měsíci (Bio5), minimální teplotě v nejchladnějším měsíci (Bio6) a srážkovém úhrnu v nejsušším měsíci (Bio14).

Na základě těchto bioklimatických proměnných v rozlišení 1 x 1 km² lze vygenerovat rastr znázorňující environmentální vhodnost prostředí pomocí již zmíněné funkce *generateSpFromFun*. Jako odezvoivá funkce ilustrující vztah druh – environmentální prostředí je použita Gaussova křivka, manuálně specifikovaná u každé bioklimatické vrstvy následovně:

- (i) obecný vzorec Gaussovy křivky

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

kdy μ vyjadřuje průměrnou hodnotu (mean), σ vyjadřuje standartní odchylku (standart deviation) a $f(x)$ vyjadřuje hodnotu výsledné vrstvy

- (ii) maximální teplota v nejteplejším měsíci (Bio 5)

$$f_{(Bio5)} = \frac{1}{100\sqrt{2\pi}} e^{-\frac{(x-200)^2}{2 \times 100^2}}$$

průměrná hodnota (μ) 200, standartní odchylka (σ) 100

- (iii) minimální teplota v nejchladnějším měsíci (Bio 6)

$$f_{(Bio6)} = \frac{1}{100\sqrt{2\pi}} e^{-\frac{(Bio5-100)^2}{2 \times 100^2}}$$

průměrná hodnota (μ) 100, standartní odchylka (σ) 100

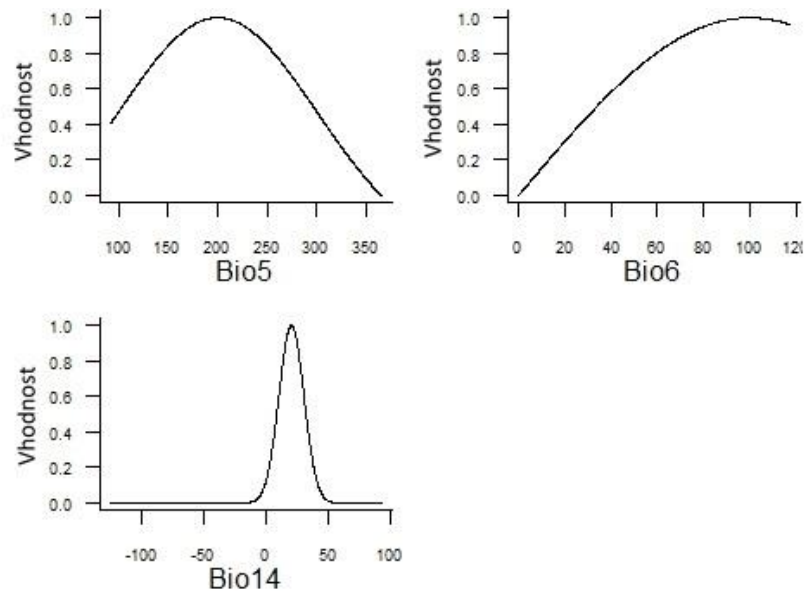
- (iv) srážkový úhrn v nejsušším měsíci (Bio 14)

$$f_{(Bio14)} = \frac{1}{10\sqrt{2\pi}} e^{-\frac{(Bio5-20)^2}{2 \times 10^2}}$$

průměrná hodnota (μ) 20, standartní odchylka (σ) 10

Přičemž průměrná hodnota vydělená 10 ilustruje nejvhodnější podmínky pro výskyt druhu a směrodatná odchylka vydělená 10 ohraničuje podmínky environmentálního prostředí, které je druh schopný tolerovat.

Výsledný vztah druh – environmentální prostředí je možno zobrazit pomocí funkce *plotResponse*, která je obsažena v extenzi *virtualspecies* (Obr. 7).

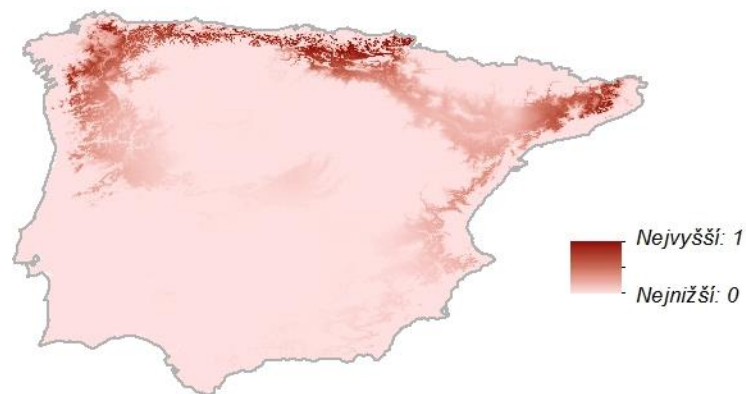


Obr. 7 Ukázka definované Gaussovy křivky pro jednotlivé bioklimatické vrstvy použité v této práci. Osa X znázorňuje environmentální proměnnou. Osa Y znázorňuje pravděpodobnost výskytu.

(Zdroj: autor)

Následným vynásobením bioklimatických vrstev dojde k vytvoření rastru vyjadřujícího vhodnost environmentálního prostředí (Obr. 8).

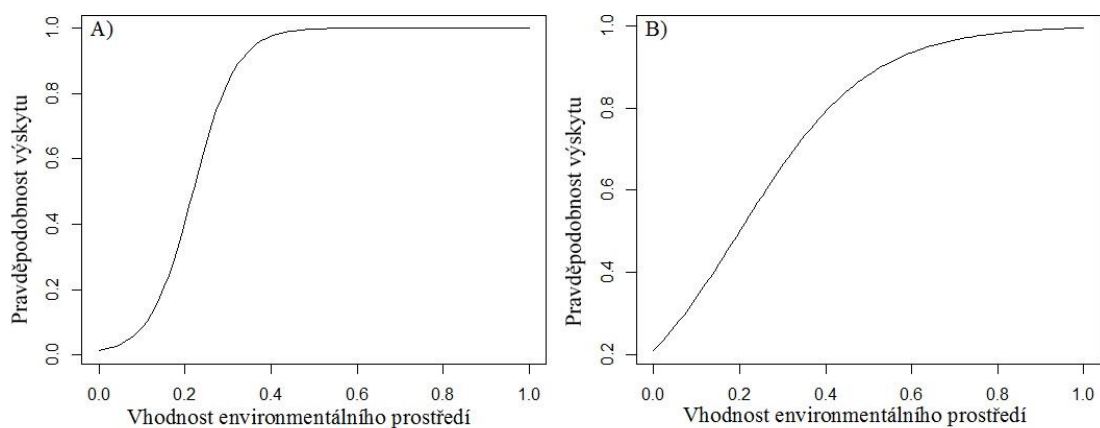
$$\text{Vhodnost environmentálního prostředí} = f_{(\text{Bio5})} \times f_{(\text{Bio6})} \times f_{(\text{Bio14})}$$



Obr. 8 Ilustrace vhodného environmentálního prostředí pro výskyt virtuálního druhu na Iberském poloostrově na základě zadaných specifikací. Hodnoty nabývají od 0 (nevhodné prostředí pro výskyt druh) až k 1 (nejvhodnější prostředí druhu).

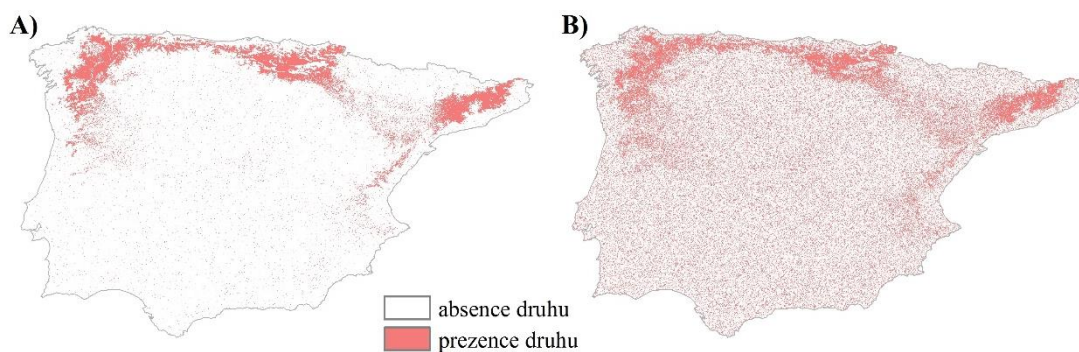
(Zdroj: autor)

Druhým krokem je konverze environmentální vhodnosti prostředí na prezenci – absenci druhu. Zde je vhodné aplikovat metodu pravděpodobnostního přístupu. Hodnotu pravděpodobnosti výskytu (druhovou prevalenci) jsem stanovil pro účely mé práce na 0,1 pro variantu A a 0,3 pro variantu B (viz Obr. 9 a 10). Během binomického výpočtu, realizovaného pro každou jednotlivou buňku, bude tedy prezence přidělena každé 1 a 3 buňce z 10 v rámci stanovené hodnoty.



Obr. 9 Logistická funkce použitá k transformaci vztahu druh – prostředí na data simulující prezenci – absenci druhu. Varianta A s prevalencí 0,1 má hodnoty α (sklon křivky) $-0,05$ a β (inflexní bod) má 0,2. Varianta B s prevalencí 0,3 má α (sklon křivky) hodnotu $-0,15$ a β (inflexní bod) má hodnotu 0,2.

(Zdroj: autor)



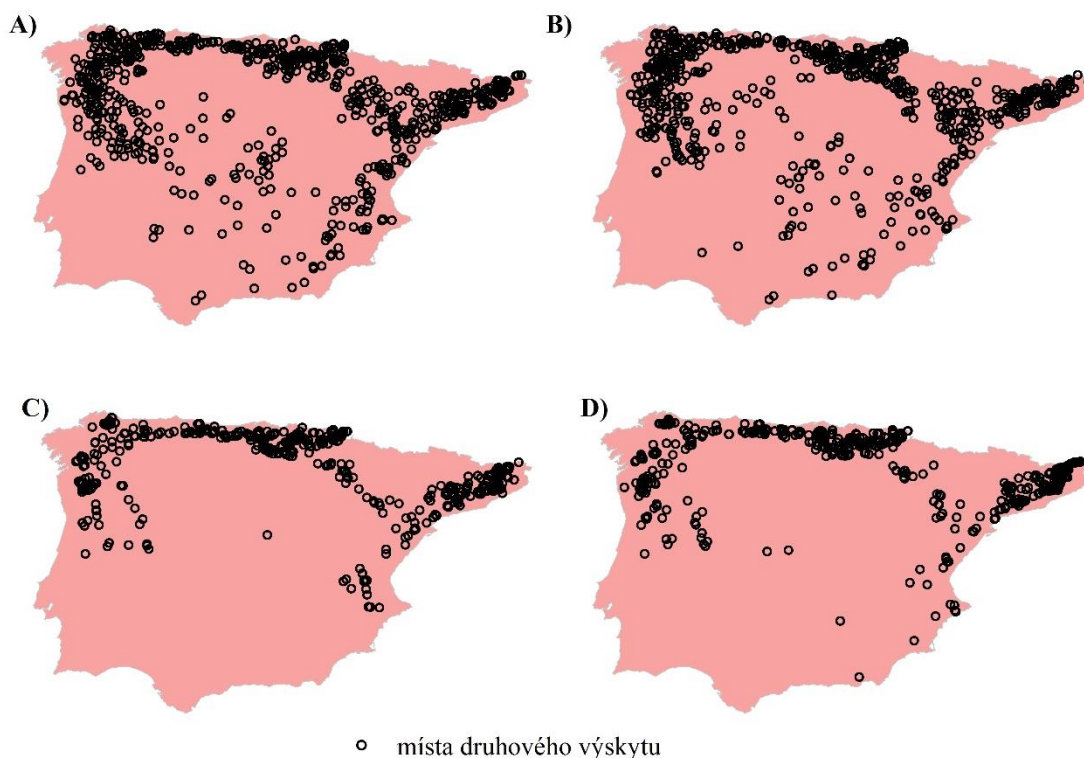
Obr. 10 Kontrastní příklad vlivu konverzní křivky na distribuci stejného virtuálního druhu. Aplikace logistické funkce umožňuje simulovat reálnější distribuci druhu v porovnání s lineární nebo prahovou metodou. Ilustrace A) ukazuje prezenci – absenci druhu v případě použití výše popsané logistické křivky (Obr. 9 ilustrace A). Ilustrace B) oproti tomu ukazuje konverzi druhu na prezenci – absenci za použití odlišné logistické křivky (Obr. 9 ilustrace B), která je na hranici mezi logistickou a lineární konverzí.

(Zdroj: autor)

3.3 Data výskytu

K testování použité techniky tvorby modelu či jeho validaci je potřeba v posledním kroku vytvořit vrstvu zachycující data o výskytu druhu. Toho lze docílit funkcí *sampleOccurrences*, která umí vzorkovat jak data prezenčně – absenční tak pouze

prezenční. Jak bylo již uvedeno v kapitole pojednávající o biologických datech, většina dnes dostupných biologických datových sad obsahuje pouze informace o výskytu druhu (prezenční data). Právě proto jsou i pro účely této práce použita pouze prezenční data, generovaná na základě virtuálního druhu definovaného v předchozí kapitole. Každý výskyt je vytvořen vždy dvakrát, a to na základě druhové prevalence založené na hodnotách 0,1 a 0,3. Jako první byla vytvořena náhodně generovaná prezenční data obsahující 700 míst s výskytem druhu (viz Obr. 11A, 11B).



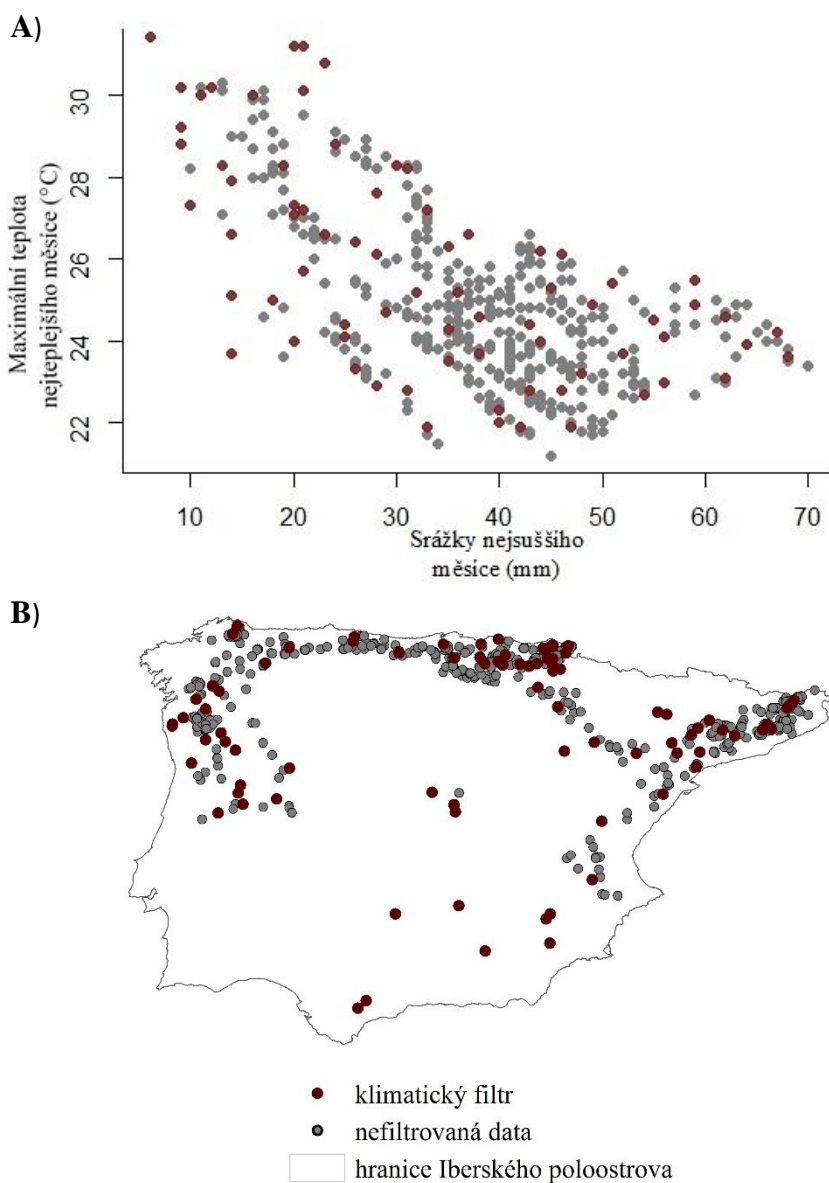
Obr. 11 Geografická distribuce: (i) náhodně generovaného výskytu druhu s prevalencí 0,1 (A) a prevalencí 0,3 (B), (ii) generovaného výskytu druhu se 40krát větší intenzitou pro chráněné oblasti (C prevalence 0,1 a D prevalence 0,3) na Iberském poloostrově.

(Zdroj: autor)

Reálné vzorkování dat je nicméně spjato s jistou mírou chybovosti a to i v případě terénních průzkumů založených na systematickém sběru dat. V případě často využívaných globálních databází nejčastěji dochází ke vzniku hojně vzorkovaných míst (národní parky, okolí osídlených lokalit etc.) v kontrastu s málo vzorkovanými lokalitami (především místa vzdálená od jakýchkoliv běžných antropogenních aktivit). K simulaci reálnějších dat bylo proto vytvořeno 470 míst s druhovým výskytem, jenž byl modifikován tak, aby se generoval se 40krát větší intenzitou v chráněných oblastech (národní parky, ptačí oblasti etc.) ležících na Iberském poloostrově (<<http://www.protectedplanet.net/>>). Tedy v místech s častým pohybem především odborné veřejnosti (viz obr 11C, 11D).

3.4 Klimatické filtrování dat

Pomocí funkce *gridSample* byl na datové sady aplikován klimatický filtr. Tato funkce vybírá body z rastrové vrstvy X-Y pomocí předem definovaného gridu. K napodobení studií reálných druhů, u nichž nejsou k dispozici všechny potřebné klimatické vrstvy nebo nejsou v případě některých druhů známy, byly použity pouze dvě environmentální proměnné namísto tří, použitých při tvorbě virtuálního druhu. Osa Y znázorňuje maximální teplotu nejteplejšího měsíce, osa X pak srážky v nejsušším měsíci. Opakovaný výskyt v podobných klimatických podmínkách byl vyřazen. Takto definovaný klimatický filtr (Obr. 12) byl aplikován na všechny použité datové sady.



Obr. 12 Klimatický filtr aplikovaný na datovou sadu zobrazený v environmentálním prostoru (A) a geografickém prostoru (B). Zde je ilustrován pouze jeden příklad (470 vygenerovaných míst výskytu druhu na základě prevalence 0,1). Ostatní datové sady viz příloha C.

(Zdroj: autor)

3.5 Druhové distribuční modely

K vytvoření modelu druhové distribuce pro všechny výše vytvořené datové sady byl aplikován MaxEnt ve verzi 3.3.3k (<www.cs.princeton.edu/~schapire/Maxent>), který je v současné době často využíván v odborných studiích, pracujících s prezenčními daty (Kramer-Schadt et al. 2013; Syfert et al. 2013; Fourcade et al. 2014; Valera et al. 2014). Volitelné funkce MaxEntu byly ponechány na defaultních hodnotách. Výjimku tvoří testovací data, nastavená na 25 % z celkového množství prezenčních dat. Všechny tři klimatické vrstvy použité při tvorbě virtuálního druhu byly použité jako prediktory (maximální teplota v nejteplejším měsíci, minimální teplota v nejchladnějším měsíci, srážkový úhrn v nejsušším měsíci).

Na základě postupu, jež stanovil DeLong et al. (1988) a vytvořené binární mapy vypočítal MaxEnt automaticky AUC (diskriminační ukazatel přesnosti distribučních modelů) pro tréninková a testovací data.

Díky virtuálnímu druhu můžeme výsledný model porovnat s virtuální realitou reprezentovanou vhodností environmentálního prostředí. Čímž srovnáme kalibrační schopnost modelů. Abychom mohli komparovat modely pomocí párového *t*-testu, byly výsledné filtrované i nefiltrované rastry relativní druhové distribuce odečteny od již uvedeného rastru ilustrujícího vhodnost environmentálního prostředí virtuálního druhu. Pro každou dvojici datových sad (filtrovaná versus nefiltrovaná data) byla individuálně spočítána *p*-hodnota se stanovenou hladinou významnosti $\alpha = 0,05$. Tyto testy byly implementovány s cílem testování nulové hypotézy (H_0) založené na předpokladu, že rozdíl mezi realitou a filtrovaným a realitou a nefiltrovaným modelem (respektive jejich střední hodnoty) se neliší.

K určení síly statistické závislosti mezi virtuální realitou a výslednými modely byl použit Pearsonův korelační koeficient, jenž lze zjednodušeně vyjádřit:

$$r = \frac{S_{xy}}{S_x S_y}$$

kde je S_x směrodatná odchylka proměnné X , S_y směrodatná odchylka proměnné Y a S_{xy} takzvaná kovariance proměnných X a Y (Sheskin 2004).

Na závěr byly srovnány reakce druhu na změnu environmentálních proměnných mezi virtuální realitou a výslednými modely.

4 Výsledky práce

4.1 AUC

Na základě srovnání AUC (viz Tab. 9) filtrovaných a nefiltrovaných modelů lze potvrdit, že klimatické filtrování zvyšuje diskriminační schopnost modelů, jak tvrdí Valera et al. (2014) ve své studii. Anomálií je model, vytvořený na základě virtuálního druhu s prevalencí 0,3 a 470 záznamy druhové prevalence. Komparace všech modelů ukazuje, že modely s nižším počtem záznamů druhového výskytu a modely založené na nižší hodnotě druhové prevalence dosahují vyššího AUC. Hodnota AUC < 0,8 u všech modelů indikuje dobrou až výbornou predikční schopnost modelů. AUC nicméně pouze ukazuje na vyšší schopnost diskriminovat filtrované body (rozlišení prevalence absence), což však neznamená, že modely také reflektují přesněji realitu.

Tab. 9 Výsledné hodnoty AUC a párového T testu

	Druhová prevalence 0,1				Druhová prevalence 0,3			
	470	470 KF	700	700 KF	470	470 KF	700	700 KF
AUC	0.95	0.95	0.89	0.9	0.95	0.94	0.86	0.88
Střední hodnota	-0.0039	-0.03	-0.0875	-0.0871	-0.0004	-0.044	-0.0872	-0.1154
sd	0.0357	0.0553	0.1064	0.1033	0.0333	0.0664	0.1066	0.1245
n	1 438 965	1 438 965	1 438 965	1 438 965	1 438 965	1 438 965	1 438 965	1 438 965
d	5.83E-05	9.04E-05	1.00E-04	0.00016	5.44E-05	0.0001	0.00018	0.0002
95 % int. Spodní	-0.0039	-0.03	-0.0876	-0.0873	-0.0005	0.0441	-0.0874	-0.1156
Horní	-0.0038	-0.0299	-0.0873	-0.0869	-0.0004	-0.0439	-0.087	-0.1153
P hodnota	< 0.001		0.005		< 0.001		< 0.001	

(Zdroj: autor)

d – vzdálenost hranic konfidenčního intervalu od průměru

470/700 – počet zaznamenaných výskytů v datové sadě

KF – datová sada s aplikací klimatického filtru

n – celkový počet buněk rastru

sd – směrodatná odchylka

4.2 Párový t -test

Dvouvýběrový t -test byl použit k testování nulové hypotézy (H_0) založené na předpokladu, že rozdíl mezi realitou a filtrovaným versus realitou a nefiltrovaným modelem (respektive rozdíl jejich středních hodnot) se u všech provedených testů neliší (je nulový). Hladina významnosti $\alpha = 0,05$.

Na základě výsledné p hodnoty se nulová hypotéza zamítá, což znamená, že jeden z dvojice porovnávaných modelů je vždy statisticky významně přesnější (viz Tab. 9). Komparací středních hodnot lze zjistit, že druhové distribuční modely bez aplikace klimatického filtrování dosahují vyšší kongruence s realitou.

Výjimkou je datová sada se 700 záznamy výskytu druhu generovaná na základě druhové prevalence 0,1, kde dosáhl model bez klimatického filtrování (*střední hodnota* = -0.0875) nižší shody s realitou oproti modelu s klimatickým filtrováním (*střední hodnota* = -0.0871).

Srovnání všech středních hodnot ukazuje, že modely u kterých byla použita datová sada s nižším počtem záznamů výskytu druhu, jsou shodnější s realitou.

4.3 Korelační koeficient

Výsledné hodnoty korelačních koeficientů zobrazené pomocí korelační matice ukazují na významnou statistickou korelaci mezi virtuální realitou a všemi výslednými modely (Tab. 10), což značí vysokou míru kongruence. I přes významnou korelaci lze najít mezi modely rozdíly vzniklé odlišným způsobem modelování a použitím různých datových sad.

Za prvé, výsledné hodnoty ukazují na vyšší validační schopnost nefiltrovaných modelů. Za druhé modely s nižší druhovou prevalencí dosahují vyšší hodnot korelačního koeficientu oproti modelům, pro jejichž tvorbu byly využity datové sady s vyšší druhovou prevalencí. Další důležitý rozdíl pak ukazuje srovnání datových sad podle počtu záznamů o výskytu druhu, kdy lepších výsledků, respektive vyšší kongruence s virtuální realitou dosahují modely s nižším počtem prezenčních záznamů.

Signifikantní rozdíl mezi jednotlivými korelačními koeficienty pak potvrzuje p hodnota, která je při komparaci mezi jednotlivými modely a virtuální realitou $< 0,001$.

Tab. 10 Tabulka Pearsonových korelací ve formě korelační matice

		0,1	0,3	0,1	0,3	0,1	0,3	0,1	0,3
		470	470	700	700	470KF	470KF	700KF	700KF
	virtuální realita								
0,1	virtuální realita	1	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
0,1	470	0.9745	1	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
0,3	470	0.972	0.9715	1	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
0,1	700	0.8878	0.8767	0.8601	1	p<0.001	p<0.001	p<0.001	p<0.001
0,3	700	0.8838	0.8695	0.8494	0.9929	1	p<0.001	p<0.001	p<0.001
0,1	470KF	0.9661	0.9553	0.9526	0.9488	0.9464	1	p<0.001	p<0.001
0,3	470KF	0.949	0.9332	0.9455	0.9551	0.9476	0.9859	1	p<0.001
0,1	700KF	0.8814	0.855	0.8416	0.9735	0.9723	0.9408	0.9498	1
0,3	700KF	0.8383	0.8168	0.8185	0.956	0.951	0.904	0.9465	0.953

(Zdroj: autor)

470/700 – počet zaznamenaných výskytů v datové sadě

KF – datová sada s aplikací klimatického filtru

0,1/0,3 – hodnota druhové prevalence

$p < 0.001$ – p hodnota

5 Diskuze

Nedávné studie (například Duputié et al. 2014; Guillera-Aroita et al. 2015) zkoumající vliv vstupních dat na validitu druhových distribučních modelů dokazují, že přesná vstupní data, jejich dostupnost a nedostatečná kvalita v korelaci s obtížným vývojem kontrolních metod určujících přesnost použitých modelů představují nejvíce limitující faktory při modelování druhové distribuce. Řešením, umožňující zkoumat vliv těchto faktorů na výsledné modely, je použití virtuálních druhů, u nichž lze kontrolovat odpověď druhu na gradienty prostředí (Meynard & Kaplan 2012), vzorec druhového výskytu (Hirzel et al. 2001; Zurell et al. 2010) nebo vliv filtrovacích metod redukujících počet míst s druhovým výskytem (Valera et al. 2014) tak, jako tomu bylo v této práci.

I přes rozšířenost prahové metody (Hirzel et al. 2001; Valera et al. 2014) byla při definování odpovědi druh – gradient prostředí v této studii použita při generování virtuálního druhu pravděpodobnostní metoda, která je odbornou veřejností (Elith & Graham 2009; Santika 2011; Meynard & Kaplan 2012; Moudrý 2015) stále více podporována a to z důvodů negativních vlivů prahové metody na celý proces modelování druhové distribuce (viz kapitola 2.6.2).

K odstranění nedostatků v prezenčních datech, která jsou ovlivněna nerovnoměrným sběrem dat – například s převahou záznamů na snadno dostupných lokalitách jsou při modelování druhové distribuce používány různé filtry. Nejčastěji se jedná o klimatické a geografické filtrování datových sad (například Tessarolo et al. 2014; Valera et al. 2014). Nicméně Moudrý (2015) tvrdí, že filtrování dat bez ohledu na jejich charakteristiku, tak jako to provedla Valera et al. (2014) ve své studii je vždy nahodilé. Podobné misinterpretace způsobené důvěrou autorů ve vysoké AUC hodnoty jsou však poměrně časté (viz Lahoz – Monfort et al. 2014; Moudrý 2015).

Vyhodnocení výsledných modelů pomocí AUC potvrdilo, že aplikace klimatického filtrování vede k vyšší diskriminační schopnosti modelů, což potvrzuje závěry studie Valery et al (2014) a to i v případě, kdy byl virtuální druh generován pomocí pravděpodobnostní metody. Vysoké hodnoty AUC však pouze indikují vyšší způsobilosti diskriminovat filtrované body, respektive odlišit prezenci absenci ve vybraném vzorku dat, ale neumožňují reflektovat realitu. A to z toho důvodu, že výsledné modely nejsou komparovány s charakteristikou vstupních dat.

Proto doporučuje Lawson et al. (2014) použít k validaci druhových modelů širokou škálu kontinuálních kalibračních metrik nebo kalibračních grafů, jež ilustrují výkon modelů ve srovnání s realitou (příklad kalibračních grafů lze nalézt ve studii Lahoz – Monfort et al. 2014). Rozmach těchto metod pak může v budoucnu znamenat odklon od používání AUC jako jednoho z hlavních ukazatelů výkonu druhových distribučních modelů.

V reakci na absenci takového srovnání výsledných modelů a vstupních dat u Valery et al. (2014) byla virtuální realita komparovaná s výslednými filtrovanými a nefiltrovanými modely pomocí párového t-testu a korelačního koeficientu. Výsledky obou prováděných testů indikují, že i přes vyšší dosažené hodnoty AUC, jsou modely založené na klimatickém filtrování oproti nefiltrovaným modelům více inkongruentní ve srovnání s virtuální realitou. To prokazuje předpoklad této práce, založený na hypotéze, že klimatické filtrování snižuje výkon modelů druhové distribuce.

Studie ukazuje překvapivé výsledky, v případě srovnání nefiltrovaných modelů v závislosti na velikosti použité datové sady. Jak v případě AUC, párového *t*-testu a korelačního koeficientu dosahují totiž vyšší validace modely, u kterých byly použity datové sady s menším počtem záznamů druhového výskytu. To může indikovat, že v případě druhů s jednoduchou ekologickou nikou, je lepší použít malé datové sady. Na druhou stranu větší datové sady dosahují více konzistentních výsledků s menšími rozdíly ve středních hodnotách. Velikost datové sady a její vliv na výsledný model by proto měla být ještě prozkoumána s rozmanitějšími daty.

6 Závěr

Výsledky této studie zkoumající vliv klimatického filtrování datových sad na výsledné druhové distribuční modely pomocí virtuálního druhu umožňují definovat několik závěrů a námětů pro budoucí výzkum. Především, se jednoznačně podařilo prokázat, že klimatické filtrování nevede k modelům, jež by dokázaly reflektovat realitu s vyšší přesností. Dosahují sice vyšší diskriminační schopnosti, ale na druhou stranu i výrazně nižší kalibrační schopnosti. Kromě toho lze říci, že datové sady s menším počtem druhové prezence vedou i přes vyšší rozdíly ve středních hodnotách k validnějším modelům. Prezentované závěry práce by však měly být chápány pouze jako předběžný pokus, jehož cílem bylo dokázat, že v případě validace modelů pomocí kalibračních metod nedosáhnou filtrované modely lepších výsledků v porovnání s těmi nefiltrovanými. Nezbytný je další výzkum, jenž by ověřil hlavní závěry za použití (a) jiného statistického softwaru, (b) většího počtu opakování při tvorbě modelu a za (c) robustnějšího množství datových sad odlišných: (i) logistickou křivkou použitou ke konverzi environmentální vhodnosti prostředí na prezenci/absenci druhu a to zejména použitím strmější/pozvolnější logistické křivky a (ii) větší variabilitou datových sad z hlediska záznamů o druhovém výskytu.

Otázkou samostatného výzkumu je pak získání odpovědi na zjištění, že modely založené na menším počtu záznamů druhového výskytu dosahují lepších výkonů, než ty, u nichž byly použity datové sady s vyšším počtem záznamů. A to i přesto, že ze statistického hlediska by tomu mělo být naopak.

Přehled literatury a použitých zdrojů

ACKERLY D. D., Community Assembly, Niche Conservatism, and Adaptive Evolution in Changing Environments. *International Journal of Plant Sciences*. 2003, 164(3), 165-184. DOI: 10.1086/368401. ISSN 1058-5893.

AGUIRRE-GUTIÉRREZ J., CARVALHEIRO G. L., et al., Fit-for-Purpose: Species Distribution Model Performance Depends on Evaluation Criteria – Dutch Hoverflies as a Case Study. *PLoS ONE*. 2013, 8(5), e63708-. DOI: 10.1371/journal.pone.0063708. ISSN 1932-6203.

ALLOUCHE O., TSOAR A. & KADMON R., Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*. 2006, 43(6), 1223-1232. DOI: 10.1111/j.1365-2664.2006.01214.x. ISSN 00218901.

ARAÚJO M. B. & GUISAN A., Five (or so) challenges for species distribution modelling. *Journal of Biogeography*. 2006, 33(10), 1677-1688. DOI: 10.1111/j.1365-2699.2006.01584.x. ISSN 0305-0270.

ARAÚJO M. B. & PETERSON T. A., Uses and misuses of bioclimatic envelope modelling. *Ecology*. 2012, 93(7), 1527-1539. DOI: 10.1890/11-1930.1. ISSN 0012-9658.

AUSTIN M. P., Searching for a model for use in vegetation analysis. *Vegetatio*. 1980, 42(1-3), 11-21. DOI: 10.1007/BF00048865. ISSN 0042-3106.

AUSTIN M.P., An Ecological Perspective on Biodiversity Investigations: Examples from Australian Eucalypt Forests. *Annals of the Missouri Botanical Garden*. 1998, 85(1), 2-. DOI: 10.2307/2991991. ISSN 00266493.

AUSTIN M.P., Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*. 2002, 157(2-3), 101-118. DOI: 10.1016/S0304-3800(02)00205-3. ISSN 03043800.

AUSTIN M.P. & VAN NIEL P. K., Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*. 2011, 38(1), 18. DOI: 10.1111/j.1365-2699.2010.02416.x. ISSN 03050270.

BARBET-MASSIN M., JIGUET F., et al., Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*. 2012, 3(2), 327-338. DOI: 10.1111/j.2041-210X.2011.00172.x. ISSN 2041210x.

BOMBI P. & D'AMEN M., Scaling down distribution maps from atlas data: a test of different approaches with virtual species. *Journal of Biogeography*. 2012, 39(4), 640-651. DOI: 10.1111/j.1365-2699.2011.02627.x. ISSN 03050270.

DELONG E. R., DELONG M. D. & CLARKE-PEARSON L. D., Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988, 44(3), 837-. DOI: 10.2307/2531595. ISSN 0006341x.

- DUPUTIÉ A., ZIMMERMANN E. N. & CHUINE I., Where are the wild things? Why we need better data on species distribution. *Global Ecology and Biogeography*. 2014, 23(4), 457-467. DOI: 10.1111/geb.12118. ISSN 1466822x.
- EDWARDS C. T., CUTLER D. R., ZIMMERMANN E. N., GEISER L. & ALEGRIA J., Model- based stratifications for enhancing the detection of rare ecological events. *Ecology*. 2005, 86(5), 1081-1090. DOI: 10.1890/04-0608. ISSN 0012-9658.
- ELITH J. & GRAHAM H. C., et al., Novel methods improve prediction of species' distributions from occurrence data. *Ecography*. 2006, 29(2), 129-151. DOI: 10.1111/j.2006.0906-7590.04596.x. ISSN 09067590.
- ELITH J. & GRAHAM H. C., Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*. 2009, 32(1), 66-77. DOI: 10.1111/j.1600-0587.2008.05505.x. ISSN 09067590.
- ELITH J. & LEATHWICK R. J., Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*. 2009, 40(1), 677-697. DOI: 10.1146/annurev.ecolsys.110308.120159. ISSN 1543-592x.
- ENGLER R., GUISAN A. & RECHSTEINER L., An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*. 2004, 41(2), 263-274. DOI: 10.1111/j.0021-8901.2004.00881.x. ISSN 0021-8901.
- FERRIER S. & GUISAN A., Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*. British Ecological Society, 2006, (43), 393–404.
- FOURCADE Y., ENGLER O. J., RÖDDER D., et al., Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *PLoS ONE*. 2014, 9(5), e97122-. DOI: 10.1371/journal.pone.0097122. ISSN 1932-6203.
- GIANCRISTOFARO R. A. & SALMASO L., Model performance analysis and model validation in logistic regression. *Statistica*. 2003, 63(2), 375-396. ISSN 1973-2201.
- GOODCHILD M. F. Scale in GIS: An overview. *Geomorphology*. 2011, 130(1-2), 5- 9. DOI: 10.1016/j.geomorph.2010.10.004. ISSN 0169555x.
- GUILLERA-ARROITA G., LAHOZ-MONFORT J. J., ELITH J. et al. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*. 2015, 24(3), 276-292. DOI: 10.1111/geb.12268. ISSN 1466822x.
- GUISAN A. & ZIMMERMANN E. N., Predictive habitat distribution models in ecology. *Ecological Modelling*. 2000, 135(2-3), 147-186. DOI: 10.1016/S0304-3800(00)00354-9. ISSN 03043800.
- GUISAN A. & THUILLER W., Predicting species distribution: offering more than simple habitat models. *Ecology Letters*. 2005, 8(9), 993-1009. DOI: 10.1111/j.1461-0248.2005.00792.x. ISSN 1461-023x.

- GUISAN, A., LEHMANN A., FERRIER S. et al., Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*. 2006, 43(3), 386 - 392. DOI: 10.1111/j.1365-2664.2006.01164.x. ISSN 0021-8901.
- HANSKI I., Metapopulation dynamics. *Nature*. Macmillan Publishers Ltd., 1998, 39(6), 41-49.
- HIJMANS, R. J., CAMERON E. S., L. PARRA L. J. et al., Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*. 2005, 25(15), 1965-1978. DOI: 10.1002/joc.1276. ISSN 0899-8418.
- HIRZEL A. H., HELFER V., & METRAL F., Assessing habitat-suitability models with a virtual species. *Ecological Modelling*. 2001, 145(2-3), 111-121. DOI: 10.1016/S0304-3800(01)00396-9. ISSN 03043800.
- HUTCHINSON, G. E., The Paradox of the Plankton. *The American Naturalist*. Chicago: University of Chicago Press, 1961, 95(882), 9.
- JIMÉNEZ-VALVERDE A., LOBO M. J. & HORTAL J., Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*. 2008, 14(6), 885-890. DOI: 10.1111/j.1472-4642.2008.00496.x. ISSN 13669516.
- JIMÉNEZ-VALVERDE A., Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*. 2012, 21(4), 498-507. DOI: 10.1111/j.1466-8238.2011.00683.x. ISSN 1466822x.
- KIENAST F., BOLLIGER J. & ZIMMERMANN N. E., *Advanced Landscape Ecology: Lecture Notes*. 2012.
- KRAMER-SCHADT S., NIEDBALLA J., PILGRIM D. J. et al., The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*. 2013, 19(11), 1366-1379. DOI: 10.1111/ddi.12096. ISSN 13669516.
- LAHOZ-MONFORT J. J., GUILLERA-ARROITA G. & WINTLE A. B., Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*. 2014, 23(4), 504-515. DOI: 10.1111/geb.12138. ISSN 1466822x.
- LAWSON C. R., HODGSON A. J., WILSON R. J. et al., Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*. 2014, 5(1), 54-64. DOI: 10.1111/2041-210X.12123. ISSN 2041210x.
- LEATHWICK J. R., ROWE D., RICHARDSON J. et al., Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*. 2005, 50(12), 2034-2052. DOI: 10.1111/j.1365-2427.2005.01448.x. ISSN 00465070.
- LEROY B., MEYNARD N. Ch., BELLARD C. & COURCHAMP F., Virtualspecies, an R package to generate virtual species distributions. *Ecography*. 2015, 001-009. DOI: 10.1111/ecog.01388. ISSN 09067590.

- LEVINS R. The Strategy of Model Building In Population Biology. *American Scientist*. 1966, 54(4), 421-431.
- LI X. & WANG Y., Applying various algorithms for species distribution modelling. *Integrative Zoology*. 2013, 8(2), 124-135. DOI: 10.1111/1749-4877.12000. ISSN 17494869.
- LOBO J., JIMÉNEZ-VALVERDE A. & HORTAL J., The uncertain nature of absences and their importance in species distribution modelling. *Ecography*. 2010, 33(1), 103-114. DOI: 10.1111/j.1600-0587.2009.06039.x. ISSN 09067590.
- LOISELLE B. A., JØRGENSEN M. P., CONSIGLIO T. et al., Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*. 2007, 35(1), 105-116. DOI: 10.1111/j.1365-2699.2007.01779.x. ISSN 0305-0270.
- MANEL S., WILLIAMS C. H. & ORMEROD S. J., Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*. 2001, 38(5), 921-931. DOI: 10.1046/j.1365-2664.2001.00647.x. ISSN 00218901.
- 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. DOI: 10.1111/j.1600-0587.2011.07157.x. ISSN 09067590.
- MEYNARD Ch. N., et al., Using virtual species to study species distributions and model performance. *Journal of Biogeography*. 2013, 40(1), 1-8. DOI: 10.1111/jbi.12006. ISSN 03050270.
- MILLER J. & FRANKLIN J., Modelling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*. 2002, 157(2-3), 227-247. DOI: 10.1016/S0304-3800(02)00196-5. ISSN 03043800.
- MILLER J., FRANKLIN J. & ASPINALL R., Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*. 2007, 202(3-4), 225-242. DOI: 10.1016/j.ecolmodel.2006.12.012. ISSN 03043800.
- MILLER J., Species Distribution Modelling. *Geography Compass*. 2010, 4(6), 490-509. DOI: 10.1111/j.1749-8198.2010.00351.x. ISSN 17498198.
- MOUDRÝ V., Modelling species distributions with simulated virtual species. *Journal of Biogeography*. 2015, 42(8). DOI: 10.1111/jbi.12552. ISBN 10.1111/jbi.12552.
- OSBORNE P. E., ALONSO J. C. & BRYANT R. G., Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology*. 2001, 38(2), 458-471. DOI: 10.1046/j.1365-2664.2001.00604.x. ISSN 0021-8901.
- PARRA J. L., GRAHAM C. C. & FREILE J. F., Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*. 2004, 27(3), 350-360. DOI: 10.1111/j.0906-7590.2004.03822.x. ISSN 09067590.

- PEARSON R. G., RAXWORTHY J. Ch., NAKAMURA M. & PETERSON T. A., Predicting species distributions from small numbers of occurrence records. *Journal of Biogeography*. 2007, 34(1), 102-117. DOI: 10.1111/j.1365-2699.2006.01594.x. ISSN 03050270.
- PHILLIPS S., DUDÍK M., ELITH J. et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*. 2009, 19(1), 181-197. DOI: 10.1890/07-2153.1. ISSN 1051-0761.
- SANTIKA T. Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography*. 2011, 20(1), 181-192. DOI: 10.1111/j.1466-8238.2010.00581.x. ISSN 1466822x.
- SEO C., THORNE J. H., HANNAH L. & THULLER W., Scale effects in species distribution models: implications for conservation planning under climate change. *Biology Letters*. 2009, 5(1), 39-43. DOI: 10.1098/rsbl.2008.0476. ISSN 1744-9561.
- SHESKIN David. *Handbook of parametric and nonparametric statistical procedures*. 3rd ed. Boca Raton: Chapman, c2004. ISBN 15-848-8440-1.
- SOBERÓN J., Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*. 2007, 10(12), 1115-1123. DOI: 10.1111/j.1461-0248.2007.01107.x. ISSN 1461-023x.
- STOCKWELL D., The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*. 1999, 13(2), 143-158. DOI: 10.1080/136588199241391. ISSN 1365-8816.
- STOCKWELL D. R. B. & PETERSON T. A., Effects of sample size on accuracy of species distribution models. *Ecological Modelling*. 2002, 148(1), 1-13. DOI: 10.1016/S0304-3800(01)00388-X. ISSN 03043800.
- SYFERT M. M., SMITH J. M., COOMES A. D. & ROBERTS L. D., The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. *PLoS ONE*. 2013, 8(2), e55158-. DOI: 10.1371/journal.pone.0055158. ISSN 1932-6203.
- TESSAROLO G., RANGEL F. T., ARAÚJO B. M. et al., Uncertainty associated with survey design in Species Distribution Models. *Diversity and Distributions*. 2014, 20(11), 1258-1269. DOI: 10.1111/ddi.12236. ISSN 13669516.
- VÁCLAVÍK T., *Modelování habitatu a distribuce druhů* [online]. 2013 [cit. 2016-03-13]. Dostupné z: http://tova.euweb.cz/downloads/08_KEKO_sdm.pdf
- VAN NIEL K., LAFFAN S. & LEES G. B., Effect of error in the DEM on environmental variables for predictive vegetation modelling. *Journal of Vegetation Science*. 2004, 15(6), 747-756.
- VARELA S. et al., Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*. 2014, no-no. DOI: 10.1111/j.1600-0587.2013.00441.x. ISSN 09067590.

WELK E., Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: Purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling*. 2004, 179(4), 551-567. DOI: 10.1016/j.ecolmodel.2004.04.020. ISSN 03043800.

XIAOQUAN K. et al., *Package 'sdmvspecies': Create Virtual Species for Species Distribution Modelling*. 2015.

ZANIEWSKI A. E., LEHMANN A. & McCOVERTON J., Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*. 2002, 157(2-3), 261-280. DOI: 10.1016/S0304-3800(02)00199-0. ISSN 03043800.

ZURELL D. et al., The virtual ecologist approach: simulating data and observers. *Oikos*. 2010, 119(4), 622-635. DOI: 10.1111/j.1600-0706.2009.18284.x. ISSN 00301299.

Seznam obrázků

Obr. 1 Odlišnost analytického, empirického a mechanického modelu	14
Obr. 2 Komparace základní a realizované niky v n-rozměrném prostoru	15
Obr. 3 Ukázka Gaussovy křivky	16
Obr. 4 Vztah mezi geografickým a environmentálním prostorem a mapou.....	17
Obr. 5 Srovnání prahové metody a pravděpodobnostního přístupu.....	30
Obr. 6 Schéma logistické křivky	32
Obr. 7 Ukázka definované Gaussovy křivky pro jednotlivé bioklimatické vrstvy. ...	35
Obr. 8 Ilustrace vhodného environmentální prostředí virtuálního druhu.....	35
Obr. 9 Logistická funkce použitá k transformaci vztahu druh – prostředí.....	36
Obr. 10 Příklad vlivu konverzní křivky na distribuci virtuálního druhu.....	36
Obr. 11 Geografická distribuce virtuálního druhu na Iberském poloostrově.....	37
Obr. 12 Příklad klimatického filtrování	38

Seznam tabulek

Tab. 1 Příklady možného využití SDMs	13
Tab. 2 Příklad zdrojů biologických dat používaných při tvorbě SDMs.....	19
Tab. 3 Příklad zdrojů environmentálních dat používaných při tvorbě SDMs.....	20
Tab. 4 Nejčastěji používané softwary a algoritmy	24
Tab. 5 Obecná chybová matice	25
Tab. 6 Indexy sloužící ke klasifikaci přesnosti prediktivní schopnosti modelu	26
Tab. 7 Běžně používané diskriminační metriky.....	26
Tab. 8 Příklady využití uměle vytvořených druhů.....	27
Tab. 9 Výsledné hodnoty AUC a párového T testu	40
Tab. 10 Tabulka Pearsonových korelací ve formě korelační matice	42

Seznam příloh

- Příloha A: Mapa rozsahu jednotlivých klimatických vrstev na území Iberského poloostrova
- Příloha B: Boxplot ilustrující rozsah uměle generované prezence virtuálního druhu v jednotlivých klimatických vrstvách
- Příloha C: Vizualizace aplikace klimatického filtrování na datové sady
- Příloha D: Komparace relativní predikce druhové distribuce realizované pomocí filtrovaných a nefiltrovaných dat

Příloha A: Mapa rozsahu jednotlivých klimatických vrstev na území Iberského poloostrova

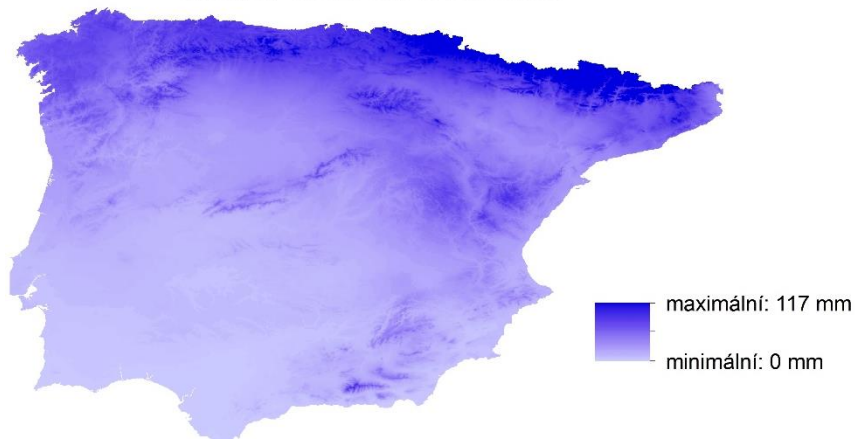
Maximální teplota v nejteplejším měsíci



Minimální teplota v nejchladnějším měsíci

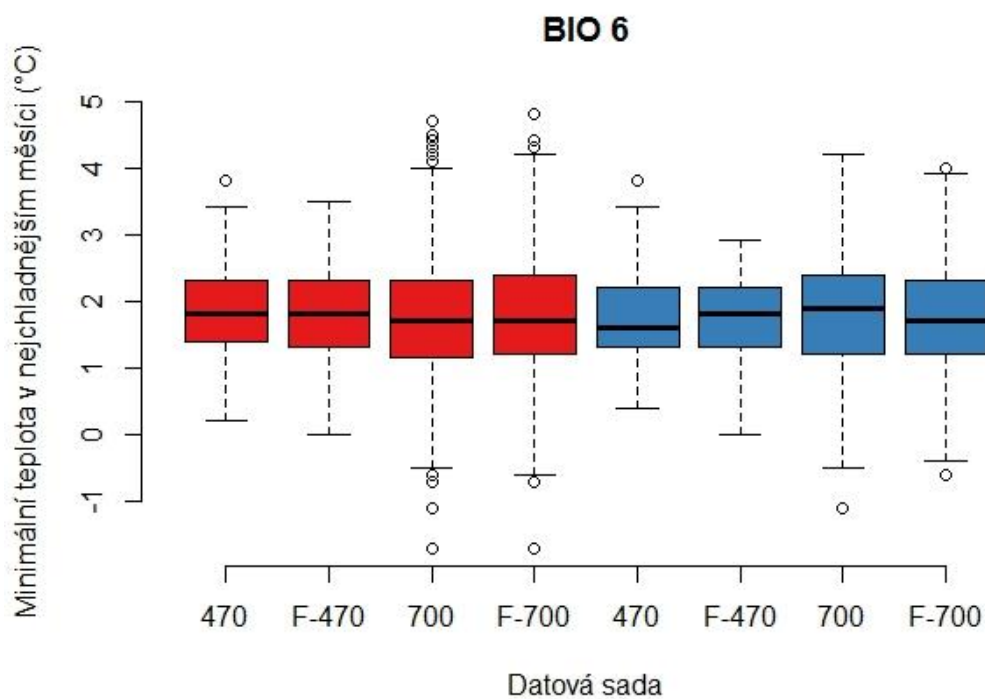
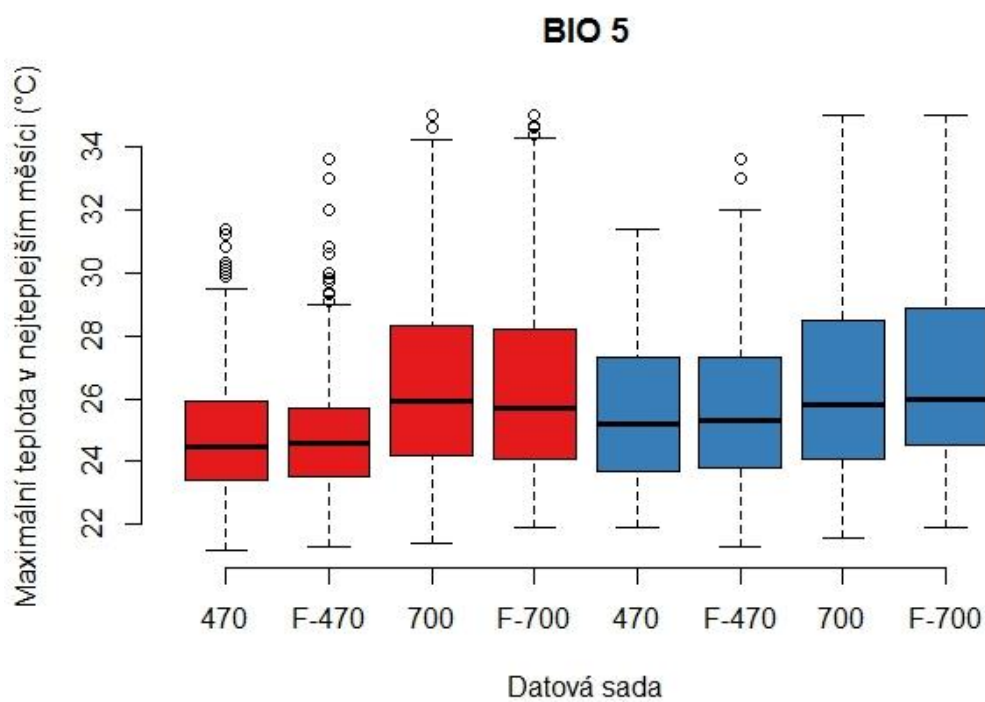


Srážkový úhrn v nejsušším měsíci

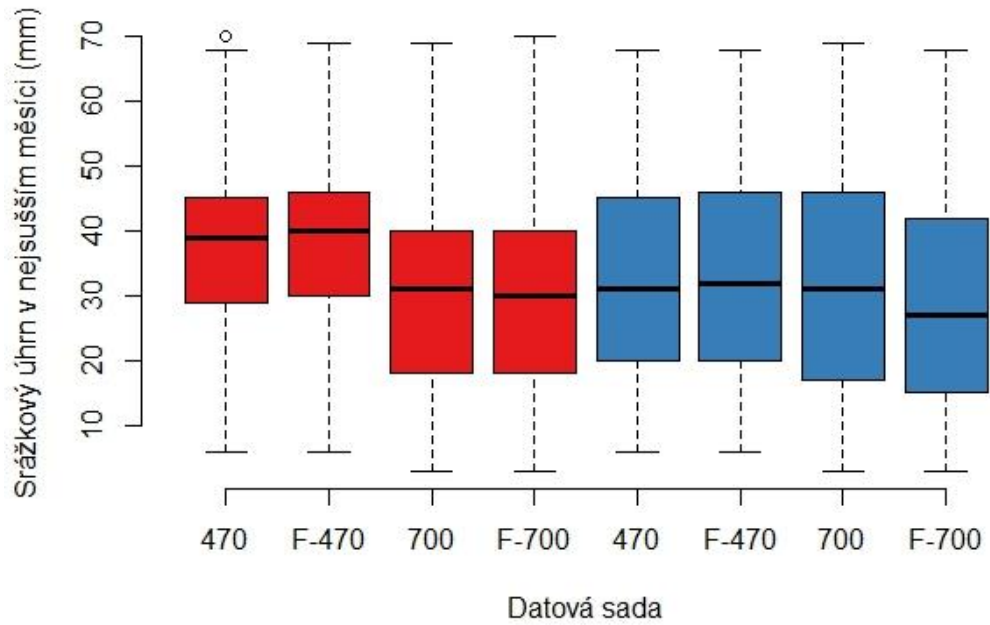


0 350 700 km

Příloha B: Boxplot ilustrující rozsah uměle generované prezence virtuálního druhu v jednotlivých klimatických vrstvách



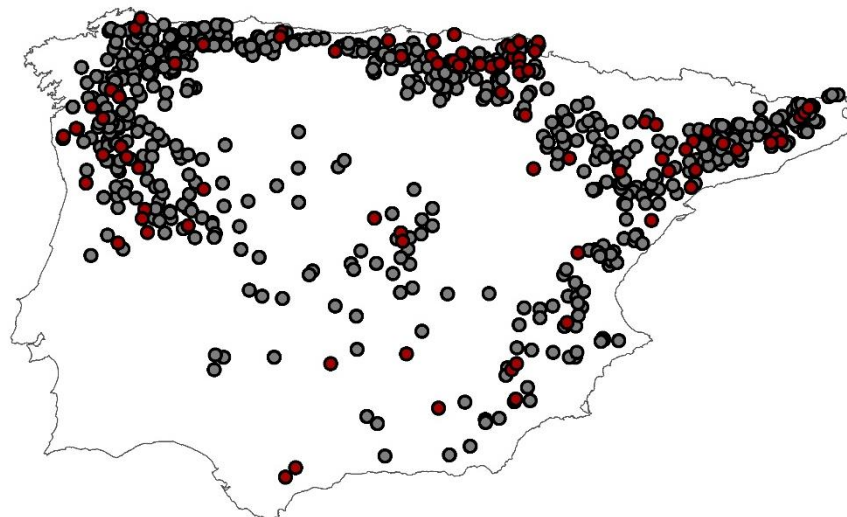
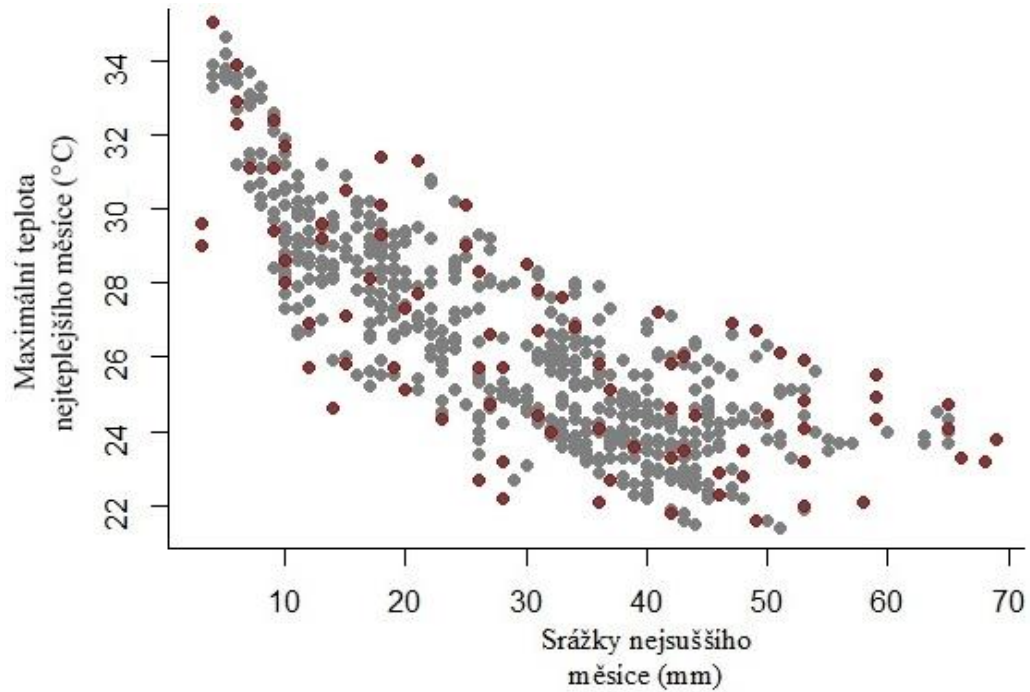
BIO 14



Osa X znázorňuje použitou datovou sadu vyjádřenou číslem (počet míst druhové prevalence), kde F symbolizuje datové sady s klimatickým filtrem. Osa Y ukazuje rozsah hodnot jednotlivých klimatických proměnných. Červená barva symbolizuje datové sady generované na základě druhové prevalence 0,1, modrá pak 0,3.

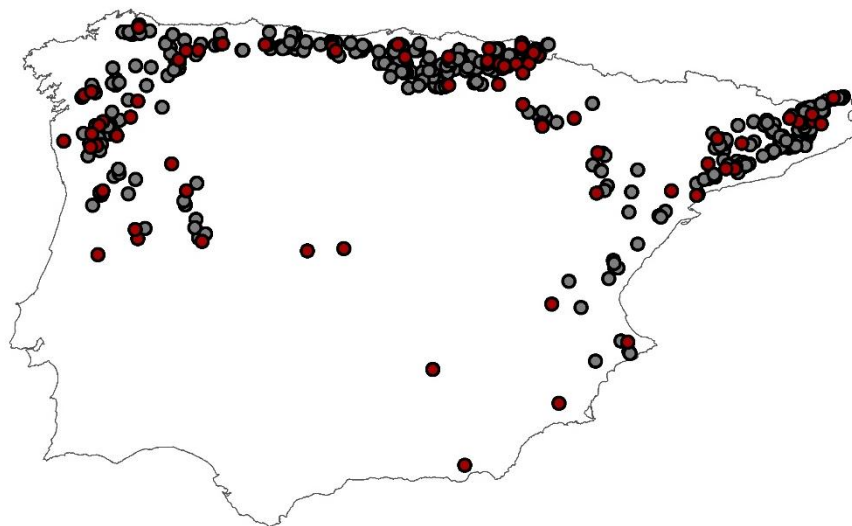
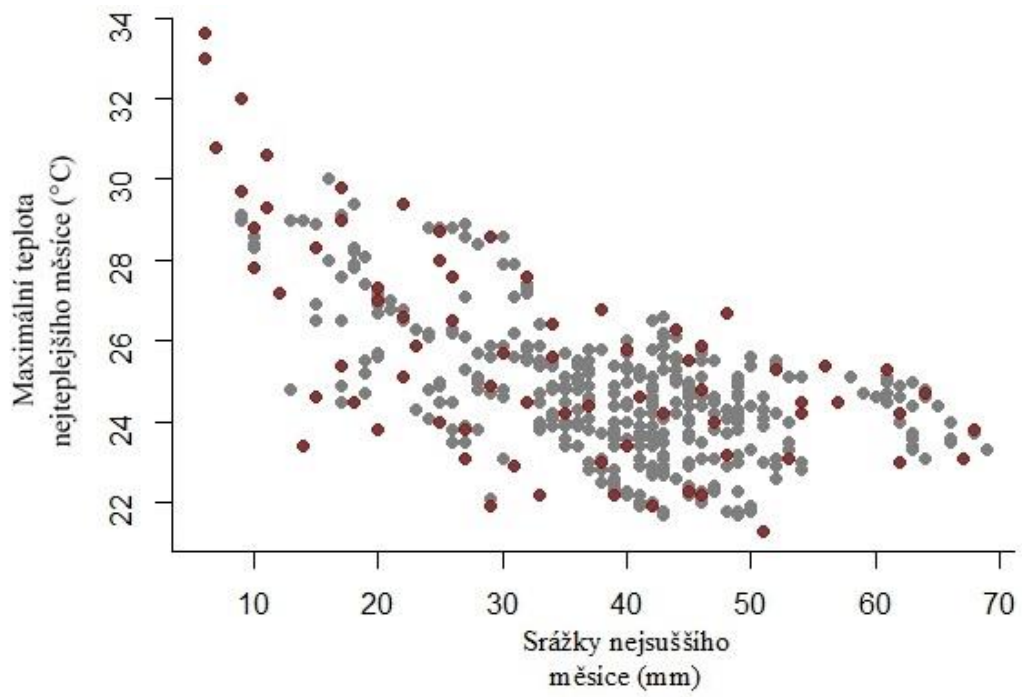
Příloha C: Vizualizace aplikace klimatického filtrování na datové sady

1) 700 prezenčních bodů generovaných na základě druhové prevalence 0,1



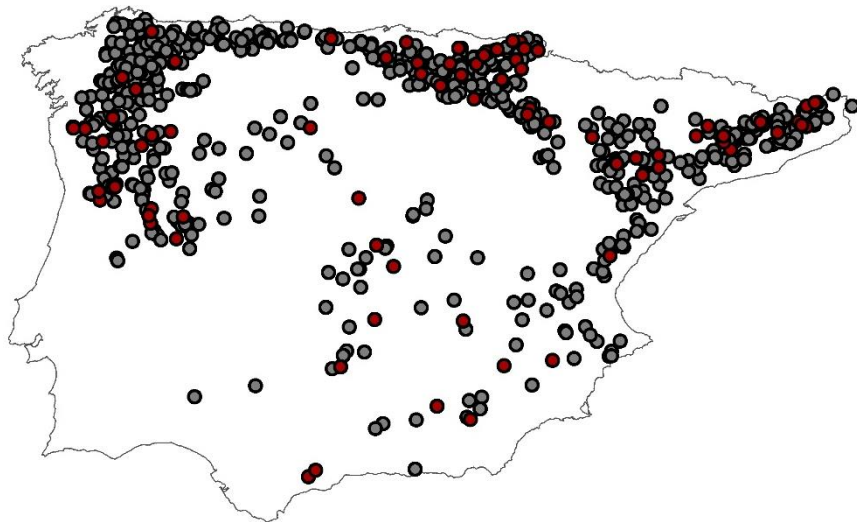
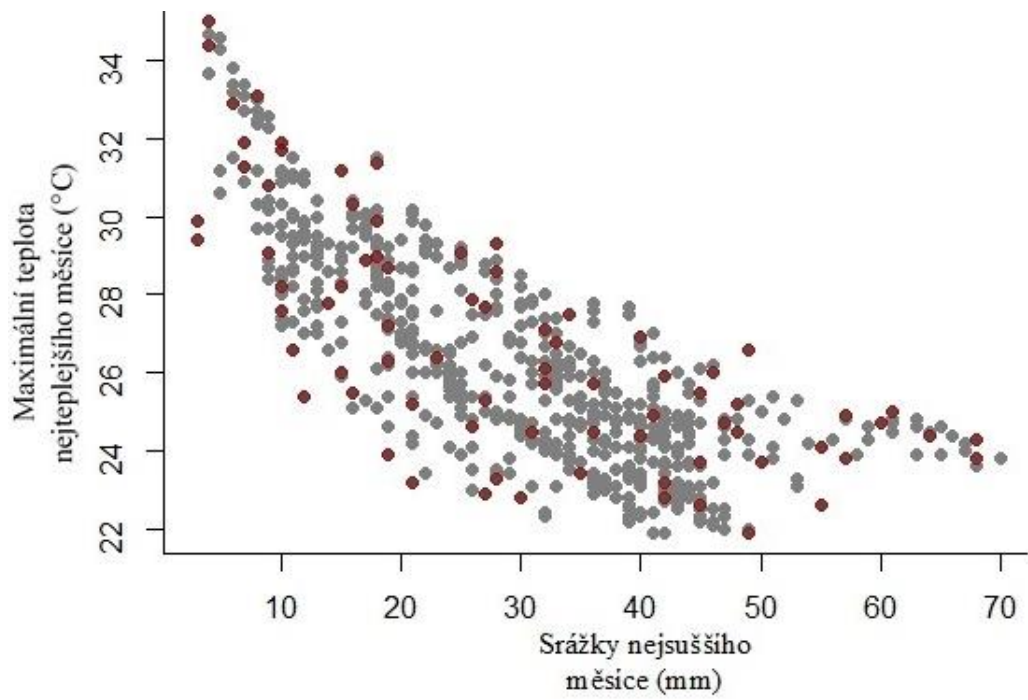
- klimatický filtr
- nefiltrovaná data
- hranice Iberského poloostrova

2) 470 prezenčních bodů generovaných na základě druhové prevalence 0,3



- klimatický filtr
- nefiltrovaná data
- hranice Iberského poloostrova

3) 470 prezenčních bodů generovaných na základě druhové prevalence 0,3

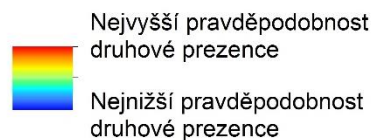
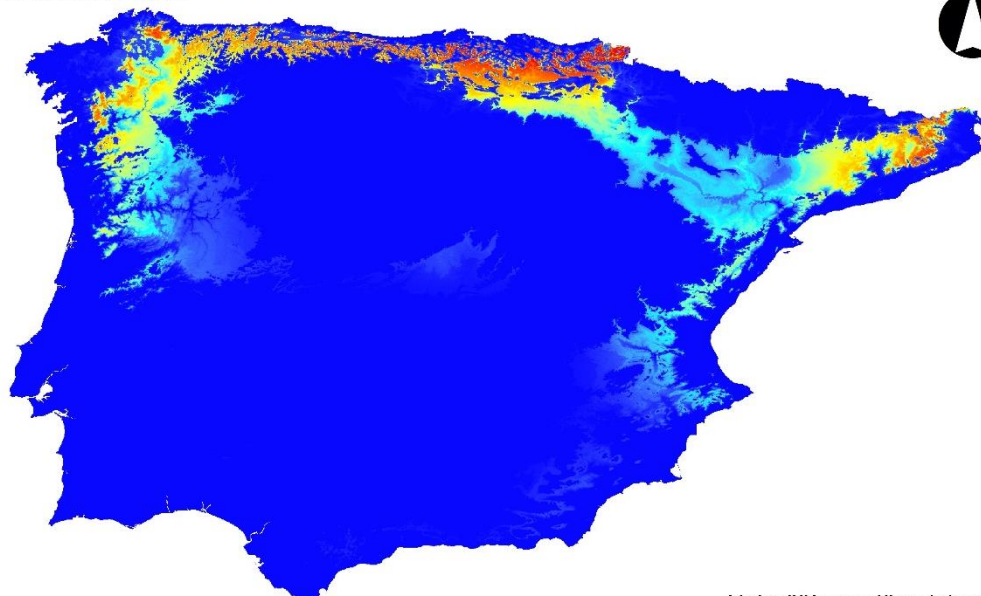


- klimatický filtr
- nefiltrovaná data
- hranice Iberského poloostrova

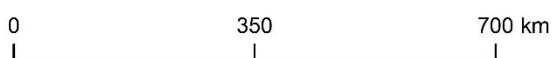
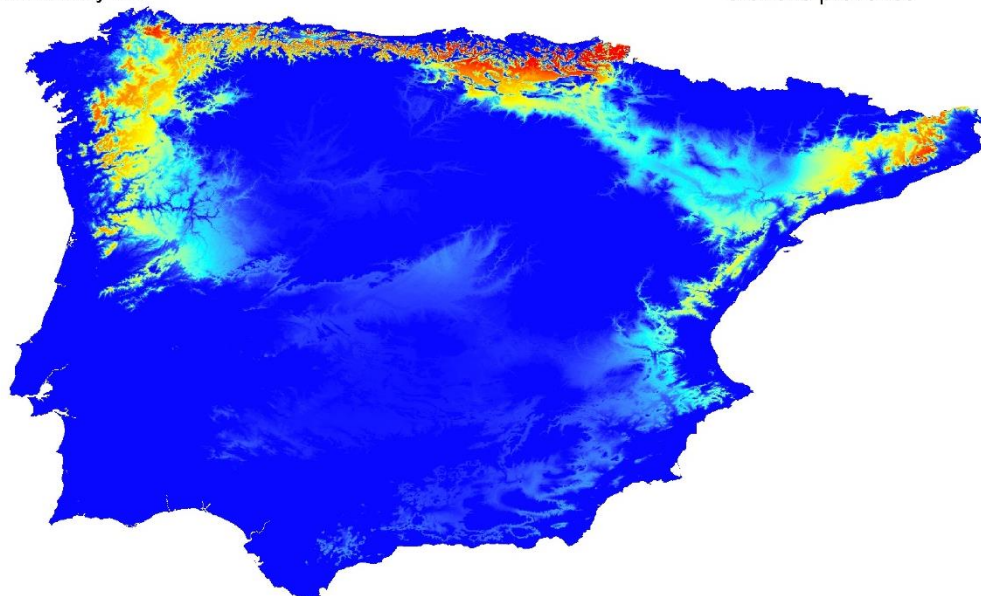
Příloha D: Komparace relativní predikce druhové distribuce realizované pomocí filtrovaných a nefiltrovaných dat

VÝSLEDNÝ MODEL RELATIVNÍ DRUHOVÉ DISTRIBUCE (470 míst s prezencí druhu, druhová prevalence 0,1)

Nefiltrovaná data

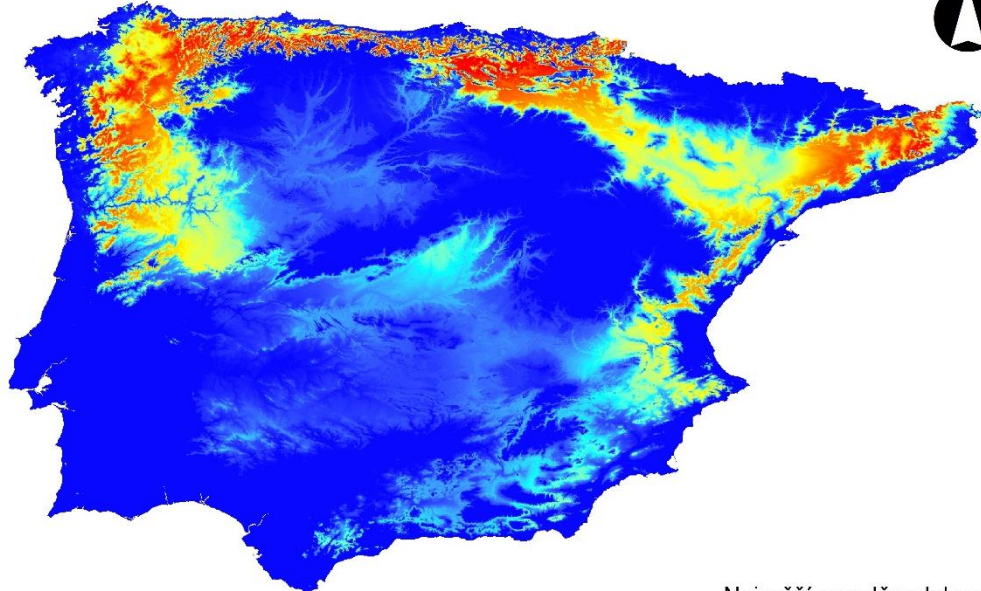


Klimatický filtr




VÝLEDNÝ MODEL RELATIVNÍ DRUHOVÉ DISTRIBUCE
(700 míst s prezencí druhu, druhová prevalence 0,1)

Nefiltrovaná data

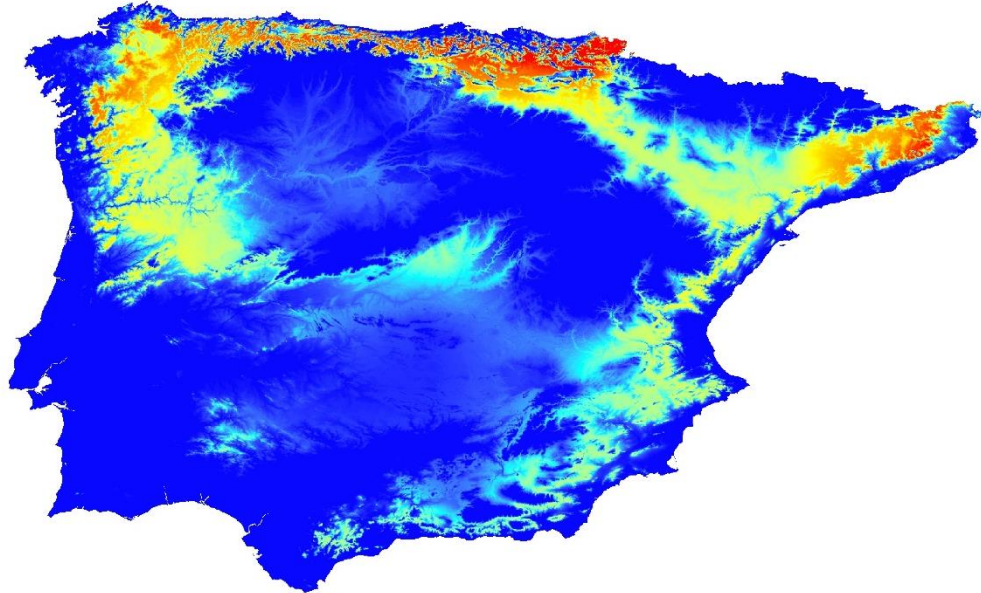


Klimatický filtr

Nejvyšší pravděpodobnost
druhov^é prezence



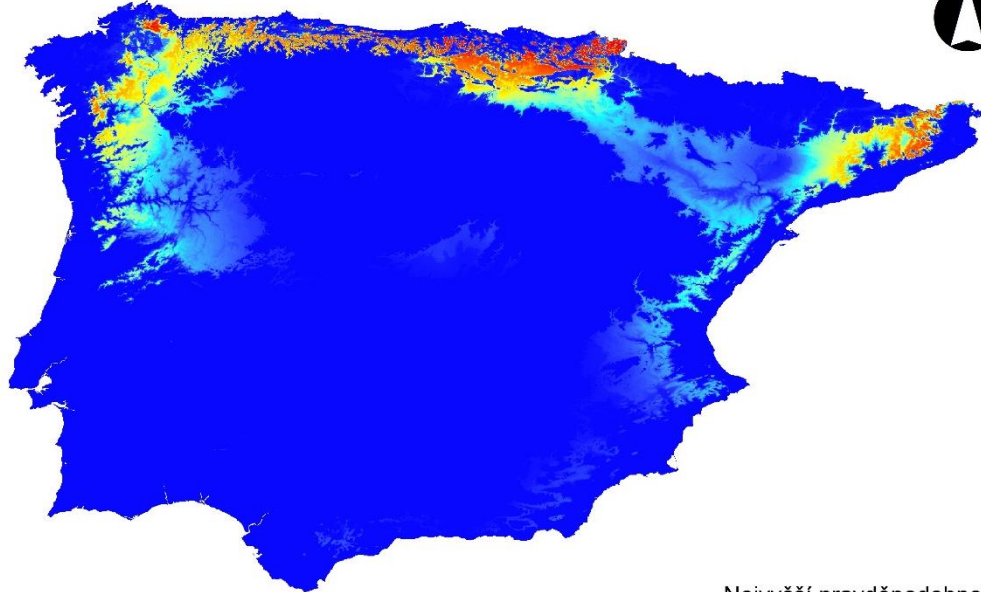
Nejnižší pravděpodobnost
druhov^é prezence



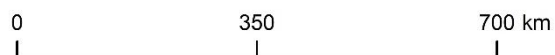
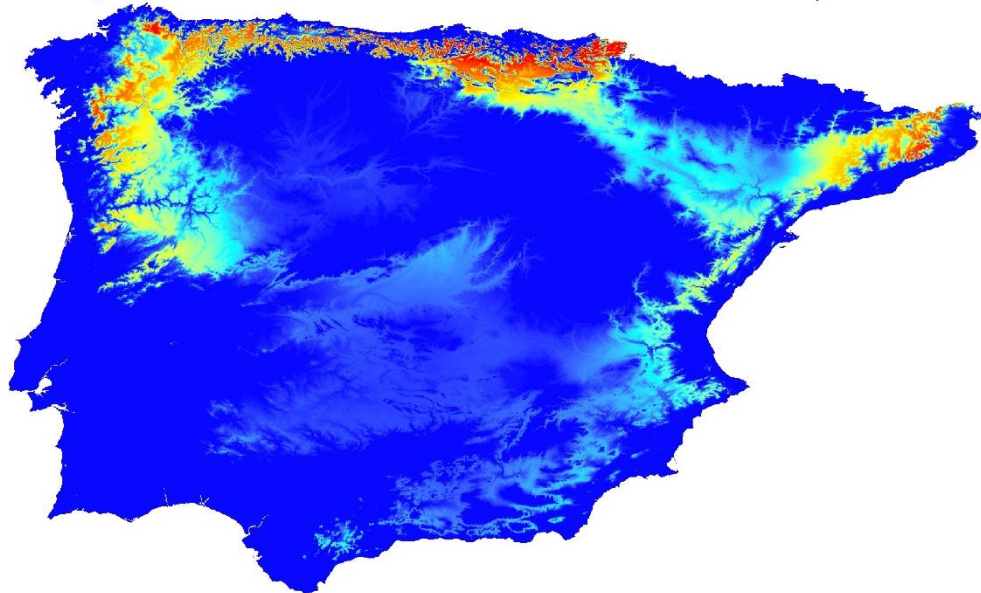
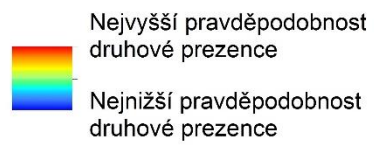
0 350 700 km

VÝSLEDNÝ MODEL RELATIVNÍ DRUHOVÉ DISTRIBUCE
(470 míst s prezencí druhu, druhová prevalence 0,3)

Nefiltrovaná data

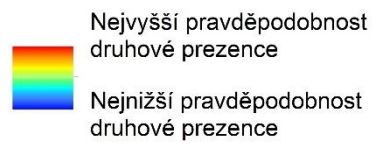
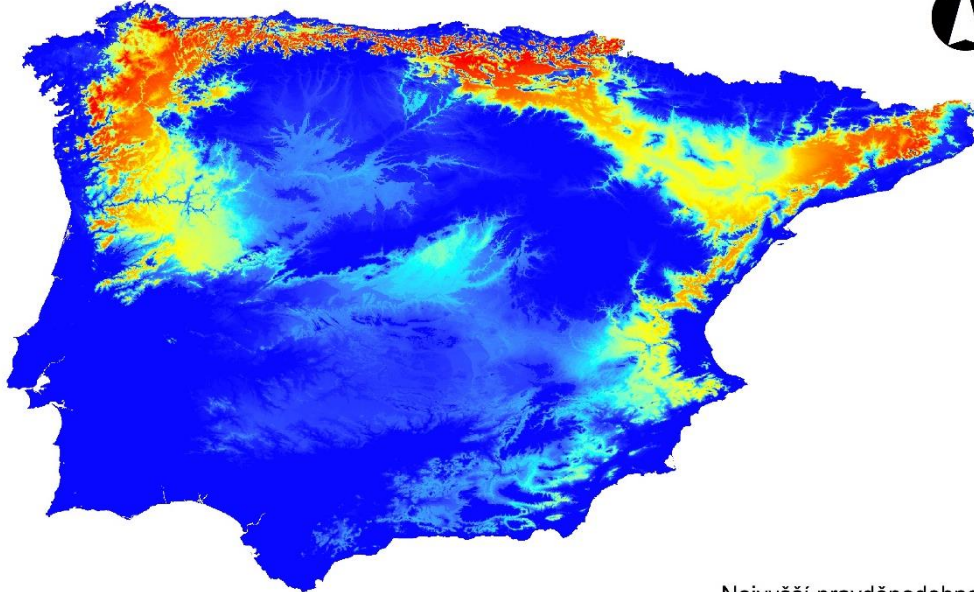


Klimatický filtr



VÝSLEDNÝ MODEL RELATIVNÍ DRUHOVÉ DISTRIBUCE
(700 míst s prezencí druhu, druhová prevalence 0,3)

Nefiltrovaná data



Klimatický filtr

