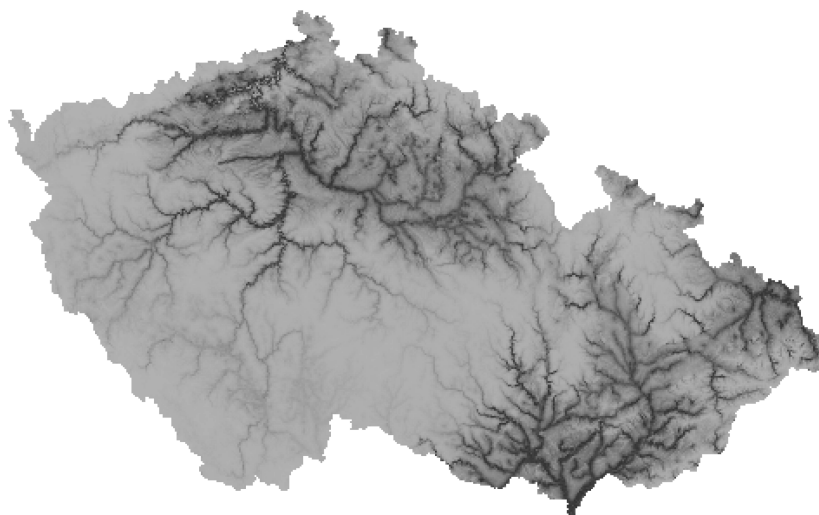


JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA BOTANIKY



**MODELOVÁNÍ POTENCIÁLNÍHO ŠÍŘENÍ INVAZNÍCH
DRUHŮ ROSTLIN V ČR: POROVNÁNÍ METOD A JEJICH
IMPLEMENTACÍ, DOSTUPNOST DAT A VLIV EKOLOGIE
DRUHU NA PŘESNOST PREDIKCE**

MAGISTERSKÁ PRÁCE

Pavel Brych

2009

VEDOUCÍ PRÁCE:

RNDr. Stanislav Mihulka, PhD.

ODBORNÝ KONZULTANT:

Mgr. Záboj Hrázský

Bibliography

Brych, P., 2009: Modelování potenciálního šíření invazních druhů rostlin v ČR: Porovnání metod a jejich implementací, dostupnost dat a vliv ekologie druhu na přesnost predikce. [Potential spatial distribution models of invasive plants in Czech republic: Comparison of methods and their implementations, data availability and effect of species ecology on prediction results. Mgr. Thesis, in Czech.] – 38 p., Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic

Annotation

Available algorithms, software and predictors data were summarized. Prediction accuracy of eight recent algorithms for potential species distribution modelling was compared by AUC. 34 invasive plant species were used for model creation. Effect of biotope preferences, life time and main long-distance transport vector type on models accuracy was examined.

Prohlašuji, že svoji diplomovou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách.

Ve Cvikově dne 25. 4. 2009

Pavel Brych

Abstract

Eight potential species distribution modelling algorithms (ANN, CTA, GBM, GAM, GLM, GARP, MARS, MDA) implemented in R statistical software (BIOMOD package) and OpenModeller were compared by quality of produced models (measured with de-facto standard AUC criteria). 272 models with 34 invasive plant species divided to several ecological categories were used and effect of this ecological characteristics on predictability of such species were examined. Only commonly available environmental predictors data were used (DTM, distance to water and climatic variables from Worldclim project)

Most stable and accurate results were given by GBM, which is consistent with previous comparisons, especially Elith et al. 2006. Standard generalized methods (GLM performed practically equally with GAM) were also very accurate and stable and it's recommended to always use them et least for comparison. Most variable results were obtained by GARP (and ANN) mainly due high proportion of failed models. Accuracy of MARS models was moderate, MDA and CTA performed poorly.

Three ecological characteristics were used: annual/perennial, invaded biotope (alluvial/other) and main vector of long distance transport (anemochory, zoochory, hydrochory and other). None of these characteristics were affected models accuracy significantly, but there were some noticeable differences especially in case of some modelling algorithms.

Poděkování

Chtěl bych touto cestou poděkovat zejména Zábaji Hrázskému za všechny konzultace a pomoc při zpracování této práce. Pokud mají získané výsledky nějaký smysl, je to z velké části jeho zásluhou. Dále děkuji Stanislavu Mihulkovi za metodické vedení a pomoc a tradičně všem ostatním, zejména mým nejbližším, díky jejichž toleranci jsem mohl této práci věnovat tolik času, abych ji dokázal dokončit.

Tato práce vznikla v rámci projektu pro vědu a výzkum Ministerstva životního prostředí SPII2D1/37/07 – Dynamika šíření invazních druhů rostlin v ČR za různých scénářů globální klimatické změny (2007-2010, MZP/SP).

Obsah

I.Úvod.....	1
I.1. Teorie modelování.....	2
I.2. Vznik invazí.....	2
I.3. Vstupní data.....	3
I.4. Modelovací algoritmy a software.....	5
I.5. Stanovené cíle.....	8
II. Metodika.....	9
II.1. Výběr vhodných algoritmů a jejich implementace.....	9
II.2. Vybrané invazní druhy.....	12
II.3. Software.....	14
II.4. Vstupní vrstvy.....	16
II.5. Vyhodnocení modelů.....	18
III. Výsledky.....	20
IV. Diskuse.....	25
IV.1. Modelovací techniky.....	25
IV.2. Vstupní datové vrstvy.....	25
IV.3. Ekologické skupiny.....	26
V. Závěr.....	27
VI. Literatura.....	28
VII. Přílohy.....	33

I. Úvod

Studiu invazních druhů rostlin i živočichů je z mnoha důvodů již dlouhou dobu věnována značná pozornost. Mezi ty prakticky nejvýznamnější, se snadno viditelným dopadem a vzájemně propojené patří zejména nebezpečí narušení, poškození či úplného zničení nebo přetvoření invadovaných ekosystémů a potenciální velké hospodářské škody, plynoucí ať už přímo z poškození přirozených ekosystémů nebo působené například na kulturních plodinách a podobně. Tyto změny mohou vést nejen ke ztrátám obrovských přírodních hodnot, ale i k velkým ztrátám finančním či přímo k ohrožení obyvatel, pokud je invazí zasažen nějaký druh nebo prostředí klíčové k jejich obživě.

Dlouhodobě také probíhá výzkum snažící se odhalit příčiny invazního chování druhů, zjistit a vyjádřit nějakým způsobem, zda a za jakých podmínek se daný druh může projevit jako invazní a z druhé strany výzkum faktorů ovlivňujících náchylnost ekosystémů k invazi a schopnost ekosystému či jednotlivých druhů s invazí se nějakým způsobem vyrovnávat.

Mohutný rozvoj a dostupnost výkonné výpočetní techniky, pracovních nástrojů i dostupnost kvalitních a podrobných dat umožňuje stále přesnější a detailnější zpracování a studium vztahů invadujících druhů k původním druhům i všem faktorům prostředí a průběhu invaze. To nám umožňuje jednak studovat a odhalovat tyto vztahy, jednak na základě matematických analýz a pravděpodobnosti předvídat rozsah a postup invaze v čase. Matematické modely typu GAM (generalized additive models), GLM (generalized linear models), CART (classification and regression trees), ANN (artificial neural network) a mnoho dalších se dnes ve spojení s GIS (geographic information systems) používají k podobným účelům v podstatě běžně ve sféře vědeckých studií a výzkumu, jejich složitost a sofistikovanost však často brání jejich praktickému využití v procesu rozhodování a přijímání příslušných opatření. Je to dáno zejména složitostí samotných metod analýzy, složitostí přípravy dat a nedostupností uživatelsky přívětivých nástrojů pro takovou práci. Toto vše se ale postupně mění.

Příkladů úspěšných modelů invazních druhů můžeme nalézt mnoho. Práce Zhu et al. 2007 odhaluje například mj. místa hrozící invaze druhem *Eupatorium andenophorum* v Číně s využitím metody GARP (genetic algorithm for rule-set production). Stejný algoritmus využívá známý projekt Lifemapper pro modelování celosvětového rozšíření mnoha druhů (Stockwell et al. 2006) či Anderson et al. 2006 k predikci oblastí Kansasu ohrožených invazí

u nás běžného druhu *Lythrum salicaria*. Endress et al. 2007 zas pomocí klasifikačních a regresních stromů zjišťují faktory ovlivňující abundanci a dominanci invazního druhu *Potentilla recta* a tak dále.

1.1. Teorie modelování

Modelování potenciálního rozšíření všech možných druhů organismů obecně na základě různě složitých matematických modelů a statistického zpracování jejich odpovědí na různé gradienty či faktory prostředí je dnes poměrně často používané a rozšířené. Jejich využití pro modelování šíření invazních druhů má mnoho specifických požadavků na zvolení vhodných metod, vstupních dat a interpretaci výsledků.

Obecně existují dva možné přístupy. Výrazně složitější a mladší metodou pokud jde o praktické využití jsou dynamické modely, založené na detailních znalostech ekologie, šíření a interakcí daného druhu a prostředí. Principem je v podstatě simulaci šíření jednotlivých jedinců, například pomocí metod buněčných automatů a pod. Jde o mechanistické modely operující s průběhem šíření v čase (Beerling et al. 1995 v Robertson et al. 2003). Ukázkou takové práce je modelování šíření *Heracleum mantegazzianum* (Nehrbass et al. 2007). Tyto metody jsou však spíše „experimentální“, příliš náročné na nutné znalosti daného druhu a bylo by obtížné vytvořit obecný model aplikovatelný na více druhů. Jejich využití je spíše vyvrcholením poznání ekologie daného druhu.

Metody založené na analýze vztahu aktuálního rozšíření (prostorové distribuce) druhu (např. presence/absence, biomasa, pokryvnost...) k různým faktorům prostředí, tzv. statické korelativní modely hledající časově nezávislou rovnováhu (Beerling et al. 1995 v Robertson 2003) jsou v tomto ohledu mnohem univerzálnější a robustnější. Nezabývají nás sice nutností snažit se co nejlépe chápat ekologii daného druhu, ale pomáhají nám tyto vztahy a vlastnosti odhalovat a k dostatečně přesné predikci jejich znalost mnohdy nevyžadují. Jejich využití má navíc poměrně dlouhou tradici, jedná se o metody dostatečně prověřené a dostupné pokud jde o softwarovou výbavu i nutná vstupní data.

1.2. Vznik invazí

Faktorů ovlivňujících vznik a úspěšnost invaze je velké množství, obecně bývají uváděny například diverzita původních druhů, intenzita přírodních či antropogenních disturbancí, dostupnost zdrojů, produktivita ekosystému, absence přirozených nepřátel

invadujícího druhu... Současné studie rovněž potvrzují, že úspěch a míra invaze je přímo korelována s intenzitou šíření semen či jiných propagulí u rostlin a pod., přičemž podmínky úspěšné invaze mohou být velmi komplexní, ovlivněné mnoha faktory současně. Zejména pochopení zdroje a cest šíření spor invazních druhů umožňuje efektivní plánování managementu a omezení negativních dopadů potenciální či probíhající invaze (např. framework pro odhad nebezpečí invaze druhů vyskytujících se v horních částech povodí toků jako vektorů šíření spor, jejichž nižší části toků se nacházejí v zájmovém území, Foxcroft et al. 2007). K nejvýznamnějším koridorům umožňujícím průnik a šíření invazních druhů patří komunikace a vodní toky (Pyšek et Prach 1994 v Foxcroft et al. 2007).

Půjdeme li ještě dále, nejefektivnější způsob jak omezit případné „škody“ způsobené invazí je zajistit, aby k ní nedošlo, např. zabráněním importu či šíření druhů, u kterých je nebezpečí, že se stanou invazními, příliš vysoké. Je velmi obtížné dosáhnout při takto obecných odhadech rozumné míry generalizace, můžeme však vycházet z faktu, že základním faktorem pro odhad potenciální invaze je klimatická podobnost lokality a informace, zda se daný druh již někde jinde chová jako invazní (Richardson et Thuiller 2007).

1.3. Vstupní data

Pro vytvoření modelu bývají obvykle používána data presenční/absenční, údaje o absolutní pokryvnosti či biomase. To může vést k podhodnocení významu v místech, kde např. celková pokryvnost je výrazně nižší než 100 % (třeba některé aridní oblasti a pod.). Z tohoto pohledu se jeví jako vhodnější použití relativní pokryvnosti či biomasy (Lundholm et Larson 2004 v Endress et al. 2007).

U invazních druhů v „nepůvodním“ prostředí je problematický už výběr vstupních dat prezence/absence sledovaného druhu. Invaze může z různých důvodů probíhat v prostředí, které se v mnoha parametrech (více či méně významných pro daný druh) může lišit od prostředí považovaného pro něj za přirozené. Dochází tak k posunu realizovaných nik takovým způsobem, že výsledný model rozšíření potenciální invaze vytvořený pouze na základě dat z lokalit jejich přirozeného výskytu může být nepřesný či podhodnocený (Broennimann et al. 2007). Pokud je to možné, je tedy nutné kombinovat tato data z lokalit přirozeného výskytu s informacemi z lokalit počínající invaze či jiných podobných lokalit, kde invaze již probíhá či proběhla. V mnoha případech může být výhodnější nebo nezbytné

použít pouze data z invadovaného území, ať už z důvodů přílišné odlišnosti přirozeného prostředí nebo nedostupnosti těchto dat. Modely založené na takových datech jsou pak výrazně přesnější (Loo et al. 2007).

Rovněž environmentální data z přední fronty invaze jsou vhodná spíše jako doplňující, zachycují totiž pravděpodobně jen malou část niky daného druhu a často bývají autokorelované s gradienty korespondujícími se směrem invaze (Welk 2004 v Anderson et al. 2006). Prostorová autokorelace v datech může vznikat i z mnoha jiných důvodů, včetně vnějších přírodních a historických faktorů omezujících pohyb jedinců, způsobu rozmnožování, nebo je způsobena statisticky nevhodným vzorkováním a pod. (problém zejména položek z herbářů (Loiselle et al. 2008), fytoecologických databází a pod., kdy snadno dostupná, zajímavá či jinak vybočující místa bývají zachycena mnoha vzorky a naopak rozsáhlá území např. botanicky „nezajímavá“ nemají jediný záznam). V případě invazních druhů je navíc herbáře a fytoecologické databáze nutné považovat jen za doplňující zdroj dat i z toho důvodu, že biotopy ve větší míře zasažené nějakou invazí bývají považovány za „degradované“, mívají nižší druhovou bohatost a jsou tak na okraji zájmu. Podobné databáze proto v případě mnoha invazních druhů určitě nezahrnují podstatnou část areálů jejich výskytu. Přehled v současnosti dostupných a používaných metod pro odstranění či redukci prostorové autokorelace v datech nabízí např. Dormann et al. 2007, další srovnání a informace je možné nalézt v článku Segurado et al. 2006. Vhodným řešením tohoto problému může být využití vegetačního indexu odvozeného ze satelitních snímků namísto přímých environmentálních proměnných (Anderson et al. 2006). Nedostatečné vzorkování a ignorování potenciálního nerovnovážného stavu invadujících druhů vůči prostředí vede navíc ke zvýšenému riziku chyby falešného zamítnutí výskytu (Zhu et al. 2007).

Dalším velmi důležitým predikčním faktorem je třeba historické využití lokality, které umožňuje v určitých případech predikovat potenciální invazi, aniž bychom znali konkrétní ekologické faktory (Endress et al. 2007). Ačkoliv přímé či nepřímé „přirozené“ environmentální faktory jsou ve většině případů rozhodující pro výskyt či absenci daného druhu, historické a např. kompetiční faktory nelze zanedbávat (Costa et al. 2008). Pro invazní druhy mohou být antropogenní faktory dokonce statisticky průkaznějšími a vhodnějšími prediktory, než klasické přímé či nepřímé environmentální proměnné (Sánchez-Flores 2007).

Většina současných studií opomíjí důležité faktory druhové distribuce, jako například migraci či biotické i environmentální interakce (Guisan et al. 2006). Přitom tyto mohou být důležité nejen v mikroekologickém ale i makroekologickém měřítku (Araújo et Luoto 2007). Z teoretického hlediska je také problém fakt, že většina studií nestanovuje před provedením samotného modelování předpokládané odpovědi druhů, často ani neověřuje, zda fitované odpovědi a vypočítané distribuce mají nějaké biologické/ekologické opodstatnění (více např. Guisan et Thuiller 2005; Austin 2007). Detailněji se těmito i mnohými dalšími otázkami souvisejícími např. s měřítkem/rozlišením zabývá Guisan et Thuiller 2005. V praxi nám ale často jde především o co nejpřesnější předpověď možného rozšíření daného druhu a odhad rychlosti tohoto šíření, případně jeho potenciální rychlosti v závislosti na různých faktorech.

Výrazného zvýšení přesnosti predikce lze dosáhnout kombinací tradičních modelů založených na ekologických nikách a modelů zahrnujících vektory šíření daného invazního druhu. Např. Herborg et al. 2007 toto využívá při modelování introdukce a invaze asijského kraba *Eriocheir sinensis*, který je na území Ameriky transportován spolu s balastní vodou, používanou jako stabilizační zátěž pro prázdné nákladní lodě, kde právě trasy těchto lodí jako vektor šíření výrazně omezují potenciálně invadované lokality.

1.4. Modelovací algoritmy a software

Existuje poměrně velké množství metod pro vytváření modelů potenciálního rozšíření druhů, od klasických a v základu jednoduchých GAM či GLM po relativně mladé metody umožňující nejen zachytit velmi komplexní odpovědi druhů, ale i jejich interakce a pod. (např. neuronové sítě, CART, MARS (multivariate adaptive regression splines), GARP, BRT (boosted regression trees), MAXENT (maximum entropy). Jak ukazuje Pande et al. 2007, k predikci pravděpodobnosti úspěšné invaze cizího druhu lze rovněž využít jednoduchou mapovou algebru („map algebra“).

Dalším problémem při výběru vhodné metody je skutečnost, že neexistuje žádný jednoduchý způsob, jak zjistit, která z nich je nejlepší. Hodnocení modelů vytvořených jednotlivými metodami je obtížné i pro klasické modelování potenciálního rozšíření, natož při aplikaci na invazní druhy či změny klimatu (Araújo et Guisan 2006; Elith et al. 2006). Základní potíž je v tom, že v případě reálných dat nasbíraných v terénu nikdy neznáme „pravdu“ o tom, jaká je jejich skutečná distribuce nebo odpověď druhu na zvolené proměnné.

Logickým řešením je testování schopností jednotlivých technik (a jejich srovnání) na uměle vytvořeném modelu, jehož parametry známe. Problém je ovšem v tom, že netušíme, jak moc se tento model blíží situacím, se kterými se potýkáme v reálných datech (Austin et al. 2006).

Ještě složitější je situace v případě, že potřebujeme vytvářet modely přenositelné v čase nebo prostoru (například při modelování důsledků klimatických změn). Obecně lze říci, že v těchto případech je výhodnější použít jednodušší metody s menší tendencí k overfittingu (například GLM) a využívat spíše přímé prediktory (Randin et al. 2006), protože nepřímé (např. nadmořská výška) mohou za různých podmínek znamenat např. různou teplotu či vlhkost. I přesto však takovéto modely po přenosu selhávají (Randin et al. 2006), zejména bioklimatické modely jsou přenositelné jen velmi omezeně (Mitikka et al. 2008).

Při modelování potenciálního rozšíření invadujícího druhu je pravděpodobné, že se často budeme potýkat s problémem malého počtu pozorování (v počáteční fázi invaze, pokud nebudeme mít k dispozici např. analogické území v pokročilejší fázi invaze). Obecně se uvádí, že právě taková situace je silnou stránkou GARP, nedávná srovnání však naznačují, že alespoň pokud jde o současnou „standardní“ implementaci v programu DesktopGarp, jde o metodu překonanou. Poněkud lepších výsledků dosahuje nově implementovaná verze v softwaru OpenModeller (Elith et al. 2006). Zajímavou alternativou v případě malého množství pozorování je algoritmus MAXENT (Phillips et al. 2006 v Pearson et al. 2007). Ve srovnání Desktop GARP vs. MAXENT při modelování madagaskarských ještěrek vychází MAXENT jednoznačně jako úspěšnější navíc s tou výhodou, že jde o deterministický algoritmus (Pearson et al. 2007). K podobnému závěru dospěli Sérgio et al. 2007.

Ať už na základě testování na počítačem vygenerovaných družích se známou odpovědí na vygenerované proměnné či na reálných datech, existuje množství srovnání různých modelovacích technik (např. Thuiller 2003; Elith et al. 2006; Tsoar et al. 2007, Meynard et Quinn 2007), nejčastěji na základě hodnot AUC (area under curvature). Pomineme-li, že AUC možná není tím ideálním kritériem pro testování výsledků modelovacích technik (Lobo 2008), po prostudování těchto článků dojdeme k závěru, že žádný jednoznačný soud učinit nelze. Ačkoliv např. dle Elith et al 2006 moderní modelovací techniky přesvědčivě a konzistentně překonávají starší a zavedené metody a Willems et al. 2008 uvádí ANN modely jako úspěšnější než GLM, Meynard 2007 doporučuje využití GLM resp. GAM, které v jeho srovnání s GARP a CRT vychází lépe (zejména kvůli své větší univerzálnosti a menší tendenci k „overfittingu“. Heikkinen et al. 2007 poukazují však na to, že například při studiu

chování druhů při změnách klimatu či přírodních podmínek obecně může být úspěšnost těchto modelů výrazně nižší. Robustnější metody jako GAM jsou navíc méně náchylné k ovlivnění prostorovou autokorelací (Segurado et al. 2006). Rovněž Heikinnen et al. 2007 ukazuje, že GAM/GLM jsou na základě AUC robustnější ve srovnání s ANN a MARS, doporučuje však zejména MARS používat spolu s GLM/GAM (viz Muñoz et Felicísimo 2004). Konečně Leathwick et al. 2006 dochází k závěru, že GAM a MARS jsou naprosto srovnatelné co se týče statistického vyhodnocení a volbu vhodné metody spíše na dalších požadavcích (výpočetní náročnost, transparentnost, atd...). Zajímavý je také algoritmus ENFA (ecological niche factor analysis), použitý např. v Sérgio et al. 2007. Umožňuje sledovat přispění jednotlivých proměnných do modelu a odhalovat tak zajímavé ekologické souvislosti. Funguje na principu PCA.

V současné době je vždy vhodné použít více modelovacích technik a porovnat je na konkrétním případě či konkrétních modelovaných druzích alespoň v nějaké pilotní studii a podle výsledků zvolit nejvhodnější metodu/metody. Z výše zmíněných i jiných prací totiž jednoznačně vyplývá, že výsledky (úspěšnost, přesnost,...) modelování se výrazně liší při použití různých algoritmů a parametrů/postupů.

Při použití stejných parametrů ale různých modelů navíc často dostaneme různé výsledky (Araújo et Guisan 2006). Výsledky některých prací také ukazují, že distribuční vlastnosti modelovaných druhů (Tsoar et al. 2007; Thuiller 2003) a schopnosti/zkušenosti člověka provádějícího modelování (Austin et al. 2006) mohou mít větší vliv na přesnost výsledného modelu, než zvolená metoda modelování.

1.5. Stanovené cíle

Cílem této úvodní studie je nalezení vhodných algoritmů, jejich implementací a zdrojů vstupních dat, které by teoreticky bylo možné využít k sestavení jakéhosi obecného frameworku, umožňujícího uživatelské vytváření map potenciálního rozšíření invazních druhů. Následně zejména srovnání vybraných metod z hlediska kvality jimi vytvářených modelů a vyhodnocení vlivu vybraných ekologických charakteristik testovacích invazních druhů na možnosti jejich predikce. Hlavní otázky jsou:

1. Jaké algoritmy jsou nejvhodnější pro predikci invazních druhů rostlin na území ČR?
2. Jaké konkrétní implementace umožní jejich co nejširší využití a integraci?
3. Jaká je dostupnost potenciálních vstupních dat pro území celé ČR?
4. Ovlivňují vybrané ekologické charakteristiky predikovatelnost invazních druhů rostlin?

II. Metodika

II.1. Výběr vhodných algoritmů a jejich implementace

Zvolené algoritmy musí být dostatečně robustní zejména vůči problematické distribuci vstupních dat o rozšíření druhů a jejich množství. Jejich výpočetní náročnost musí být taková, aby bylo možné získat výsledky v rozumném čase. Implementace zvolených algoritmů musí být uživatelsky přijatelná a pokud možno kompletní, včetně vyhodnocení úspěšnosti a dalších parametrů na výstupu. Algoritmy musí být dostupné ve formě aplikací či knihoven, jejichž licence umožní co nejširší využití a které bude možné integrovat s dalším softwarem (např. GIS). Zvolená softwarová řešení by měla mít v rámci možností zajištěnou podporu ze strany vývojářů (mělo by jít o živé projekty, ať už komerční, free či opensource).

Z dostupných zdrojů byl sestaven přehled používaných algoritmů a jejich implementací (tab. 1), výběr vhodných algoritmů byl prováděn primárně na základě aktuálních odborných článků. Nejkomplexnější přehled včetně kvalitního otestování na mnoha druzích uvádí Elith et. al, 2006. Tabulka 1 uvádí přehled běžně používaných modelovacích metod, jejich implementací a referencí.

Tabulka 1: Přehled používaných modelovacích metod, jejich implementací a referencí. Kurzívou označené metody nebyly vybrány k dalšímu testování.

Metoda	Implementace	Poznámka	Reference
Generalized linear model (GLM)	R		McCullagh and Nelder (1989), Lehmann et al. (2003), Thuiller (2003)
Generalized additive model (GAM)	R		Hastie (1991), Lehmann et al. (2003), Thuiller (2003)
<i>Environmental envelope (CLIM)</i>	<i>OpenModeller</i>	<i>Systémy založené čistě na environmental envelopes, nedosahují výsledků srovnatelných s ostatními metodami</i>	<i>Busby (1991), Kadmon et al. (2003)</i>
<i>BIOCLIM</i>	<i>BIOCLIM, DIVA-GIS</i>		<i>Busby (1991)</i>
<i>DOMAIN</i>	<i>DIVA-GIS</i>		<i>Carpenter et al. (1993)</i>
Multivariate adaptive regression splines (MARS)	R		Friedman (1991), Ennis et al. (1998), Moisen and Frescino (2002), Yen et al. (2004)
GARP (Genetic algorithm for rule-set production)	GARP, OpenModeller		Stockwell et Peters (1999), Irfan-Ullah (2007), Chen (2007), Fiaboe et al. (2006), Stockwell et al. (2006), Sweeney et al. (2007), Willems et al. (2008)
ANN (Artificial neural networks)	OpenModeller, R		Thuiller (2003)
CTA (Classification tree analysis)	BIOMOD pro R		Thuiller (2003), Munoz and Felicísimo (2004)
BRT (Boosted regression trees)	BIOMOD pro R		Elith et al. (2006)
<i>RF (Random Forest)</i>	<i>R</i>	<i>Výpočetně náročné</i>	<i>Breiman (2002), Elith et al. (2006)</i>
MDA (Mixture discriminant analysis)	R		Manel et al. (1999)

Z těchto metod byly vybrány jako nejslibnější níže popsané algoritmy. Výběr byl prováděn především na základě výše uvedené rešerše existujících prací. Překlad názvů jednotlivých metod může být v některých případech problematický, proto jsou uváděny zavedené anglické termíny, ze kterých vychází i příslušné zkratky.

Výpočty byly prováděny se standardním/doporučeným nastavením jednotlivých parametrů, nebyla prováděna žádná zvláštní kalibrace a úpravy.

Generalized linear models (GLM) jsou matematickým rozšířením lineárních modelů, založené na vztahu průměru vysvětlované proměnné a lineární kombinaci vysvětlujících proměnných – tzv. „link funkce“. Data mohou pocházet z různých rozdělení včetně normálního, binomického, Poissonova, negativního binomického či gamma. (Guisan et al. 2002).

Generalized additive models (GAM) je neparametrické rozšíření GLM. Jejich předpoklady jsou hladký průběh a aditivita proměnných. Stejně jako GLM využívají „link funkci“ k propojení průměru závislé proměnné a „vyhlazení“ funkce vysvětlujících proměnných. Hlavní výhodou GAM je jejich schopnost vyrovnat se se silně nelineární odpovědí na vysvětlující proměnné. Jsou založeny na vztahu průměru vysvětlované a lineární kombinaci vysvětlujících proměnných (Guisan et al. 2002), proto jsou GAM často považovány za založené spíše na datech než na nějakém teoretickém modelu. (Yee et Mitchell 1991 v Guisan et al. 2002). Pro vytváření modelů potenciálního rozšíření druhů jsou ve spojení s GIS využívány již dlouhou dobu (z recentních např. Loots et Koubbi 2007 či Moisen et al. 2006).

Generalized boosted models (GBM) -> Boosting regression trees (BRT) – moderní metoda, která namísto jediného složitějšího modelu snažícího se vysvětlit co největší množství variability (např. GLM) vytváří a kombinuje větší množství relativně jednoduchých modelů (v tomto případě klasifikačních stromů založených vždy na jedné vysvětlující proměnné). Tato metoda bere automaticky v potaz možné interakce. Může být náchylná k overfittingu.

GARP (Genetic algorithm for rule-set production) je možné představit si jako sadu pravidel schopnou poradit si s málo strukturovanými daty nevhodnými pro klasické statistické metody a gradientové analýzy (Sánchez-Flores 2007). GARP pracuje v jednotlivých iteracích výběru pravidel. Nejprve je zvolena jedna z možných variant (např. logistická regrese, bioklimatické pravidlo...), ta je aplikována na výběr trénovacích bodů. Na základě této aplikace je vytvořeno pravidlo, které je otestováno na testovacích bodech presence a pseudoabsence. Změna v přesnosti predikce mezi iteracemi je pak využita k vyhodnocení, zda dané pravidlo zahrnout do modelu či nikoliv. Výpočet končí po zadaném počtu iterací nebo když dojde ke kovergenci. Příkladem jeho využití pro modelování potenciální invaze na základě ekologických nik je práce Peterson 2003. Anderson et al. 2003 uvádí jeden z možných způsobů výběru optimálního modelu.

CTA (Classification tree analysis) nejsou, podobně jako GAM, založeny na předem určeném vztahu vysvětlujících a vysvětlovaných proměnných. Data jsou v jednotlivých iteracích rozdělována na skupiny na základě jednotlivých proměnných takovým způsobem, aby byla co nejvíce zachována homogenita v rámci skupiny (maximálně snížena variance) při zachování co nejmenšího počtu skupin. Výsledkem je binární strom.

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

ANN (Artificial neural networks) – neuronové sítě – jsou systémy chovající se jako nelineární regresní modely s mnoha parametry, schopné aproximovat libovolnou hladkou funkci, snažící se do určité míry simulovat funkci lidského mozku. Jde o velmi univerzální a flexibilní model pro aproximaci libovolných dat (Lek et Guegan 1999 v Luoto et Hjort 2005) bez nutnosti znát přesné vztahy, distribuce a odpovědi proměnných. Pro svou složitost jsou často chápány jako jakási černá skříňka, na jejíž vstup připojíme data a na výstupu čteme výsledky. ANN jsou stále častěji využívány nejen v modelování potenciálního šíření druhů, ale v jakékoliv analýze či klasifikaci dat. Ve srovnání s ostatními jde stále o mladou metodu, její výsledky však bývají s jinými přinejmenším srovnatelné (Luoto et Hjort 2005), často i lepší.

Mixture discriminant analysis (MDA) – klasifikační metoda založená na „mixture models“. Jde o rozšíření „linear discriminant analysis“, využívající ke klasifikaci kombinaci několika (např. normálních) rozdělení.

II.2. Vybrané invazní druhy

Výběr druhů a jejich charakteristiky (tab. 2) pro testování vlivu ekologických vlastností na možnosti predikce vychází z materiálů probíhajícího projektu SPII2D1/37/07 – Dynamika šíření invazních druhů rostlin v ČR za různých scénářů globální klimatické změny (2007-2010, MZP/SP). Chybějící data byla doplněna na základě informací z odborné literatury – Slavík et al 1988 – 2004.

Tabulka 2: Vybrané druhy, na kterých bylo provedeno porovnání modelovacích technik a vyhodnocen vliv ekologických vlastností (**Biotop:** R – ruderalní, N – nivní, L – lesní; **Délka života:** T – trvalky, J – jednoletky; **Způsob šíření:** A – anemochorní, N – antropochorní/náhodné, Z – zoochorní, H – hydrochorní).

Jméno	Biotop	Délka života	Hlavní způsob šíření na velké vzdálenosti	Počet výskytů v rastru 1x1 km
Acer negundo	R, N	T	A	3567
Ailanthus altissima	R	T	A	222
Amaranthus retroflexus	R	J	N	1089
Amorpha fruticosa	R	T	H	76
Bidens frondosa	R, N	J	Z	1350
Bunias orientalis	R	T	N	138
Conyza canadensis	R	J	A	1745
Digitalis purpurea	R	T	Z	623
Echinocystis lobata	R, N	J	H	87
Epilobium ciliatum	R, N	T	A	1644
Galinsoga parviflora	R	J	A	1078
Galinsoga quadriradiata	R	J	A	576
Geranium pyrenaicum	R	T	N	196
Helianthus tuberosus	R, N	T	N	626
Heracleum mantegazzianum	R	T	H	891
Impatiens glandulifera	R, N	J	H	2783
Impatiens parviflora	R	J	H	10938
Juncus tenuis	R	T	Z	635
Lupinus polyphyllus	R	T	N	1225
Lycium barbarum	R	T	N	352
Matricaria discoidea	R	J	N	1738
Mimulus guttatus	R, N	T	A	89
Oenothera biennis	R	T	N	306
Parthenocissus inserta	R, N	T	Z	145
Pinus strobus	R	T	A	772
Quercus rubra	L	T	Z	2544
Reynoutria sp.	R, N	T	A	353
Rhus hirta	R	T	N	27
Robinia pseudacacia	R	T	N	4372
Rudbeckia laciniata	R, N	T	N	140
Rumex alpinus	R	T	N	168
Solidago canadensis	R, N	T	A	1409
Solidago gigantea	R, N	T	A	782
Telekia speciosa	L	T	N	159
Veronica persica	R	J	N	2135

Testované ekologické charakteristiky jsou zejména způsob šíření na větší vzdálenosti a délka života (jednoleté/trvalky). Většina vybraných druhů invaduje mimo jiné nebo především ruderalní biotopy, tedy místa obvykle související s lidskou činností, často nějak narušená a opuštěná stanoviště (např. skládky, okraje cest, stavenišť, rumišť...). Část druhů se objevuje také v nivních biotopech a dva v lesích. Vzhledem k počtu druhů v jednotlivých skupinách bylo jediné možné porovnání nivních druhů vůči ostatním.

Prezenční data byla získána kombinací dat z České národní fytoecologické databáze (Chytrý et Rafajová 2003) a dat z mapování biotopů poskytnutých AOPK. Tato bodová data byla následně interpolována do rastru 1 x 1 km (každá buňka s alespoň jedním výskytem daného druhu byla označena jako prezenze). Tato výsledná rastrová vrstva převedená na číselnou matici prezenze (1) a absence (0) pak byla použita jako vstup pro sestavení a ověření jednotlivých modelů.

Většina testovaných metod vyžaduje pro správné fungování prezenční i absenční data. Absenční data jsou ale, zvláště ve velkém (celorepublikovém) měřítku velmi obtížně získatelná, pracuje se proto s tzv. pseudoabsencemi (viz např. Chefaoui et Lobo 2008). V tomto konkrétním případě byly za absence považovány všechny body mimo prezenční, resp. náhodný výběr ze všech těchto bodů.

Druhým velkým problémem je absence nezávislých dat pro testování modelů. Řešením je rozdělení datové sady na dvě části, přičemž na jedné probíhá výpočet modelu a na druhé „krosvalidace“.

Všechny druhy rodu *Reynoutria* byly zahrnuty do jedné kategorie, neboť může být obtížné mezi nimi rozlišovat a existuje předpoklad, že jejich ekologie a invazní chování je v podstatě jednotné. Modely pro *Rhus hirta* byly spočítány, ale nebyly zahrnuty do vyhodnocení kvůli malému počtu prezenčních bodů.

II.3. Software

Kvalitních softwarových nástrojů, které splňují dříve uvedené podmínky, není příliš mnoho. Konečný výběr proběhl z následujících pěti projektů:

Biomod (Thuiller 2003) je zřejmě nejaktuálnější a nejkomplexnější dostupný software pro modelování potenciálního šíření druhů. Jde o balík funkcí napsaných v jazyce R bez dalších závislostí, takže je plně multiplatformní (lze jej použít všude kde funguje R). Implementuje tyto statistické modely: GLM, GBM, GAM, CTA, ANN, SRE, MDA, MARS a RF (pro jednotlivé algoritmy využívá již existující balíky a funkce pro R). Umožňuje vytváření a testování modelů, projekci modelů do jiných časových a klimatických scénářů, sestavování křivek odpovědí druhů, testování interakcí a mnoho dalšího.

Software je doplněn rozsáhlým manuálem popisujícím jak praktické použití funkcí tak základní teorii a interpretaci všech implementovaných statistických metod (Thuiller et al. 2008).

GRASP¹ (Lehmann et al. 2003) je sofistikovaný balík včetně grafického rozhraní napsaný pro jazyk S-plus resp. R. Implementuje modelování prostorového rozšíření druhů statistickým modelem GAM od výběru proměnných, zpracování regresního modelu, predikce až po vyhodnocení a vytvoření mapy potenciálního rozšíření. Při předběžném testování na rozsáhlých datových souborech (velké území v podrobném rozlišení) narážíme na limity dané jak použitým skriptovacím jazykem (v testech využita varianta pro R) tak samotnou implementací.

OpenModeller² (Sutton et al. 2007) vznikl za účelem vytvoření jednotného, dostupného a uživatelsky přívětivého prostředí pro tvorbu a vyhodnocování ekologických modelů. Jde o multiplatformní software napsaný v C++. Jednotlivé statistické algoritmy jsou řešeny jako pluginy, takže je snadné rozšiřovat funkčnost programu o nové metody. Program umožňuje přípravu datových sad (dokonce získání dat výskytu daného druhu z několika dostupných on-line databází), sestavení a spuštění modelu, jeho vyhodnocení a zobrazení/exportu výsledných statistických a popisných dat a kartografické reprezentace vypočítané potenciální distribuce. Nabízí též rozhraní pro příkazovou řádku umožňující snadnou automatizaci a knihovny funkcí použitelné pro integraci v rámci jiných projektů.

Kromě klasických metod jako jsou ANN, AquaMaps, Bioclim a další „Environmental envelopes“ a jiné obsahuje jako jediný novou implementaci GARP (Elith et. al 2006).

Desktop GARP³ (Stockwell et Peters 1999) je klasická referenční implementace metody GARP pro modelování a vytváření map potenciálního rozšíření druhů. Jde o poměrně propracovaný desktopový software, při předběžném testování se však ukázal jako velmi nestabilní. To je možné přičíst použité testovací konfiguraci PC (Windows Vista 64bit). V současnosti ale existuje výrazně lepší implementace tohoto algoritmu v programu OpenModeller (Elith J. et al., 2006), která mimo jiné umožňuje paralelní zpracování na vícejádrových systémech, což může několikanásobně urychlit výpočet.

1) <http://www.unine.ch/cscf/grasp/grasp-r/index.html>

2) <http://openmodeller.sourceforge.net>

3) <http://www.nhm.ku.edu/desktopgarp/index.html>

Diva GIS⁴ je komplexní systém pro zpracování a analýzu prostorových dat, zejména distribuce organismů. Kromě mnoha jiných funkcí podporuje algoritmy BIOCLIM a DOMAIN pro modelování rozšíření druhů na základě ekologických nároků.

Všech pět softwarových produktů bylo porovnáno dle různých parametrů (tab. 3) a prakticky otestováno na jednoduchých modelech.

Tabulka 3: Srovnání vybraných softwarových produktů dle zvolených parametrů.

1) možnost spouštět program na více platformách, především Windows, Linux a Mac

2) Možnosti integrace s dalšími programy, automatizace úloh a pod.

3) Především možnost a složitost změn zdrojového kódu

4) Množství podporovaných algoritmů

5) Kvalita dokumentace

6) Aktuálnost vyjadřuje, zda je projekt stále ve vývoji

Software	Multiplatformnost ¹⁾	Možnosti integrace ²⁾	Flexibilita, možnosti úprav ³⁾	Množství metod ⁴⁾	Dokumentace ⁵⁾	Aktuálnost ⁶⁾
Biomod	+	+	+	+	+	+
GRASP	+	+	+	-	+	-
OpenModeller	+	+	+	+	-	+
Desktop GARP	-	-	-	-	+	-
Diva GIS	-	-	-	+	+	+

Z výše uvedeného vyplývají jednoznačně jako vhodné softwarové nástroje balík BIOMOD pro R a knihovny OpenModeller. Pomocí těchto dvou nástrojů byly vypočítány všechny modely použité v této studii.

II.4. Vstupní vrstvy

Volba prediktorů je obvykle limitována jejich dostupností pro dané území. Vzhledem k tomu, že vstupní datové vrstvy v tomto případě musí být dostupné pro celé (nebo alespoň velkou část) území České republiky, je jejich výběr skutečně omezen. Dostupností se totiž rozumí nejen to, že data někde existují, ale že je možné je za rozumných (finančních a technických) podmínek získat a využít. Následující odstavce popisují obecně dostupná data pro Českou republiku.

Základní vstupní data vždy poskytuje kvalitní **digitální model terénu**. Pro Českou republiku je bez problémů dostupný rastrový digitální model v rozlišení přibližně 50 x 60 m/pixel, to je pro velké oblasti naprosto dostačující. Z DTM mohou být odvozeny další vrstvy jako sklon, orientace a jejich varianty (teplotní index, oslunění a pod.).

4) <http://www.diva-gis.org/>

Data o krajinném pokryvu jsou dostupná ve formě aktualizované **databáze CORINE**, pro libovolné území je možné získat je on-line z příslušných internetových stránek projektu.

Podrobné **pedologické mapy** pro celé území České republiky zatím v digitální podobě dostupné nejsou, jako jistá náhrada mohou sloužit **geologické mapy** (1: 50 000 dostupné pro celé území, 1: 25 000 zatím aktualizovány a digitalizovány pouze částečně), které lze získat u České geologické služby.

Klimatická data lze snadno využít z projektu Worldclim (Hijmans et al. 2005) v rozlišení cca 1 km/pixel. Podrobná data má k dispozici Český hydrometeorologický ústav, není ale snadné je získat.

Zajímavá vstupní data by, zvláště pro některé druhy, mohla být získána např. z databáze **LPIS** či podkladů **ÚHÚL**. Bohužel tato data jsou vázána pouze na zemědělskou resp. lesní půdu a navíc velmi obtížně dostupná. Přínosné by mohly být i vrstvy vzniklé z **historického vojenského mapování**, bohužel tyto mapy zatím z velké části nejsou vektorizované a klasifikované.

Topografické podklady (např. města, komunikace, vodní toky a vodní plochy, lesy...) a od nich odvozené vrstvy jsou dobře dostupné v libovolném měřítku.

Pro zpracování modelů v rámci této studie bylo použito pouze několik základních vrstev (tab. 4). Nejdůležitější je kvalitní digitální model terénu. Nadmořská výška je základní (nepřímý) prediktor pro modelování potenciálního rozšíření, výrazně ovlivňuje a limituje všechny druhy rostlin (i ostatních organizmů). Z ní jsou pak odvozené vrstvy sklonu, expozice a oslunění. Dále byly zvoleny tři klimatické vrstvy z datové sady Worldclim (Hijmans et al., 2005), dvě celkově charakterizující (roční úhrn srážek a roční průměrná teplota) a nejnižší teplota nejchladnějšího měsíce (nízké teploty jsou pro rostliny často limitující).

Přidána byla vrstva vzdálenosti od vody (vodních ploch a vodních toků), neboť mnoho invazních druhů blízkost vody vyžaduje a pro mnohé jsou vodní toky důležitým vektorem šíření na velké vzdálenosti.

Všechny vrstvy byly převzorkovány na rozlišení rastru 1 x 1 km (použit byl software GRASS¹, metodou výpočtu průměrné hodnoty všech bodů rastru spadajících do nového čtverce.

1) GRASS Development Team, 2008. Geographic Resources Analysis Support System (GRASS) Software, Version 6.4.0 RC3. <http://grass.osgeo.org>

Tabulka 4: Vstupní datové vrstvy (pro modelování převzorkovány na rozlišení 1000 m/pixel)

Vrstva	Zdroj	Původní rozlišení	Popis
Nadmořská výška	http://grass.fsv.cvut.cz/gwiki/FreeGeodataCZ	cca 90x60 m/pixel	Podrobný model získaný metodou radarové interferometrie
Svažitost	Odvozeno z volně dostupných vrstev (http://grass.fsv.cvut.cz/gwiki/FreeGeodataCZ)		Odvozeno z nadmořské výšky softwarem GRASS
Expozice	Odvozeno z volně dostupných vrstev (http://grass.fsv.cvut.cz/gwiki/FreeGeodataCZ)		cos (orientace-225°)
Osluněnost	Odvozeno z volně dostupných vrstev (http://grass.fsv.cvut.cz/gwiki/FreeGeodataCZ)		cos (orientace-225°)*tg (svážitost)
Úhrn ročních srážek	http://www.worldclim.org	cca 1000 m/pixel	Klimatická data interpolovaná ze sítě pozemních meteorologických stanic
Nejnižší teplota nejhladnějšího měsíce	http://www.worldclim.org	cca 1000 m/pixel	Klimatická data interpolovaná ze sítě pozemních meteorologických stanic
Průměrná roční teplota	http://www.worldclim.org	cca 1000 m/pixel	Klimatická data interpolovaná ze sítě pozemních meteorologických stanic
Vzdálenost od vody	Odvozeno z volně dostupných vrstev (http://grass.fsv.cvut.cz/gwiki/FreeGeodataCZ)		

II.5. Vyhodnocení modelů

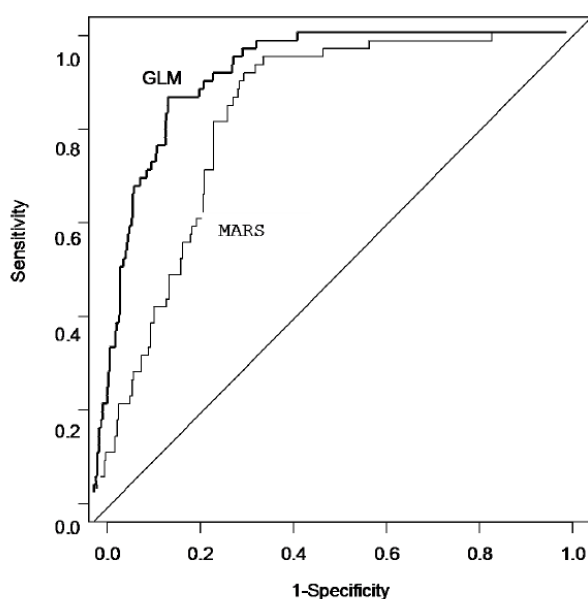
Vytvořené modely byly vyhodnoceny standardním způsobem pomocí hodnot AUC (area under curvature). Výstupem modelu je pravděpodobnost v rozmezí 0 – 1 (0 až 100 %), s jakou se v každém bodě může na základě vstupních dat daný druh vyskytovat. Chceme-li získat vrstvu vyjadřující potenciální presenci/absenci druhu, je třeba stanovit nějakou hranici

pravděpodobnosti, např. hodnoty menší než 50 % znamenají absenci, větší než 50 % prezenci. Toto stanovení může být velmi subjektivní a přitom výrazně ovlivňuje výslednou kvalitu predikce (počet správně a chybně určených prezencí a absencí). AUC představuje objektivní kritérium na stanovené hranici nezávislé. Jde o plochu pod křivkou ROC (Receiver operating curve), konstruovanou na základě množství správných a chybných prezencí při všech možných stanovených hranicích pravděpodobnosti mezi prezencí/absencí (obr. 1).

Pro model s nulovou vypovídací hodnotou je pravděpodobnost správně predikované absence stejná jako chybně predikované, AUC má hodnotu 0,5 (ROC tvoří diagonálu). Naopak model se 100% rozlišovací schopností bude mít AUC = 1 (ROC tvoří pravoúhle zalomenou křivku). Čím ostřeji a blíže levému hornímu rohu grafu, tím přesnější model s větší vypovídací hodnotou (tab. 5).

Tabulka 5: Klasifikace modelů podle hodnoty AUC

AUC	Přesnost
0,9 – 1	Výborná
0,8 – 0,9	Dobrá
0,7 – 0,8	Přiměřená
0,6 – 0,7	Špatná
0,5 – 0,6	Žádná



Obrázek 1: Příklad dvou křivek ROC umožňujících porovnání modelů (v tomto případě vychází GLM jako lepší než MARS) a výpočet AUC (Thuiller et al., 2008)

Výsledky byly zpracovány statistickým softwarem R¹ metodami klasické faktoriální analýzy variance, nested ANOVA a post-hoc metodou TukeyHSD pro zjištění rozdílů mezi jednotlivými skupinami.

1) R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

III. Výsledky

Na základě uvedených datových vrstev a prezenčních informací o druzích bylo zpracováno a vyhodnoceno celkem 272 modelů potenciálního rozšíření vybraných invazních druhů na území České republiky (tab. 6).

Tabulka 6: Hodnoty AUC pro jednotlivé druhy a modely jejich potenciálního rozšíření v rámci ČR. Šedivé hodnoty označují modely s AUC 0,5 – 0,6 (nulová predikční schopnost), **tučně** jsou označené modely s AUC > 0,9 (výborná predikční schopnost).

Druh	ANN	CTA	GAM	GBM	GLM	MARS	MDA	GARP
<i>Acer negundo</i>	0,678	0,708	0,729	0,767	0,727	0,734	0,695	0,692
<i>Ailanthus altissima</i>	0,916	0,825	0,945	0,933	0,945	0,896	0,852	0,904
<i>Amaranthus retroflexus</i>	0,873	0,799	0,877	0,889	0,875	0,862	0,711	0,844
<i>Amorpha fruticosa</i>	0,895	0,861	0,912	0,882	0,898	0,870	0,845	0,957
<i>Bidens frondosa</i>	0,788	0,694	0,798	0,804	0,796	0,793	0,609	0,500
<i>Bunias orientalis</i>	0,794	0,795	0,834	0,910	0,823	0,797	0,588	0,861
<i>Conyza canadensis</i>	0,822	0,812	0,828	0,844	0,826	0,823	0,707	0,795
<i>Digitalis purpurea</i>	0,757	0,692	0,865	0,881	0,865	0,852	0,772	0,789
<i>Echinocystis lobata</i>	0,888	0,580	0,890	0,884	0,871	0,796	0,662	0,887
<i>Epilobium ciliatum</i>	0,635	0,577	0,660	0,721	0,657	0,670	0,601	0,612
<i>Galinsoga parviflora</i>	0,772	0,700	0,798	0,809	0,796	0,783	0,640	0,794
<i>Galinsoga quadriradiata</i>	0,756	0,669	0,757	0,812	0,755	0,718	0,533	0,723
<i>Geranium pyrenaicum</i>	0,500	0,644	0,882	0,867	0,883	0,804	0,699	0,842
<i>Helianthus tuberosus</i>	0,854	0,778	0,878	0,862	0,876	0,861	0,753	0,850
<i>Heracleum mantegazzianum</i>	0,756	0,792	0,822	0,870	0,818	0,820	0,797	0,500
<i>Impatiens glandulifera</i>	0,754	0,727	0,778	0,818	0,778	0,743	0,722	0,500
<i>Impatiens parviflora</i>	0,699	0,731	0,740	0,770	0,739	0,743	0,706	0,500
<i>Juncus tenuis</i>	0,648	0,594	0,689	0,743	0,683	0,654	0,536	0,747
<i>Lupinus polyphyllus</i>	0,705	0,678	0,766	0,804	0,767	0,742	0,704	0,770
<i>Lycium barbarum</i>	0,915	0,756	0,926	0,922	0,923	0,891	0,722	0,882
<i>Matricaria discoidea</i>	0,622	0,600	0,623	0,702	0,621	0,614	0,511	0,500
<i>Mimulus guttatus</i>	0,720	0,615	0,848	0,901	0,847	0,728	0,584	0,717
<i>Oenothera biennis</i>	0,836	0,639	0,845	0,864	0,837	0,757	0,639	0,799
<i>Parthenocissus inserta</i>	0,859	0,653	0,887	0,871	0,900	0,892	0,755	0,870
<i>Pinus strobus</i>	0,761	0,699	0,755	0,842	0,752	0,733	0,721	0,500
<i>Quercus rubra</i>	0,700	0,721	0,735	0,806	0,732	0,743	0,720	0,500
<i>Reynoutria</i>	0,692	0,589	0,757	0,848	0,757	0,646	0,558	0,731
<i>Robinia pseudacacia</i>	0,810	0,806	0,823	0,837	0,822	0,820	0,806	0,749
<i>Rudbeckia laciniata</i>	0,775	0,545	0,833	0,785	0,829	0,727	0,555	0,500
<i>Rumex alpinus</i>	0,500	0,856	0,973	0,960	0,974	0,947	0,815	0,940
<i>Solidago canadensis</i>	0,762	0,733	0,772	0,813	0,773	0,756	0,732	0,731
<i>Solidago gigantea</i>	0,762	0,634	0,824	0,829	0,823	0,792	0,634	0,820
<i>Telekia speciosa</i>	0,500	0,601	0,810	0,809	0,799	0,500	0,596	0,741
<i>Veronica persica</i>	0,706	0,678	0,700	0,737	0,700	0,688	0,657	0,500

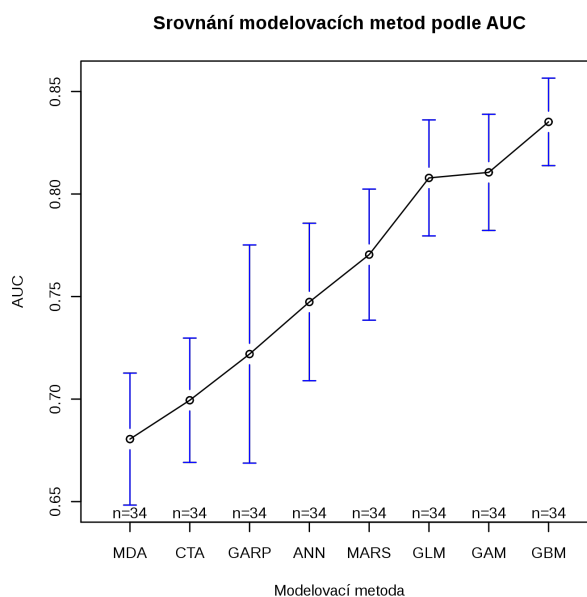
Průměrná hodnota AUC je 0,76. Téměř 74 % modelů dosahuje dobré a vyšší přesnosti (AUC > 0,7). Hodnoty AUC 0,5 se objevují u modelů, které nemají žádnou vypovídací hodnotu – s danými parametry a vstupními daty se nepovedlo nalézt žádný použitelný vztah mezi proměnnými prostředí a výskytem modelovaného druhu.

Mezi jednotlivými modelovacími technikami však existují poměrně značné rozdíly ($F_{(7,264)} = 11,245$; $p = 1,881e-12$; obr. 2 a tab. 7). MDA a CTA dosahují za testovaných podmínek konstantně nejhorších výsledků, středně kvalitních výsledků dosahují GARP, ANN a MARS. Hodnoty AUC pro GARP a ANN jsou velmi variabilní ($sd = 0,1524$ resp. $0,11$) zejména proto, že u těchto dvou metod nejčastěji došlo k selhání modelu ($AUC = 0,5$). Pokud bychom pro GARP počítali pouze modely s AUC vyšším než $0,5$, zařadil by se před MARS do skupiny s generalizovanými modely. MARS dosahuje průměrných ale stabilních

výsledků. Nejstabilnějších a nejlepších celkových výsledků dosahují tradiční metody GLM a GAM spolu s GBM, žádná z těchto metod ale neposkytuje nejvyšší hodnoty AUC ve všech jednotlivých případech. Stále proto platí, že lepší je modely kombinovat.

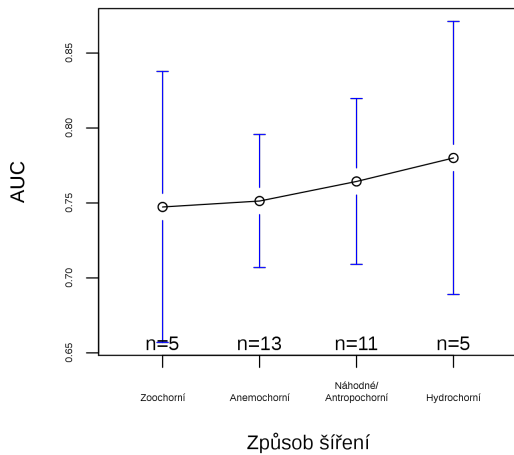
Srovnání úspěšnosti modelů podle jednotlivých ekologických skupin nepřineslo žádné průkazné výsledky. Ač z nalezených rozdílů můžeme usuzovat na určité potenciální trendy, nezdá se, že by zvolené biotopy, délka života a způsob šíření výrazným způsobem ovlivňovaly možnosti modelování té které skupiny, spíše jsou podstatné rozdíly na úrovni druhu dané jeho specifickou ekologií, historií a s tím souvisejícím vhodným nebo nevhodným výběrem vstupních vrstev a kvalitou dostupných prezenčních (či absenčních) dat.

Čtyři hlavní způsoby šíření invazních druhů na velké vzdálenosti (obr. 3) se mezi sebou v kvalitě modelů průkazně neliší ($F_{(3,30)} = 0,2269$; $p = 0,8769$). Je pravděpodobné, že hydrochorní rostliny se díky vazbě na vodní toky, které jsou prostorově dobře vymezené a zachycené v dostupných datech, dají modelovat o něco přesněji než ostatní. Rozdíly se v tomto případě týkají především některých metod (obr. 4), vzhledem k malému počtu testovaných rostlin však nejde činit žádné obecné závěry. Lze jen konstatovat, že v tomto konkrétním případě GARP při modelování hydrochorních a zoochorních druhů prakticky selhal, naopak MDA s ANN získaly nadprůměrné hodnoty AUC.



Obrázek 2: Srovnání jednotlivých metod podle kvality vytvořených modelů hodnocených pomocí AUC (body označují průměrné hodnoty s vyneseními 95% konfidenčními intervaly, spojnice bodů jsou určeny pro lepší vizualizaci rozdílů jednotlivých sousedních hodnot. Číslo n značí počet vzorků v dané kategorii.)

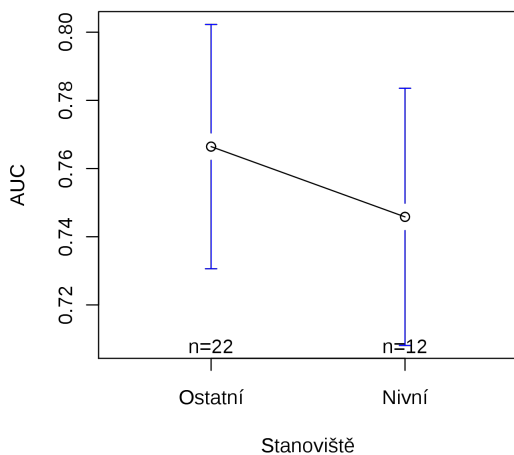
Srovnání různých způsobů šíření



Obrázek 3: Průměrné hodnoty AUC pro čtyři hlavní způsoby šíření invazních druhů rostlin na velké vzdálenosti (body označují průměrné hodnoty s vnesenými 95% konfidenčními intervaly, spojnice bodů jsou určeny pro lepší vizualizaci rozdílů jednotlivých sousedních hodnot. Číslo n značí počet vzorků v dané kategorii.)

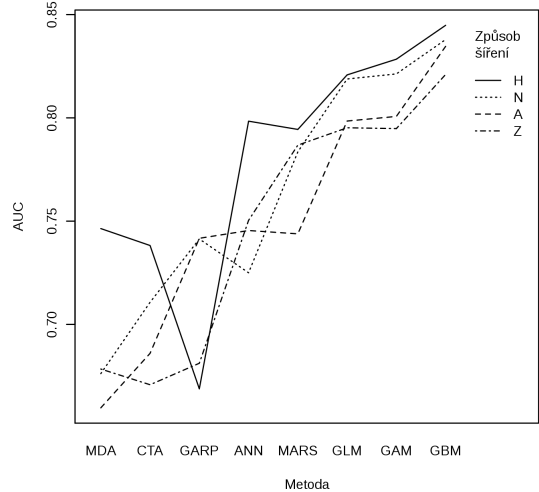
Srovnání podle stanovišť (obr. 5) bylo provedeno pouze pro nivní biotopy proti ostatním, protože ruderalní biotopy obývají téměř všechny vybrané druhy a není tedy s čím srovnávat. Nivní biotopy mají nepatrně horší výsledky než druhá skupina, rozdíl však není průkazný ($F_{(1,32)} = 0,6002$; $p = 0,4442$).

Srovnání invazních druhů nivních stanovišť a ostatních



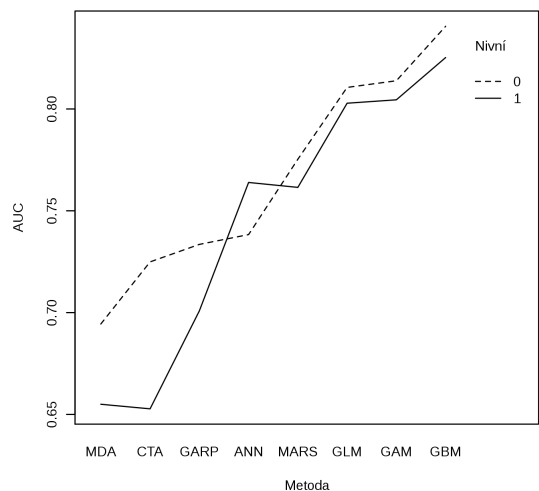
Obrázek 5: Průměrné hodnoty AUC pro invazní rostliny obývající mj. nivní biotopy a pro ostatní (především ruderalní) druhy (body označují průměrné hodnoty s vnesenými 95% konfidenčními intervaly, spojnice bodů jsou určeny pro lepší vizualizaci rozdílů jednotlivých sousedních hodnot. Číslo n značí počet vzorků v dané kategorii.)

Metody podle AUC a způsobu šíření



Obrázek 4: Porovnání průměrných hodnot AUC pro jednotlivé modely a hlavní způsoby šíření invazních druhů rostlin na velké vzdálenosti (vyjádření spojnici bodů je určeno pro lepší vizualizaci rozdílů)

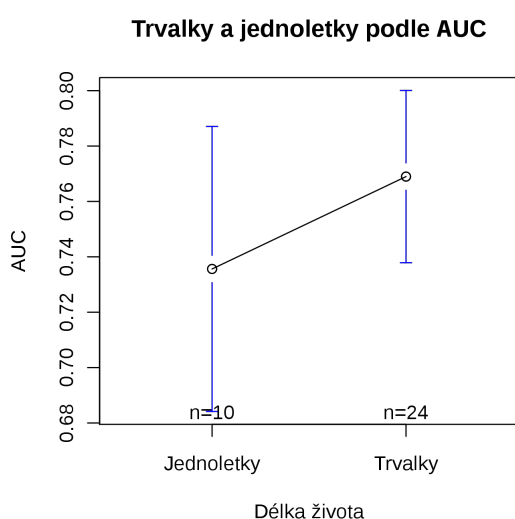
Metody podle AUC a stanoviště



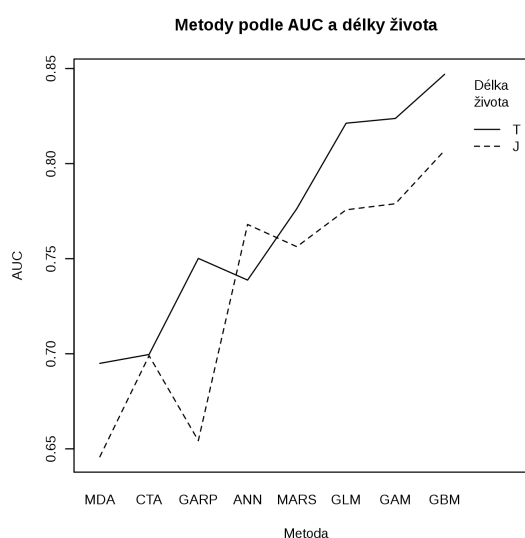
Obrázek 6: Srovnání průměrných hodnot AUC pro jednotlivé metody podle biotopu invazních druhů (1 – invazní druhy obývající mj. nivní biotopy, 0 – ostatní druhy, zejména ruderalní; vyjádření spojnici bodů je určeno pro lepší vizualizaci rozdílů)

Rozdíly mezi metodami (obr. 6) jsou patrné zejména u MDA a CTA, kde mají ruderální a ostatní druhy viditelně vyšší skóre AUC než nivní, u ostatních metod je rozdíl velmi malý a nepodstatný (ANN dokonce vykazuje mírně vyšší skóre pro nivní druhy).

Rozdíly mezi jednoletými a trvalými druhy (obr. 7) jsou nejvýraznější, ale ani v tomto případě zdaleka nejsou průkazné ($F_{(1,32)} = 1,4669$; $p = 0,2347$). Jednoleté druhy se zdají být variabilnější a hůře predikovatelné než trvalky. Vysvětlením může být větší nestálost a proměnlivost jednoletých druhů a jejich stanovišť a s tím související vyšší prostorová variabilita a zároveň obtížnější zachytitelnost v terénu a tudíž méně úplná a kvalitní prezenční data.



Obrázek 7: Průměrné hodnoty AUC pro jednoleté a trvalé invazní druhy (body označují průměrné hodnoty s vyneseními 95% konfidenčními intervaly, spojnice bodů jsou určeny pro lepší vizualizaci rozdílů jednotlivých sousedních hodnot. Číslo n značí počet vzorků v dané kategorii.)



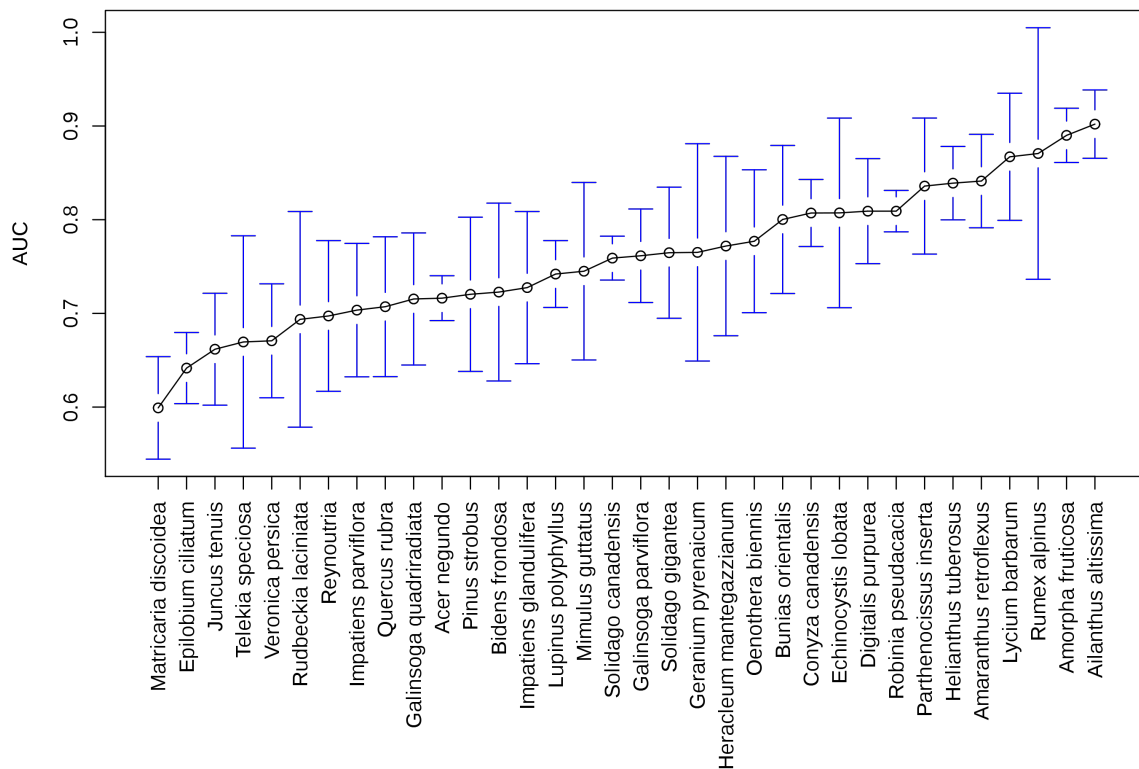
Obrázek 8: Srovnání průměrných hodnot AUC pro jednotlivé metody a délku života vybraných invazních druhů (T – trvalky, J – jednoletky; vyjádření spojnicemi bodů je určeno pro lepší vizualizaci rozdílů)

Rozdíly mezi metodami (obr. 8) jsou zde výraznější, ale v žádném z případů nejsou průkazné. Zatímco MDA a GARP v těchto konkrétních případech v modelování jednoletých druhů selhávají, CTA má pro obě skupiny výsledky prakticky shodné. ANN podobně jako v případě nivních biotopů a na rozdíl od ostatních metod vykazuje mírně vyšší AUC skóre pro jednoleté druhy. Trojice nejúspěšnějších modelů GLM, GAM a GBM produkují viditelně lepší predikce pro trvalé druhy.

Grafické srovnání druhů podle průměrných hodnot AUC (obr. 9) je doplněno pro lepší představu o variabilitě mezi samotnými druhy, přičemž přesnost modelování jednotlivých druhů není příliš důležitá a je ovlivněna množstvím faktorů zde nezachycených. Vysoké AUC neznamená, že je daný druh obecně lépe predikovatelný než ostatní, vybrané vstupní vrstvy

pouze vysvětlují prostorové rozšíření výskytů některých druhů lépe než jiných (některé druhy mohou být třeba více ovlivněny nadmořskou výškou či zrovna těmi klimatickými faktory, které byly v modelu použity).

Srovnání druhů podle AUC



Obrázek 9: Rozdíly průměrných hodnot AUC pro testované druhy invazních rostlin (body označují průměrné hodnoty s vyneseními 95% konfidenčními intervaly, spojnice bodů jsou určeny pro lepší vizualizaci rozdílů jednotlivých sousedních hodnot.)

IV. Diskuse

IV.1. Modelovací techniky

Ze srovnání modelovacích technik vychází jako vítěz trojice generalizovaných modelů GLM, GAM a GBM (BRT), přičemž nejlepších výsledků dosahuje posledně jmenovaná metoda, což se plně shoduje se závěry mnohem rozsáhlejší studie Elith et al. 2006. GLM/GAM se ukázaly jako univerzální a spolehlivé metody, což potvrzují i další studie (např. Meynard et Quinn 2007, Heikinnen et al. 2007). Díky jejich dostupnosti je proto dle mého názoru vhodné vždy je využívat jako jakési referenční modely, zejména pokud je cílem pouze kvalitní predikce prostorové distribuce daného (invazního) druhu bez specifických nároků na odhalení určitých skrytých vazeb či ověřování jiných teorií.

GLM a GAM poskytují za daných podmínek výsledky naprosto srovnatelné, v praxi by proto bylo výhodnější použít výpočetně méně náročný GLM (výpočet modelů metodou GLM byl zhruba o 50 % rychlejší než GAM).

Nejméně vyrovnané výsledky poskytly modelovací algoritmy ANN a zejména GARP. Síla GARP by proti ostatním modelům měla být ve schopnosti úspěšně modelovat potenciální rozšíření i z poměrně malého množství prezenčních dat. Naopak velké množství prezencí zřejmě při zvolených parametrech výrazně stěžuje schopnost GARP nalézt vhodná pravidla (výsledky prokazují negativní závislost hodnot AUC na počtu prezencí druhu vstupujících do modelu). Výsledky dosažené GARP tak docela neodpovídají srovnání v Elith et al. 2006, pokud bereme v úvahu všechny vytvořené modely. Pokud pomineme modely s hodnotou AUC rovných 0,5 – tedy případy, kdy žádný model nebyl vytvořen, výsledky budou v souladu s výše citovanou prací. Srovnání s jinými pracemi (Meynard et Quinn 2007, Pearson et al. 2007 a pod.) není smysluplné, v těchto pracích je použit původní algoritmus implementovaný v softwaru Desktop GARP (Stockwell et Peters 1999)

IV.2. Vstupní datové vrstvy

Výběr testovaných invazních druhů byl v tomto případě podřízen datům aktuálně dostupným v rámci projektu SPII2D1/37/07. Počty druhů v jednotlivých kategoriích proto nebyly vyvážené a v některých případech ani dostatečné. To mohlo zapříčinit slabší sílu statistických testů a jejich neprůkazné výsledky.

Vstupní data o výskytu druhů byla získána především z České národní fytoocenologické databáze (Chytrý et Rajfová 2003) a dat z aktuálně probíhajícího mapování biotopů poskytnutých Agenturou ochrany přírody a krajiny. Zejména data z fytoocenologické databáze jsou známá svou prostorovou neúplností, místy ostrůvkovitou distribucí a „přesnímkováním“ určitých botanicky zajímavých lokalit. Tento problém z části řeší právě kombinace s vrstvou mapování biotopů a pak určitá prostorová stratifikace při převzorkování dat do rastru 1 x 1 km, který je dostatečně hrubý na to, aby přesnímkování některých lokalit odstranil. I tak jsou data, zvláště v případě některých druhů, jistě neúplná a mohou být nedostatečná pro realistické modelování jejich rozšíření, což ale není cílem této studie. Pro realističtější modely zaměřené na konkrétní druhy je nutné využít další dostupné zdroje dat. Tam kde data dostupná nejsou je nutné přistoupit k terénnímu průzkumu.

Zvolené rozlišení modelů (rastr 1 x 1 km) vychází především z aktuálně dostupných datových vrstev a výpočetní náročnosti (limitující byl v tomto případě především výkon procesoru, při výpočtech ve vyšším rozlišení = s větším množstvím bodů v rastru je ale limitem i velikost operační paměti). Pro představu – čistý výpočetní čas uvedených modelů byl přibližně 72 hodin na procesoru s frekvencí 2,6 GHz, paměťové nároky se pohybovaly kolem 1 GB RAM. V dnešní době však není takový problém získat potřebné vstupní vrstvy (zejména digitální model terénu) v rozlišení umožňujícím vytvořit modely v rastru řekněme 100 x 100 m. To by mohlo mít význam zejména pro přesnější predikce v rámci menších území (výpočetní náročnost ale vzroste zhruba stonásobně).

IV.3. Ekologické skupiny

Především výsledky testování vlivu metod dálkového šíření invazních druhů na jejich predikovatelnost mohou být výrazněji ovlivněny výběrem datových vrstev vstupujících do modelů. Jednotlivé způsoby transportu se dle získaných výsledků liší jen minimálně, přičemž hydrochorní druhy získaly průměrně o něco vyšší AUC než ostatní. Příčinou může být uvedená jasná prostorová vymezenost vodních toků, ale také jednoduše fakt, že do modelů vstupovala vrstva vzdálenosti od vody. Pro lepší výsledky v této oblasti by bylo vhodné do modelů zařadit také vzdálenost od komunikací či od lidských sídlišť různé velikosti, případně překlasifikovat druhy do kategorií odpovídajících více důsledkům než jednotlivým vektorům šíření (anemochorní druhy by se z hlediska možnosti predikce mohly chovat např. jako náhodně se šířící).

V. Závěr

Na 34 druzích rostlin invazních v České republice bylo porovnáno 8 vybraných algoritmů pro modelování potenciálního rozšíření druhů. Nejpřesnějších výsledků dle de-facto standardního kritéria AUC dosahuje algoritmus GBM (BRT), což souhlasí s výsledky studie Elith et al., 2006. Mírně horších ale rovněž stabilních a dobrých výsledků dosahují GLM/GAM, přičemž mezi nimi v kvalitě modelů není podstatný rozdíl.

Slibné algoritmy ANN, MARS a GARP dosahují středně kvalitních výsledků, v případě GARP je to způsobeno především selháním velkého množství modelů. Konstantně špatných výsledků dosahují algoritmy MDA a CTA.

Pro modelování invazních druhů rostlin na území ČR jsou proto nejvhodnější GBM a GAM/GLM. Za určitých okolností by k nim mohl patřit i GARP, toto však vyžaduje další přezkoumání.

Jednoznačně nejvhodnější softwarové implementace těchto i dalších algoritmů jsou balíky v softwaru R a knihovny OpenModeller. Jejich využití je velmi flexibilní a umožňují integraci prakticky s čímkoliv.

Pro českou republiku existuje dostatečné množství dostupných dat umožňujících modelování potenciální distribuce druhů pro celé území státu v rozumném měřítku 1 x 1 km. Základem jsou digitální model terénu a klimatická data, další vrstvy související např. s vektory šíření nebo specifickými ekologickými nároky, jsou vesměs také dostupné.

Ekologické vlastnosti (způsob šíření, délka života a biotop) mají na predikovatelnost druhu žádný nebo jen malý vliv, který se nepodařilo prokázat. Nejvýraznější rozdíly byly zachyceny mezi skupinami jednoletých a trvalých rostlin.

VI. Literatura

- Anderson, R. P., Lew, D., Peterson, A. T., 2003.** Evaluating predictive models of species distributions: criteria for selecting optimal models. *Ecological Modelling* 162, 211-232
- Anderson, R. P., Peterson, A. T., Egbert S. L., 2006.** Vegetation-index models predict areas vulnerable to purple loosestrife (*Lythrum salicaria*) invasion in Kansas. *The Southwestern Naturalist* 51, 471-480
- Araújo, M. B., Guisan, A., 2006.** Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33, 1677-1688
- Araújo, M. B., Luoto, M., 2007.** The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16, 743-753
- Austin, M., 2007.** Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200, 1-19
- Austin, M. P., Belbin, L., Mayers, J. A., Doherty, M. D., Luoto, M., 2006.** Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological modeling* 199, 197-216
- Beerling, D. J., Huntley, B., Bailey, J. P., 1995.** Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *Journal of Vegetation Science* 6, 269-282
- Breiman, L., 2002.** Manual On Setting Up, Using, And Understanding Radnom Forests V3.1
- Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T., Guisan, A., 2007.** Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10, 701-709
- Busby, J. R., 1991.** BIOCLIM – a bioclimate analysis and prediction system. In: Margules, C. R., Austin, M. P. (Eds.), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Canberra, Australia, pp. 64-68
- Carpenter, G., Gillison, A. N., Winter, J., 1993.** DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* 2, 667-680
- Costa, G. C., Wolfe, C., Shepard, D. B., Caldwell, J. P., Vitt, L. J., 2008.** Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *Journal of Biogeography* 35, 637-646
- Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., Wilson, R., 2007.** Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30, 609-628

- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., Zimmermann, N. E., 2006.** Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151
- Endress, B. A., Naylor, B. J., Parks, C. G., Radosevich, S. R., 2007.** Landscape Factors Influencing the Abundance and Dominance of the Invasive Plant *Potentilla recta*. *Rangeland Ecol Manage* 60, 218-224
- Ennis, M., Hinton, G., Naylor, D., Revow, M., Tibshirani, R., 1988.** A comparison of statistical learning methods on the GUSTO database. *Stat. Med.* 17, 2501-2508
- Fiaboe, K. K. M., Fonseca, R. L., de Moraes, G. J., Ogol, C. K. P. O., Knapp, M., 2006.** Identification of priority areas in South America for exploration of natural enemies for classical biological control of *Tetranychus evansi* (Acari: Tetranychidae) in Africa. *Biological Control* 38, 373-379
- Foxcroft, L. C., Rouget, M., Richardson, D. M., 2007.** Risk Assessment of Riparian Plant Invasions into Protected Areas. *Conservation Biology* 21, 412-421
- Friedman, J., 1991.** Multivariate adaptive regression splines. *Annals of Statistics* 19, 1-141
- Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. MC. C., Aspinall, R., Hastie, T., 2006.** Making better biogeographical predictions of species distributions. *Journal of Applied Ecology* 43, 386-392
- Guisan, A., T. C. Edwards, Jr., Hastie, T., 2002.** Generalized linear and generalized additive models in studies of species distribution: Setting the scene. *Ecological Modelling* 157, 89-100
- Guisan, A., Thuiller, W., 2005.** Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8, 993-1009
- Hastie, T., 1991.** Generalized additive models. In: Chambers, J. M., Hastie, T. J. (Eds.), *Statistical Models*. S. Wadsworth and Brooks/Cole Advanced Books and Software, California, pp. 249-348.
- Heikkinen, R. K., Luoto, M., Kuussaari, M., Toivonen, R., 2007.** Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique. *Landscape and Urban Planning* 79, 347-357
- Herborg, L., Jerde, C. L., Lodge, D. M., Ruiz, G. M., MacIsaac, H. J., 2007.** Predicting invasion risk using measures of introduction effort and environmental niche models. *Ecological Applications* 17, 663-674
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., Jarvis, A., 2005.** Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965-1978
- Chefaoui, R. M., Lobo, J. M., 2008.** Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological modelling* 210, 478-486

- Chen, P., Wiley, E. O., Mcnyset, K. M., 2007.** Ecological niche modeling as a predictive tool: silver and bighead carps in North America. *Biological Invasions* 9, 43-51
- Chytrý M., Rafajová, M., 2003.** Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. *Preslia* 75, 1-15
- Irfan-Ullah, M., Amarnath, G., Murthy, M. S. R., Peterson, A. T., 2007.** Mapping the geographic distribution of *Aglaia bourdillonii* Gamble (Meliaceae), an endemic and threatend plant, using ecological niche modeling. *Biodiversity Conservation* 16, 1917-1925
- Kadmon, R., Farber, O., Danin, A., 2003.** A systematic analysis of factors affecting the performance of climatic envelope models. *Ecol. Appl.* 13, 853-867.
- Leathwick, J. R., Elith, J., Hastie, T., 2006.** Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological modelling* 199, 188-196
- Lehmann, A., Overton, J. M., Leathwick, J. R., 2003.** GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* 160, 165-183
- Lek, S., Guegan, J., 1999.** Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* 120, 65-73
- Lobo, J. M., Jiménez-Valverde, A., Real, R., 2008.** AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17, 145-151
- Loiselle, B. A., Jørgensen, M., Consiglio, T., Jiménez, I., Blake, J. G., Lohmann, L. G., Montiel, O. M., 2008.** Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* 35, 105-116
- Loo, S. E., Mac Nally, R., Lake, P. S., 2007.** Forecasting New Zealand mudsnail invasion range: Model comparisons using native and invaded ranges. *Ecological applications* 17, 181-189
- Loots, C., Koubbi, P., 2007.** Habitat modelling of *Electrona antarctica* (Myctophidae, Pisces) in Kerguelen by generalized additive models and geographic information systems. *Polar Biol* 30, 951-959
- Lundholm, J. T., Larson, D. W., 2004.** Dominance as an overlooked measure of invader success. *Biological Invasions* 6, 505-510
- Luoto, M., Hjort, J., 2005.** Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology* 67, 299-315
- Manel, D., Dias, J. M., Buckton, S. T., Ormerod, S. J., 1999.** Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology* 36, 734-747
- McCullagh, P., Nelder, J. A., 1989.** *Generalized Linear Models Monographs* odn *Statistics and Applied Probability*. Chapman and Hall, London, 261 pp.

- Meynard, C. N., Quinn, J. F., 2007.** Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography* 34, 1455-1469
- Mitikka V., Heikkinen, R. K., Luoto, M., Araújo, M. B., Saarinen, K., Pöyry, J., Fronzek, S., 2008.** Predicting range expansion of the map butterfly in Northern Europe using bioclimatic models. *Biodivers Conserv* 17, 623-641
- Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E., Edwards, T. C. Jr., 2006.** Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models and tree-based methods. *Ecological modelling* 199, 176-187
- Moisen, G. G., Frescino, T. S., 2002.** Comparing five modeling techniques for predicting forest characteristics. *Ecol. Model.* 157, 209-225.
- Muñoz, J., Felicísimo, Á. M., 2004.** Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science* 15, 285-292
- Nehrbass, N., Winkler, E., Müllerová, J., Pergl, J., Pyšek, P., Perglová, I., 2007.** A simulation model of plant invasion: long distance dispersal determines the pattern of spread. *Biol Invasions* 9, 383-395
- Pande, A., Williams, C. L., Lant, C. L., Gibson, D. J., 2007.** Using map algebra to determine the mesoscale distribution of invasive plants: the case of *Celastrus orbiculatus* in Southern Illinois, USA. *Biol Invasions* 9, 419-431
- Pearson, R. G., Raxworthy, C. J., Nakamura, M., Peterson, A. T., 2007.** Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34, 102-117
- Peterson, A. T., 2003.** Predicting the geography of species invasions via ecological niche modeling. *The quarterly review of biology* 78, No. 4, 419-433
- Phillips, S. J., Anderson, R. P., Schapire, R. E., 2006.** Maximum entropy modelling of species geographic distributions. *Ecological Modelling* 190, 231-259
- Pyšek, P., Prach, K., 1994.** How important are rivers for supporting plant invasions? Waal, L. C., Child, L. E., Wade, P. M., Brock, J. H., *Ecology and management of invasive riverside plants*, 19-26, John Wiley & Sons, New York,
- Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., Fuisan, A., 2006.** Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33, 1689-1703
- Richardson, D. M., Thuiller, W., 2007.** Home away from home – objective mapping of high-risk source areas for plant introductions. *Diversity and Distributions* 13, 299-312
- Robertson, M. P., Peter, C. I., Villet, M. H., Ripley, B. S., 2003.** Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling* 164, 153-167
- Sánchez-Flores, E., 2007.** GARP modeling of natural and human factors affecting the potential distribution of the invasives *Shismus arabicus* and *Brassica tournefortii* in 'El Pinacate y Gran Desierto de Altar' Biosphere Reserve. *Ecological modelling* 204, 457-474

- Segurado, P., Araújo, M. B., Kunin, W. E., 2006.** Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology* 43, 433-444
- Sérgio, C., Figueira, R., Draper, D., Menezes, R., Sousa, A. J., 2007.** Modelling bryophyte distribution based on ecological information for extent of occurrence assessment. *Biological conservation* 135, 341-351
- Slavík, B., et al., 1988-2004.** Květena České republiky. Academia, Praha
- Stockwell, D. R. B., Peters, D., 1999.** The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. Geograph. Inform. Sci.* 13, 143-158.
- Stockwell, D. R. B., Beach, J. H., Stewart, A., Vorntsov, G., Vieglais, D., Pereira, R. S., 2006.** The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecological modelling* 195, 139-145
- Sutton, T., Giovanni, R., Ferreira, S., 2007.** Introducing openModeller. *OSGeo Journal* 1, 1-6
- Sweeney, A. W., Beebe, N. W., Cooper, R. D., 2007.** Analysis of environmental factors influencing the range anopheline mosquitoes in northern Australia using a genetic algorithm and data mining methods. *Ecological modelling* 203, 375-386
- Thuiller, W., 2003.** BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9, 1353-1362
- Thuiller, W., Lafourcade, M., Araujo, M., 2008.** ModOperating Manual for BIOMOD, Laboratoire d'Écologie Alpine, Université Joseph Fourier, Grenoble 1
- Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R., 2007.** A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions* 13, 397-405
- Welk, E., 2004.** Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling* 179, 551-567
- Willems, W., Goethals, P., Van den Eynde, D., Van Hoey, G., Van Lancker, V., Verfaillie, E., Vincx, M., Degraer, S., 2008.** Where is the worm? Predictive modelling of the habitat preferences of the tube-building polychaete *Lanice conchilega*. *Ecological modelling* 212, 74-79
- Yee, T. W., Mitchell, N. D., 1991.** Generalized additive models in plant ecology. *J. Veg. Sci.* 2, 587-602
- Yen, P., Huettmann, F., Cooke, F., 2004.** Modelling abundance and distribution of Marbled Murrelets (*Brachyramphus marmoratus*) using GIS, marine data and advanced multivariate statistics. *Ecol. Model.* 171, 395-413.
- Zhu, L., Sun, O. J., Sang, W., Li, Z., Ma, K., 2007.** Predicting the spatial distribution of an invasive plant species (*Eupatorium adenophorum*) in China. *Landscape Ecology* 22, 1143-1154

VII. Přílohy

Tabulka 7: Rozdíly v kvalitě modelů (AUC) mezi jednotlivými metodami

(* = průkazný výsledek)

Metody	TukeyHSD (p)
ANN – CTA	0,47201
ANN – GARP	0,96266
ANN – MDA	0,09563
CTA – MDA	0,99314
GAM – ANN	0,13929
GAM – CTA *	0,00012
GAM – GARP *	0,00566
GAM – GLM	1,00000
GAM – MARS	0,69420
GAM – MDA *	0,00000
GARP – CTA	0,98083
GARP – MDA	0,65679
GBM – ANN *	0,00635
GBM – CTA *	0,00000
GBM – GAM	0,96841
GBM – GARP *	0,00008
GBM – GLM	0,94473
GBM – MARS	0,11941
GBM – MDA *	0,00000
GLM – ANN	0,18043
GLM – CTA *	0,00020
GLM – GARP *	0,00846
GLM – MARS	0,76464
GLM – MDA *	0,00000
MARS – ANN	0,97795
MARS – CTA *	0,06011
MARS – GARP	0,45628
MARS – MDA *	0,00461

Přehled dat o výskytu jednotlivých invazních druhů (prezenční data pro modelování):

