

Faktorová analýza výskytu vybraných botanických a zoologických taxonů

Zbyněk Janoška

Katedra geoinformatiky, Přírodovědecká fakulta, Univerzita Palackého v Olomouci,
Tř. Svobody 26
771 46, Olomouc, Česká republika
zbynek.janoska@centrum.cz

Abstrakt. Základním problémem krajinného mapování je otázka pořizování co nejkvalitnějších dat. Je zřejmé, že pomocí pouhých terénních pozorování není možné přesně zmapovat distribuci přírodních jevů. Předpovědní modely mohou velmi usnadnit hledání potenciálních oblastí výskytu těchto jevů. Zároveň však podstata přírodních fenoménů je komplexní a nemůže být jednoduše odhadnuta. Hlavním cílem bakalářské práce bylo najít závislost výskytu taxonů denních motýlů a taxonů vyšších cévnatých rostlin. Tato závislost byla modelována v prostředí R za pomoci metod shlukové analýzy a logistické regrese. Výsledky pak byly vizualizovány pomocí simulací v extenzi Geostatistical Analyst pro ArcGIS 9.3.

Klíčová slova: cévnaté rostliny, motýli, shluková analýza, logistická regrese

Abstract. Factors analysis of the incidence of selected botanical and zoological taxa. The basic problem of the landscape mapping is how to obtain the high-quality data. Prediction models can be very helpful for estimating areas of incidence of selected taxa. Also the structure of natural phenomenons is complex and can not be simply quantified without analysing the processes behind. Main aim of this bachelor thesis was to find dependence between the incidence of selected botanical taxa and zoological taxa from the family of lepidoptera. Statistical analysis in the R-Project environment was used to determine these linkages and then ArcGIS 9.3 (Geostatistical Analyst extension) to model the distribution of the phenomenons.

Keywords: tracheophyta, lepidoptera, cluster analysis, logistic regression

1 Úvod

Jak botanický, tak zoologický výzkum se potýkají se základním problémem, a to otázkou, jak získat co nejpřesnější a nejspolehlivější data. Je zřejmé, že není v lidských silách zmapovat všechny výskyty jednotlivých taxonů na celém území. Při mapování například stromů, velkých savců, nebo jiných „výrazných“ taxonomických skupin je možné částečně nahradit terénní pozorování za některou z distančních metod. Se změnou velikosti taxonu či „výraznosti“ jeho projevu či pobytové stopy v krajině je třeba počítat s tím, že získaná data reprezentují jen část populace daného taxonu v daném území.

Pro získání co nejpřesnějšího obrazu o výskytu sledovaného zoologického či botanického taxonu je proto potřebné hledat jiné souvislosti a nespoléhat se pouze na přímá terénní pozorování.

Při botanickém výzkumu se zdají být vhodným indikátorem výskytu druhů fyzicko-geografické podmínky, jako jsou například geologické, klimatické, hydrologické nebo půdní poměry, stejně dobře jako využití země.

Zoologický výzkum se může opírat o stejnou metodiku, jako vhodnější se ale jeví hledat souvislosti mezi výskytem botanických a zoologických taxonů, zejména u živočichů, kteří jsou býložraví a jejich pohyb v prostoru je omezen.

Jako příkladný studijní materiál se jeví vyšší cévnaté rostliny a denní motýli, kde můžeme předpokládat závislost výskytu motýlů na určitých skupinách či přímo taxonech rostlin.

2 Cíle práce

Cílem práce je analýza závislosti výskytu vybraných botanických a zoologických taxonů navzájem.

Hlavním cílem bude najít závislost výskytu taxonů denních motýlů na výskytu taxonů vyšších cévnatých rostlin.

Provedení vlastní analýzy vyžaduje vykonání dílčích přípravných činností, jako je sběr, standardizace a hodnocení dat. Samotná analýza bude mít za úkol zjistit možnou závislost a tu dále kvantifikovat.

Výskyt taxonů, u kterých bude závislost prokázána, bude dále modelován v prostředí extenze Geostatistical Analyst pro ArcGIS Desktop 9.3.

Výsledkem bakalářské práce bude statistické hodnocení dat, obsahující slovní, číselné, i grafické vyjádření, a pro vybrané taxony denních motýlů dále mapy areálu výskytu.

3 Postup práce

Vstupní data: Všechna získaná data pocházejí z databázového skladu CHKO Bílé Karpaty. Jedná se o tabulky botanických a lepidopterologických pozorování. Data byla pořizována v různých časových obdobích pod různými metodikami, proto bylo nutné nejprve data standardizovat. Většina dat byla pořizována ve formátu binárním, aby bylo možné provádět jednotlivé analýzy, všechna data byla proto na tento formát normalizována.

Zájmové území: Zájmovým územím je oblast CHKO Bílé Karpaty. Tato oblast je pokryta polygonovou vrstvou kvadrátů síťového mapování dle standardu KFME. Základní územní jednotkou, se kterou bylo pracováno je subkvadrát, což je plocha o ploše přibližně 9 km².

Základní analýzy: Data byla seskupena do kontingenčních tabulek. Jedním rozměrem byla vždy lokalita, druhým byl taxon a obsahem byl počet výskytů na daných lokalitách. Tabulky byly seskupeny zvlášť pro taxony botanických pozorování

a zvlášť pro lepidopterologická pozorování. V rámci obou tabulek byl proveden χ^2 -test k ověření nezávislosti proměnných. Analýza prokázala, že výskyt sledovaných jevů prostoru je nenáhodný a tudíž má cenu ho dále zkoumat.

Dalším krokem bylo provedení analýzy rozptylu (ANOVA) nad datasetem denních motýlů. Zkoumán byl vliv jednotlivých lokalit na průměrný výskyt motýlů. Výsledkem bylo zjištění, že jednotlivé lokality mají vliv na průměrný výskyt motýlů. Tyto analýzy byly prováděny v prostředí R.

Seskupení dat: Vstupní datasety obsahovaly velké množství mnohorozměrných dat, což bylo překážkou snadné analýzy a pro modelování bylo nutné počet rozměrů snížit. Nejprve byly z datasetu vyšších cévnatých rostlin vybrány ty taxony, které podléhají režimu ochrany. Tyto taxony jsou i v rámci terénního mapování sledovány pečlivěji a existuje zde předpoklad vyšší kvality pořízených dat. Dále byl pro každý jednotlivý taxon, botanický i zoologický, spočítán geografický střed. Pro botanická pozorování byl vypočítán prostý geografický střed, pro taxony motýlů byl spočítán střed vážený, kde vahou byl počet pozorování na jednotlivých lokalitách. Kvalita botanických pozorování výpočet váženého středu neumožňovala.

Nadále bylo přikročeno k vlastní shlukové analýze v programu QC.Expert. Bylo použito Ward-Wischartovy metody, kde mírou nepodobnosti byla metrika čtverce euklidovské vzdálenosti. Data byla seskupena do 44 shluků, z nich jen 28 vyhovovalo požadavku, aby shluk obsahoval nejméně jeden taxon botanický a nejméně jeden taxon zoologický.

Logistická regrese: Vzhledem k binárnímu formátu dat byla zvolena metoda logistické regrese jako optimální pro hledání modelu, vysvětlujícího závislost zkoumaných jevů. V prostředí R byl sestrojen model v rámci každého z 28 shluků pro každý taxon motýla.

Výsledný model jsi lze představit následovně^[1]:

$$Z = a_0 + a_1 x_{a_1} + a_2 x_{a_2} + \dots + a_p x_{a_p}, \quad (1)$$

kde odhadované koeficienty $a_0, a_1, a_2, \dots, a_p$ jsou jednotlivé parametry modelu a $x_{a_1}, x_{a_2}, \dots, x_{a_p}$ hodnoty sledovaného jevu, v tomto případě výskyt/nevýskyt botanického taxonu.

Jako relevantní byly vybrány ty parametry, u kterých byla vypočtená hladina významnosti vyšší než 90%. Pokud alespoň tři parametry modelu překročily tuto mez, byly dále modelovány areály těchto taxonů.

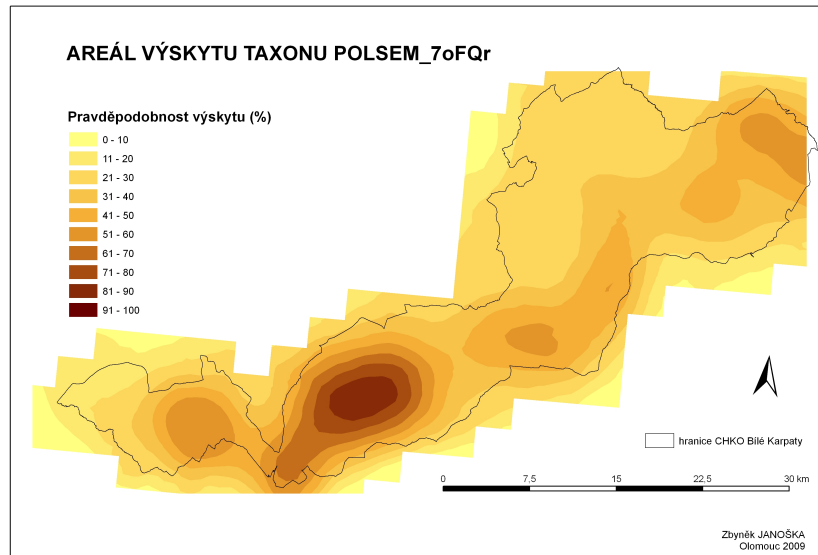
Simulace: pomocí vzorce^[1]:

$$L_{(1)} = \frac{1}{(1 + e^{-Z})} \quad (2)$$

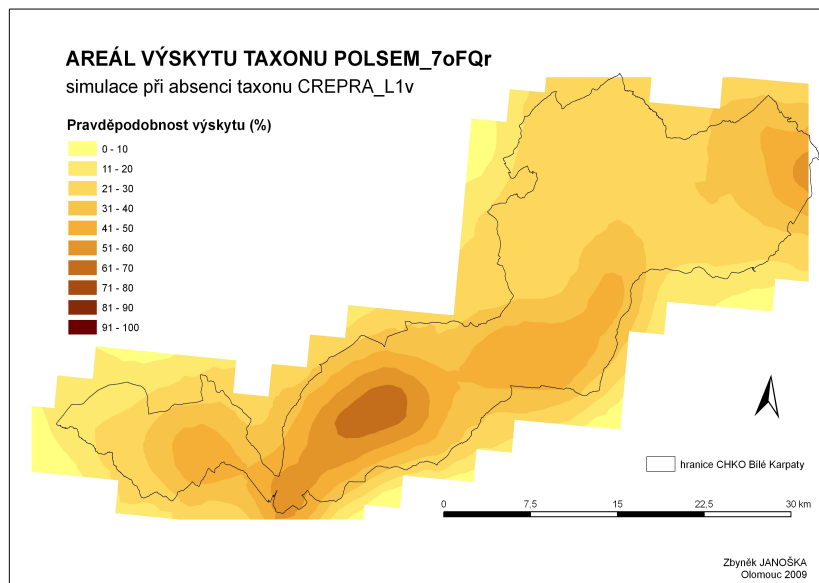
byla vypočtena pravděpodobnost výskytu na každé lokalitě. Interpolací pak byly v prostředí extenze Geostatistical Analyst pro ArcGIS 9.3 sestrojeny areály výskytu.

Protože závislost výskytu byla zjišťována na rostlinách podléhajících určitému režimu ochrany, nabízí se otázka, jak by se změnil areál výskytu v případě vyhynutí těchto druhů. Pro simulace byly vybrány taxony rostlin, které podléhají režimu ochrany C1 (kriticky ohrožený druh) nebo C2 (silně ohrožený druh).

Celkem bylo sestrojeno v rámci práce 9 map současných areálů rozšíření a 8 simulací, předpokládajících vymýcení určitých druhů. U zbylých taxonů nebyla splněna podmínka relevance alespoň tří parametrů modelu.



Obr. 1. areál výskytu taxonu POLSEM_7oFQr



Obr. 2. simulace areálu výskytu při absenci taxonu CREPRA_L1v

4 Závěr

Při vědeckém výzkumu se všechny výsledky odvíjejí od kvality vstupních dat. Problém pořizování dat byl zmíněn v úvodní kapitole, jako možné řešení byla zmíněna předpověď výskytu sledovaných jevů. Při dalším sběru dat je pak možné zaměřit se na konkrétní území a ne hledat „naslepo“. Grafické výstupy by měly pomoci při výběru takových lokalit.

Dalším výstupem jsou obecně formulované rovnice, vyjadřující závislost výskytu denních motýlů na vyšších cévnatých rostlinách. Tyto rovnice mohou pomoci zejména při simulacích, kdy budeme zjišťovat, jak by se změnil areál výskytu daného druhu při dosažení určitých podmínek, například kdyby došlo k vyhynutí jistého druhu apod.

Na závěr bych rád uvedl, že veškeré výsledky práce jsou jen statistické předpovědi, u nichž nikdy nelze dosáhnout absolutní přesnosti. Jejich kvalitu bude nutné posoudit pozorováním v terénu, avšak hlavní jejich význam by měl spočívat v přibližném odhadu lokalit, na které by se vědci, zabývající se sledovanými fenomény, měli při jejich dalším studiu zaměřit.

Rovněž by výsledky neměly být interpretovány tak, že výskyt motýlů je závislý pouze na výskytu rostlin. Lepší je interpretovat výsledky tak, že ukazují na podobnost daných taxonů, ale nevysvětlují závislosti. Je pravděpodobné, že podobné taxony, tak jak byly v práci zjištěny, nejsou závislé na sobě navzájem, ale na společných abiotických podmínkách. Toto bohužel nešlo ze vstupních dat ověřit ani vyvrátit. Je možné, že absence taxonů rostlin tak, jak byla modelována, by se na výskytu motýlů neodrazila, ale pro zjištění detailnějších vazeb by bylo třeba použít pokročilejší metody statistické analýzy a zkoumat vliv více faktorů.

Reference

[1] MELOUN, M., MILITKÝ, J., HILL, M., *Počítačová analýza vícerozměrných dat v příkladech* – Akademie věd ČR, Praha 2005