

Oponentský posudek na magisterskou práci

Název práce: **Modelování potenciálního šíření invazních druhů rostlin v ČR: Porovnání metod a jejich implementací, dostupnost dat a vliv ekologie druhu na přesnost predikce**

Autor: Pavel Brych

Oponent: František Zemek

Předložená práce má poskytnout přehled metod vhodných pro modelování invazních druhů rostlin v ČR, provést srovnání těchto metod pro konkrétní druhy, včetně hodnocení existence/dostupnosti dat pro tento účel.

Po **stránce obsahové**, zejména naplnění dílčích částí, se mně jeví práce značně nevyvážená.

Úvodní část, která zahrnuje rešerše řešené problematiky, od teorie modelování invazních druhů, přes přehled nejčastěji používaných metod včetně jejich základních charakteristik, výhod/nevýhod jejich dosavadního použití, vazbu jednotlivých algoritmů na charakter dat, je podána velmi přehlednou a srozumitelnou formou, v dostatečném rozsahu.

Cíle práce jsou již méně srozumitelné, protože není jasné, zda cílem má být skutečně „sestavení jakéhosi „obecného frameworku“, umožňujícího uživatelské vytváření map potenciálního rozšíření invazních druhů“ nebo srovnání existujících implementovaných algoritmů na zvoleném, docela úzkém výběru ekologických charakteristik, z obecně dostupných dat pro území ČR.

Kdo a na jaké odborné úrovni bude předpokládaný uživatel, jaký bude mít přístup k jakým datům, apod.?

Následné části práce spíše nasvědčují, že vlastní práce byla zúžena na srovnání existujících implementovaných algoritmů. Tato úloha může být pro kteréhokoli řešitele již samo velmi náročný úkol, protože předpokládá pochopení široké škály algoritmů, zvládnutí různorodého software a demonstrace výsledků, což řešitel dostatečně dokumentoval hlavně ve výsledkové části práce.

Úvodní tři části **Metodiky** dobře navazují na přehled metod z rešerše a seznamují se software, který je k dispozici pro řešení úlohy.

Za velmi kusou a nedostatečnou považuji část věnovanou **vstupním datům** do modelů. Přitom se určitě jedná pro výsledek takového úlohy minimálně o stejně důležitý krok, jako výběr modelů. Přihlížím k tomu, že rozsah řešení byl omezen prostředky projektu, jehož je diplomant spoluřešitelem. Toto omezení ale nebrání řešiteli, aby buď nedeklaroval jako jednu z otázek (č. 3 v cílech projektu) zpracování „dostupnosti potenciálních“ vstupních dat pro území celé ČR nebo aby skutečně provedl aktuální inventarizaci dat GIS pro řešení úlohy na různé informační/prostorové úrovni.

Co je zde rozuměno pojmem „potenciální“?

Jsou zde v podmiňovacím stavu uvedena data (vrstvy), která nejsou řešitelem dále využita (např. CORINE, pedologie, LPIS, atd.). Schází detailnější popis, proč byla dále zpracovaná data vybrána, jakou mají vazbu na šíření vybraných druhů.

Je to průnik základních ekologických charakteristik všech druhů, které byly zpracovány?

V Tab. 4 postrádám bližší definici o parametru „osluněnosti“, obvykle se rozumí suma energie dopadající ze Slunce na danou plochu, závisí i na poloze Slunce v průběhu roku a zeměpisné šířce. Podle autorem použitého vzorce je hodnota jeho oslunění stejná pro stejnou expozici a svažitost, nezávisle na zeměpisné šířce místa.

Jak byly připraveny vrstvy vzdálenost od vody? O jaký řád toku se jedná, když pracujete s pixelem 1x1 km?

Několikrát je v rešerši a metodách zmiňována nutnost přesného digitálního modelu reliéfu terénu, plně souhlasím. Nakonec je vše (i deriváty DMT) ale průměrováno na 1x1 km. Jako základní zdroj DTM je jednou uváděno rozlišení 50x60 m/pixel, podruhé (v Tab. 4) 90x60 m/pixel.

Stejně tak je zmiňována důležitost zohlednění autokorelace v modelech, u některých druhů určitě velice významný parametr pro modelování šíření druhu.

Bylo toto nějak v modelech využito? Využívají některé z modelů kontext okolí nebo je vše per pixel? Protože práci schází větší zdůvodněnost vazeb mezi sledovanými druhy a výběrem ekologických parametrů v metodické/datové části, jsou i výsledková část a diskuse zaměřeny na srovnání rozdílů mezi algoritmy pro jednotlivé druhy/funkční skupiny pojaty trochu mechanisticky, ale statisticky správně.

Formální stránka práce je na velmi dobré úrovni, práce je psána stručným, ale výstižným způsobem. Totéž platí i o prezentaci tabulek a grafů. Citovaná literatura je úplná, zde se vyskytuje jen několik překlepů, např. při výčtu více citací v textové části by měl být jiný oddělovač než „v“, podobně jako při dvou autorech jsou jejich jména v textu citace obvykle spojována „and“, „&“, nikoli „et“. Odkazy na aktuální webové zdroje jsou užitečné.

Celkové hodnocení práce

I přes výše zmíněné nedostatky, hlavně v metodické části práce, se domnívám, že diplomant prokázal v předloženém elaborátu schopnost zpracovat náročné, široce pojaté téma, orientovat se a prakticky uchopit celou škálu modelů. Využitelnost výsledků práce vidím spíše ve vytvořeném přehledu algoritmů a jejich implementaci než jako „kuchařku“, podle níž by mohl někdo „bezstarostně“ uchopit daný modul a „vytvářet mapy potenciálního rozšíření invazních druhů“.

Dle mého mínění práce splňuje požadavky kladené na magisterské práce předkládané na PŘF JU, proto ji doporučuji k obhajobě a hodnotím klasifikačním stupněm:

VELMI DOBRĚ

Datum: 24. května 2009

Podpis oponenta: František Zemek

Posudek na magisterskou práci P. Brycha (2009), Modelování potenciálního šíření invazních druhů rostlin v ČR: porovnání metod a jejich implementací, dostupnost dat a vliv ekologie druhu na přesnost predikce

Autor si jako cíle práce zvolil porovnat schopnost různých typů statistických modelů předpovídat výskyt invazních druhů na území České republiky, prozkoumat použitelnost software implementujícího jednotlivé modely pro potřeby rozsáhlejšího projektu, jehož součástí tato práce byla, a také zjistit, zda se výkon různých modelů liší podle ekologických vlastností invazních druhů.

Po přečtení této práce jsem nabyt dojmu, že autor uvedeným úkolům věnoval dostatek času a získal odpovídající výsledky. Na co ale zbylo času méně je dopilování prezentace těchto výsledků do uspořádané a dobře čtivé podoby. Organizace textu i jednotlivé formulace jsou na mnoha místech dosti odbyté a oponentovi nezbyvá než věřit, že je tomu tak "jenom" proto, že si autor neponechal na dokončení práce dostatek času, nikoliv že by lepšího výstupu nebyl schopen. Za příklad formulačních nedostatků může sloužit tento text z úvodu (1. odstavec na straně 3): "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" nebo věta ze závěru práce: "Ekologické vlastnosti (...) mají na predikovatelnost druhu žádný nebo jen malý vliv, který se nepodařilo prokázat".

Celková organizace textu také pokulhává na více místech, např. v literárním úvodu (strana 4 kolem řádky 17) autor v témže odstavci přeskakuje z tématu autokorelace na téma nereprezentativnosti dat a zase zpět. Oba problémy se sice často vyskytují u stejných datových zdrojů, ale jsou přeci jen odlišné (první nelze řešit beze sběru dalších dat, druhý lze do značné míry potlačit volbou vhodné statistické metody).

K vlastnímu obsahu práce mám několik dotazů a komentářů:

- (1) Po přečtení textu o prostorové autokorelaci na straně 4 mám pocit, že autor považuje za příčinu autokorelace nevhodné vzorkování. Domnívá se autor, že pokud by vzorkování bylo náhodné případně pravidelné (např. s použitím družicových dat), že by byl problém autokorelace odstraněn? Mohl by také definovat, co si pod pojmem "prostorová autokorelace v datech" představuje?
- (2) Mohl by autor vysvětlit, co je "mapová algebra" (strana 5, konec prvního odstavce sekce 1.4)? Možná, že je to termín zcela běžný v oboru geografických informačních systémů, mně ale nic neříká, i když mi význam slov "mapa" a "algebra" znám je.
- (3) Kde autor našel vysvětlení, že zkratka AUC znamená "area under curvature"? Já jsem si myslel, že třetím slovem má být "curve", tedy křivka – nikoliv zakřivení, a krátké hraní s Googlem mne utvrzuje v tom, že si to myslí i mnoho dalších autorů.
- (4) Na stranách 10 a 11 se opakuje formulace týkající se modelů "založených na vztahu průměru vysvětlované proměnné a lineární kombinaci vysvětlujících proměnných". Vypadalo by to jako nevinný překlep (vztah musí být k něčemu), ale proč by se pak objevil vícekrát ...
- (5) V tabulce 2 na straně 13 mne překvapila definice borovice vejmutovky (*Pinus strobus*) coby ruderálního druhu. Co autora vedlo k tomu, že nebyl druhu přiřazen typ biotopu "lesní", tak jako například u *Quercus rubra*?

- (6) Autor zmiňuje v metodické části práce vhodnost krosvalidace výsledků, ale nepodařilo se mi jednoznačně určit, zda jím provedené vyhodnocení kvality predikcí pomocí AUC charakteristiky bylo založeno na krosvalidaci nebo na pouhé aplikaci parametrizovaných modelů na výchozí data. Pokud to druhé, mohl by autor stručně diskutovat, jaký vliv mohlo mít toto rozhodnutí na závěry práce?
- (7) Obrázky 3 až 8 prezentují v grafické podobě rozdíly kvality statistických modelů mezi různě definovanými skupinami druhů (podle způsobu šíření, podle biotopů, či podle životní strategie). Ani u jedné z těchto klasifikací ale nemohla být zamítnuta nulová hypotéza, že mezi skupinami nejsou systematické rozdíly, a tedy že jde o pouhou náhodnou odlišnost. Z tohoto pohledu mi přijde rozsáhlá grafická prezentace docela nadbytečná a při případné tvorbě článku ji vřele doporučuji vynechat.
- (8) Tato práce je zaměřena výlučně na metody, které předpovídají výsledné (konečné) rozšíření invazních druhů, zatímco některé z druhů použitých k testování metod se ještě dále šíří. Použité metody také provádějí predikci nezávisle pro každý bod sítě, bez ohledu na kontext okolních bodů. Mohl by autor stručně okomentovat, jak se k těmto možným dvěma problémům staví jiní badatelé, případně jak by je bylo možné řešit?

Na závěr bych svůj dojem z práce rád shrnul tak, že práce má víceméně jen formální nedostatky, nicméně požadavek správných a čtivých formulací nelze zcela opomenout. Při kvalitní obhajobě proto navrhuji práci hodnotit stupněm velmi dobře.



České Budějovice, 17. května 2009

Doc. RNDr. Petr Šmilauer, Ph.D.

katedra biologie ekosystémů PFF JU