

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA APLIKOVANÉ GEOINFORMATIKY A ÚZEMNÍHO
PLÁNOVÁNÍ



PROSTOROVÁ NEURČITOST GEODAT V ANALÝZÁCH
DISTRIBUCE VYBRANÝCH DRUHŮ PTÁKŮ
DIPLOMOVÁ PRÁCE

VEDOUCÍ PRÁCE: Ing. Petra ŠÍMOVÁ, Ph.D.

DIPLOMANT: Bc. Markéta NEDBÁLKOVÁ

2013

PROHLÁŠENÍ

Prohlašuji, že jsem tuto diplomovou práci na téma „Prostorová neurčitost geodat v analýzách distribuce vybraných druhů ptáků“ vypracovala samostatně pod vedením Ing. Petry Šimové, Ph.D. a že jsem uvedla všechny literární prameny, ze kterých jsem čerpala.

V Praze dne 22.4.2013

.....

PODĚKOVÁNÍ

Chtěla bych poděkovat všem, kteří mi byli nápomocní při psaní mé diplomové práce, jmenovitě pak mojí vedoucí práce Ing. Petře Šímové, Ph.D. za poskytnuté rady a vedení i Ing. Katce Gdulové za pomoc s daty.

Dále bych ráda poděkovala svojí mamince a celé rodině za neustálou podporu během celého studia.

V Praze dne 22.4.2013

.....

ABSTRAKT

V první části této diplomové práce byla zpracována literární rešerše zabývající se neurčitostí, jejím dělením v rámci ekologie či jejími projevy v procesu modelování druhové distribuce.

V další části práce byla zkoumána závislost výskytu kalouse ušatého (*Asio otus*) na pěti různých environmentálních proměnných (prediktorech), jejichž hodnoty byly získány z map rozdílného rozlišení. Účelem bylo především vyhodnotit, zda rozlišení použitých map ovlivnilo hodnoty proměnných a výsledky zkoumané závislosti.

Jako prediktory výskytu byly použity obvody a obsahy lesů i zemědělské půdy a délka jejich společné hranice na území ČR. Data o výskytu druhu byla převzata z analogového podkladu Atlasu hnízdního rozšíření ptáků v České republice (Šťastný & kol., 2006). Jako geodata s rozdílným měřítkem byla použita data CORINE Land Cover 1990 a 2000 v měřítku 1:100 000, DMÚ 25 s měřítkem 1:25 000, doplněné o LPIS v měřítku 1:10 000.

Pro odhad parametrů jednotlivých prediktorů v sestavených modelech byla použita logistická regrese, jakožto metoda pro vysvětlení závislosti diskrétní veličiny (výskyt/nevýskyt druhu) na spojitých nezávisle proměnných. Jednotlivé modely pak byly srovnány podle výsledných distribučních funkcí a překryvů intervalů spolehlivosti jednotlivých regresních parametrů.

Přestože se použitá geodata lišila ve svém rozlišení, u třech prediktorů nebyly nalezeny větší rozdíly v hodnotách odvozených z těchto geodat. U všech geodatabází byla nalezena závislost výskytu kalouse ušatého na ploše a obvodu zemědělské půdy a na ploše lesů, u DMÚ/LPIS dokonce i na obvodu lesů.

KLÍČOVÁ SLOVA:

neurčitost, modely druhové distribuce, SDM, logistická regrese, kalous ušatý, *Asio otus*

ABSTRACT

In the first part of this Master thesis a literature review about uncertainty, taxonomy of uncertainty for ecology and uncertainty in species distribution modelling was written.

Secondly, the dependence of the long-eared owl (*Asio otus*) on five different environmental variables (predictors), whose values were obtained from maps of different resolution, was investigated. The aim was to evaluate whether the resolution of maps affects values of the variables and therefore the results of studying dependence.

In this study we have used perimeters and areas of forest and agricultural polygons and total length of their common boundary in the Czech Republic as the core predictors of species presence. Data on species occurrence has been obtained from the atlas *Atlas hnízdního rozšíření ptáků v České republice* (Šťastný & kol., 2006). Geodata with different resolutions were used; namely CORINE Land Cover 1990 and 2000 with a scale of 1:100 000, DMÚ 25 which scale is 1:25 000 and LPIS with a scale of 1:10 000.

To estimate the parameters of the individual predictors in established models, logistic regression was used as a method for explaining discrete dependent variable (species presence/absence) according to continuous explanatory variables. The models were then compared by the resulting distribution functions and overlaps of confidence intervals of each parameter.

Although the geodata had varying resolution, there were only small differences in the values of the 3 predictors gained from these geodata. Dependence between long-eared owl occurrence and environmental parameters were found for all geodata sets for 3 predictors - area and perimeter of agricultural polygons and area of forest, and in the case of DMÚ/LPIS for forest perimeter moreover.

KEYWORDS:

uncertainty, species distribution models, SDM, logistic regression, long-eared owl, *Asio otus*

OBSAH

1. ÚVOD	8
2. CÍLE PRÁCE.....	9
3. LITERÁRNÍ REŠERŠE	10
3.1 POPIS DRUHU	10
3.1.1 KALOUS UŠATÝ (<i>ASIO OTUS</i>).....	10
3.2 POPIS DAT	13
3.2.1 ATLASOVÉ ÚDAJE	13
3.2.2 DMÚ 25	14
3.2.3 LPIS	14
3.2.4 CORINE LAND COVER 1990 A 2000 (CLC 90, CLC 2000).....	15
3.3 NEURČITOST	16
3.3.1 VÝVOJ POSTAVENÍ NEURČITOSTI VE VĚDĚ	16
3.3.2 OBECNÉ DĚLENÍ NEURČITOSTI.....	17
3.3.3 DĚLENÍ NEURČITOSTI V EKOLOGII.....	18
3.3.4 NEURČITOSTI V MODELOVÁNÍ DISTRIBUCE DRUHŮ.....	22
3.4 STATISTICKÉ METODY	31
3.4.1 LOGISTICKÁ REGRESE.....	32
4. METODIKA	37
4.1 ATLASOVÉ ÚDAJE.....	37
4.2 GEODATABÁZE	39
4.2.1 CORINE LAND COVER.....	39
4.2.2 LPIS + DMÚ.....	40
4.3 STATISTICKÉ VYHODNOCENÍ.....	42
5. VÝSLEDKY	44
5.1 VIZUÁLNÍ POROVNÁNÍ GEODATABÁZÍ.....	44
5.2 STATISTICKÁ VÝZNAMNOST ODHADNUTÝCH PARAMETRŮ	47
5.3 POROVNÁNÍ ROZSAHU INTERVALŮ SPOLEHLIVOSTI A JEJICH PŘEKRYVŮ.....	48
6. DISKUZE	52
7. ZÁVĚR.....	55
8. POUŽITÁ LITERATURA	56
9. PŘÍLOHY	61
PŘÍLOHA 1	61
PŘÍLOHA 2	62
PŘÍLOHA 3	64
PŘÍLOHA 4.....	66
PŘÍLOHA 5.....	67
PŘÍLOHA 6.....	68

1. ÚVOD

Modelování prostorové distribuce druhů je, jakožto důležitý nástroj pro porozumění základním vztahům mezi druhy a jejich životním prostředím, výchozí pro jakékoli ekologické a biogeografické výzkumy. Přesnost modelů prostorové distribuce druhů může být ovšem ovlivněna různými zdroji neurčitosti (Araújo & Guisan, 2006; Moudrý & Šímová, 2012).

Obecně se dá říci, že zdrojem neurčitosti je nedostatek informací (Martínek, 2007), protože i s pomocí měřících přístrojů nebudou získané informace nikdy kompletně a přesně popisovat ani fyzikální charakteristiky geografického životního prostředí (Duckham, 2002), nemluvě o biotických interakcích a znalostech fungování dynamických přírodních systémů. Proto je nutné prozkoumat koncepční nejasnosti a biotické i algoritmické nejistoty používaných modelů, aby se zvýšila spolehlivost modelovaných výsledků (Araújo & Guisan, 2006).

Pro tvorbu modelů distribuce druhů jsou potřebná environmentální data charakterizující prostředí, v jakém se druh nachází, a informace o výskytu druhu. Ovšem i přes značně významnou roli dat o druhové distribuci je geografie života na Zemi velmi špatně zdokumentována (Jetz & kol., 2012). Ptáci jsou tak jedním z mála taxonů, pro který jsou, díky velkému veřejnému zájmu, data o jejich výskytu zpracovaná dobře (Venier & kol., 2004). I tato práce vychází z *Atlasu hnízdního rozšíření ptáků v České republice 2001-2003*, který je již třetím hnízdním atlasem zpracovaným pro naše území a Česká republika se tak v mapovacích pracích řadí k evropské špičce (Šťastný & kol., 2006).

Pokus provedený v rámci této práce vychází z myšlenky Guisana & Thuillera (2005), že stejné environmentální parametry posbírané v různém rozlišení mohou mít pro druhy rozdílný význam a prostorová neurčitost vstupních dat se pak může šířit do modelových predikcí (Jager & King, 2004). Pokus byl tedy zaměřený na vyhodnocení, zda je nějaká z geodatabází (respektive měřítko některé z bází) CORINE Land Cover a DMÚ25/LPIS vhodnější pro získání hodnot prediktorů výskytu kalouse ušatého na území České republiky.

2. CÍLE PRÁCE

Celá práce vychází z hypotézy, že zvolené měřítko, v jakém byla vytvořena mapa environmentálních proměnných, výrazně ovlivní hodnoty těchto proměnných, neboli hodnoty environmentálních prediktorů využívaných pro určení výskytu druhu. Pokud se pro odvození prediktoru použijí geodata rozlišných měřítek, hodnoty téhož prediktoru by se dle hypotézy měly lišit, a proto by měla být rozdílná i modelovaná distribuce druhu.

Proto je jedním z cílů práce zhodnotit na příkladu dat o výskytu kalouse ušatého a s využitím geodat rozlišných měřítek (CORINE Land Cover 1:100 000 versus DMÚ 25/LPIS 1:25 000/1:10 000) platnost této hypotézy.

Mezi další cíle práce pak patří vyhodnocení, jaký z použitých environmentálních prediktorů nejlépe předpovídá výskyt kalouse ušatého, či zhodnocení, která geodata (respektive jaké mapové měřítko) jsou pro takovou studii vhodnější a zda mapy hrubších měřítek budou skutečně poskytovat méně přesné výsledky nebo zda je možné je pro odvození prediktorů bez obav použít.

3. LITERÁRNÍ REŠERŠE

3.1 Popis druhu

Pro potvrzení / vyvrácení hypotézy, zda měřítko mapy ovlivňuje výsledky výzkumů biotopových efektů na prostorovou distribuci druhů, byla vybrána data o výskytu kalouse ušatého.

3.1.1 Kalous ušatý (*Asio otus*)

Kalous ušatý (viz obr. 1, viz Příloha 6) taxonomicky náleží do třídy ptáků (*Aves*), řádu sov (*Strigiformes*) a čeledi puštíkovitých (*Strigidae*). Svým vzhledem je podobný výru velkému, protože má také péřová „ouška“. Je ale menší, jeho výška je průměrně 35-40 cm. Je relativně štíhlý (hmotnost se pohybuje mezi 250-300 g) se silnými křídly s rozpětím 90 až 100 cm. Má velké oranžové oči a nenápadný černý zobák (Šoltésová, 2007). Jeho základní zbarvení je rezavohnědé s tmavými podélnými skvrnami, které ve spodní části těla přechází v příčky (Šťastný, Bejček & Hudec, 2006). Kalous se řadí mezi druhy CITES uvedené v apendixu II. (Shao & Liu, 2006).



Obr. 1: Kalous ušatý (Bělina, 2010)

Kalous ušatý je holarktický rozšířený druh v mírném pásmu Eurasie, Severní Ameriky a v severní části Afriky (Bejček & Šťastný, 1999). V České republice se vyskytuje prakticky na celém území v hnízdní populaci stálé a přelétavé, kdy se po vyhnízdění rozptyluje ve všech směrech do vzdáleností i 100 km. Méně často je možné pozorovat tažné populace kalousů, u kterých převládá jihozápadní směr tahu (Martiško, 1999).

Kalous žije ve většině nadmořských výšek od nížin až do hor (i když nejhojněji hnízdí do výšky 600 m n. m.). Obývá lesy - převážně jejich okrajové oblasti sousedící s otevřenou zemědělsky využívanou krajinou, skupiny stromů na otevřených prostranstvích, remízky a keřové pásy, doprovodnou zeleň potoků i

městské parky, sady a hřbitovy. Hnízdí většinou do výšky 15 m nad zemí. Hnízda si sám nestaví ani neupravuje, obývá stará hnízda dravců či krkavcovitých ptáků. Méně často může zahnízdit i v budce či přímo na zemi (Šťastný, Bejček & Hudec, 2006).

Výběr hnízdišť je vázán na dostatek potravních zdrojů (které tvoří z 90% hraboš polní (*Microtus arvalis*)) a hnízdních příležitostí (dostatek prázdných či umělých hnízd). Na vhodných lokalitách můžeme najít i 10 hnízdních párů na 100 km², v případě gradace populace hraboše polního se tento počet může i ztrojnásobit (Martiško, 1999).

Aktivita kalousů je nejvyšší za stmívání a v noci a je rozdělena do dvou vln oddělených asi tříhodinovou půlnoční pauzou. Jak již bylo uvedeno výše, hlavní složkou kořisti kalousů je hraboš polní, jehož je hlavním predátorem. Loví v nízkém letu a svou kořist vyhledává pomocí sluchu (Mlíkovský, 1998). Pouze při nízkých stavech hrabošů loví i drobné ptáky, obojživelníky či ostatní drobné savce (Závalský, 2004).



Obr. 2: Kalous ušatý s uloveným hrabošem a s mláďaty (Bohdal, 1984).

Tok u těchto ptáků probíhá většinou od února. Samci se v tomto období opakovaně ozývají temným hlubokým tlumeným „hůu“ v intervalu asi 3 vteřin. Může být také slyšet jejich tleskání křídel pod břichem o sebe. Samice odpovídají slyšitelným „hée“. Pár se vytváří obvykle jenom na jednu sezónu. Zahnízdí pravidelně jednou za rok, kdy snůška vajec probíhá od druhé poloviny března. Při příznivých trofických podmínkách můžou kalousi zahnízdít i dvakrát - začátkem března a v květnu (Martiško, 1999; Šťastný, Bejček & Hudec, 2006).

Samice snáší v průměru 4 - 6 čistě bílých, slabě lesklých, kulovitých vajec s rozměry asi 4 cm x 3 cm a hmotností 20 g. Na vejcích sedí sama a samec jí přináší potravu. Asi po 4 týdnech se postupně líhnou mláďata, která se z hnízda ozývají tesklivým mňoukavým „sí“ nebo „píje“ (Šťastný, Bejček & Hudec, 2006). Samice je ještě asi 9 dní zahřívá, poté se oba rodiče zdržují přes den mimo hnízdo na nedalekém stromě. Mláďata ve stáří asi 21 – 25 dní vyskakují z hnízda na větev nebo padají na zem, kdy ještě nejsou vzletní, ale umí dobře šplhat. Vzletní začínají být od stáří 5 týdnů a pohlavní dospělosti dosahují ve 2 letech (Martiško, 1999).

Úmrtnost mláďat v prvním roce života je až 52%, kdy mohou zahynout buď pádem ze stromu nebo po útoku predátora především v době, kdy ještě nejsou vzletná. Jejich přirozenými predátory jsou kuny, jeřáb lesní, káně lesní i výr velký (Martiško, 1999). Záhlavský (2004) také uvádí jako příčinu značných ztrát u dospělých kalousů automobilovou dopravu.

I přes velkou úmrtnost mláďat je kalous ušatý naší druhou nejpočetnější sovou (hned po puštíkově obecném). Jeho stavy jsou v České republice poměrně stálé a celkový odhad jeho početnosti byl 4000 – 7000 hnízdních párů v letech 1985 – 1989, v období 2001 - 2003 obdobně 4000 – 8000 párů. V Praze je dokonce sovou nejhojnější (Šťastný, Bejček & Hudec, 2006).

3.2 Popis dat

3.2.1 Atlasové údaje

Atlasové údaje o výskytu kalouse ušatého byla převzatá z *Atlasu hnízdního rozšíření ptáků v České republice z roku 2006* (autoři Karel Šťastný, Vladimír Bejček a Karel Hudec), který si klade za cíl zmapovat výskyt a početnosti všech druhů ptáků, jež se vyskytují v době mapování na našem území.

Šťastný, Bejček & Hudec (2006): Mapování probíhalo postupně ve třech fázích:

- V období 1973 – 1977 se mapovalo ve vymezených plochách 100 km² (vždy čtverce 10 x 10 km). Na území České republiky tak připadlo 846 kvadrátů, v kterých po dobu 5 let provádělo terénní práce 913 pozorovatelů.
- Pro mapování v letech 1985 – 1989 došlo ke změně velikosti kvadrátů kvůli unifikování sítě pro mapování všech druhů organismů. Nové rozměry čtverců byly 6' severní šířky x 10' východní délky vyznačených geografickými souřadnicemi. Jednalo se o plochy zhruba 133,2 km² (11,1 km x 12 km). Na území ČR nově připadalo 679 kvadrátů. Ovšem kvadráty, které zasahují méně než polovinou plochy území republiky, byly vyřazeny a výsledky mapování z nich byly přiřazeny k sousedním čtvercům. V Atlasu se tak nakonec pracuje s 628 čtverci. Terénní práce probíhaly také 5 let.
- Poslední mapování proběhlo v období 2001 – 2003. Kvadráty zůstaly nepozměněné, ale zkrátila se doba terénních prací na 3 roky (z důvodu dostatečného počtu pracovníků a jejich zkušenosti).

Terénní pozorování byla zapisována do záznamových druhových karet jednotlivých kvadrátů dle mezinárodních kódů (viz *Příloha 1*). Do mapových výstupů byly použity stupně průkaznosti B (možné hnízdění), C (pravděpodobné hnízdění) a D (prokázané hnízdění). Nasbíraná data byla z druhových karet následně zpracována do digitální podoby pomocí programů ArcView GIS 3.3 ESRI a ESRI ArcMap 9.1 (Šťastný, Bejček & Hudec, 2006).

Výskyt kalouse ušatého byl během všech tří pozorovacích období podobný. V letech 1973 – 1977 byl zaznamenán v 81% kvadrátů (čili v 688 z 843 celkem) a v druhém období v 79% čtverců (v 498 z celkového počtu 628). V letech 2001 - 2003 byl zaznamenán výskyt v 78% (492) kvadrátů, kdy 20% z toho byly údaje o možném hnízdění (99 čtverců), 17% bylo pravděpodobné hnízdění (85 čtverců) a 63% z celkového výskytu tvořilo prokázané hnízdění (308 kvadrátů).

3.2.2 DMÚ 25

DMÚ 25 je digitální model území v měřítku 1:25 000 vytvořený vojenským topografickým ústavem (VTOPÚ) Dobruška (Tichý, 2012). Jedná se o soubor vektorových topografických dat vytvořený digitalizací všech topografických prvků z map s měřítkem 1:25 000.

Účelem jeho vzniku bylo (mimo jiné) vytvoření zdrojové databáze geografických dat pro tvorbu vojenského státního mapového díla i jiných geografických produktů (Břoušek & Laža, 2006).

V této databázi jsou topografické informace rozděleny do sedmi vektorových vrstev na vodstvo, komunikace, sídla, síťová vedení, hranice, reliéf terénu a rostlinný a půdní kryt. Data je možné získat ve formátech ARC/INFO Coverage, ARC/INFO Library a ESRI Shapefile a jsou dostupné v základních souřadnicových systémech S-JTSK, S-42 a WGS 84. Odchyłka dat je udána v rozmezí 50 cm až 20m (Langr, 2001).

3.2.3 LPIS

Český LPIS (Land Parcel Identification System) je referenční registr využití zemědělské půdy v České republice spadající pod Ministerstvo zemědělství. Vznik evidence i proces aktualizace změn v evidenci byl uzákoněn v novele č. 128/2003 Sb. zákona č. 252/1997 o zemědělství.

Důvodem vzniku byla jednak v 90. letech potřeba vytvoření evidence zemědělské půdy pro účely kontroly celého spektra státních dotací poskytovaných na jednotku plochy a dále především podmínka Evropské unie kvůli uvolňování evropských zemědělských dotací, kdy ČR měla zavést ucelený systém v prostředí geografického informačního systému pro identifikaci zemědělských pozemků za základě skutečného užívání půdy.

Uplatnění ale kromě ověřování žádostí o dotace má i jako podklad pro vedení evidence o použití hnojiv či přípravků na ochranu rostlin, podklad pro omezení hospodaření kvůli erozi či jako nástroj v oblasti monitoringu výskytu škodlivých organismů.

První off-line verze registru půdních bloků vznikla koncem roku 2002. Uživatelské bloky zemědělské půdy byly zakresleny do ortofotomap z let 1999 až 2001. Zakreslení bylo následně verifikováno pomocí ověřování zákresu hranic půdních bloků na papírových mapách v měřítku 1:10 000.

Druhou a současnou generací LPIS již bylo on–line řešení vytvořené firmou Sitewell s.r.o. v roce 2004, protože v off-line verzi nebylo možné zajistit dostatečně rychlou aktualizaci a plnou integritu dat.

LPIS se skládá z tzv. „farmářských bloků“, což jsou základní evidenční jednotky reprezentující vždy souvislou plochu zemědělské půdy s jednotnou kulturou obhospodařované jedním uživatelem v jednom režimu obhospodařování. Každý takový blok je zakreslen nad ortofotem a obsahuje informace ohlašované farmářem i vypočtené systémem z digitálního modelu terénu či průniku s jinými geografickými vrstvami. Klíčovým prvkem celého systému je centrální databáze vytvořená v prostředí Oracle Spatial (Sitewell s.r.o., 2004; Ministerstvo zemědělství, 2011).

3.2.4 CORINE Land Cover 1990 a 2000 (CLC 90, CLC 2000)

CORINE (Coordination of information on the environment) Land Cover je projekt, který spadá pod program CORINE fungující od roku 1985. Program CORINE zastřešuje:

- shromáždění informací o stavu životního prostředí, které jsou významné pro všechny členské státy;
- koordinaci při kompilování dat a organizaci informací jak v rámci členských států, tak na mezinárodní úrovni;
- zajištění, že informace jsou odpovídající a data kompatibilní.

CORINE Land Cover (CLC) je jednotná vektorová databáze vytvořená na území celkem 12 států Evropy v měřítku 1:100 000, která pokrývá rozlohu 2,3 mil km². Nejmenší mapová jednotka má plochu 25 ha (Commission of european communities, 1994). Cílem projektu bylo vytvoření metodologicky jednotné a pravidelně aktualizované databáze evropského krajinného pokryvu (Cenia, 2012). Výstupní mapy vegetačního pokryvu byly rozdělené do pěti hlavních tříd a celkem 44 kategorií (viz *Příloha 2*).

Do této doby byly pro území České republiky vytvořeny celkem tři databáze - a to pro roky 1990, 2000 a nejnovější pro rok 2006. Navíc vznikly databáze obsahující změny ploch polygonů jednotlivých tříd mezi lety 1990 - 2000 a 2000 – 2006 (Cenia, 2012).

CLC90 – tato databáze byla zpracována firmou Gisat podle satelitních snímků družice LANDSAT, které byly nasnímány mezi lety 1986 – 1995. Dokončena byla v roce 1996.

CLC2000 – Po převzetí pravomoci za program CORINE Evropskou agenturou životního prostředí (European Environment Agency - EEA) byl zahájen projekt aktualizace evropské databáze I&CLC2000 (zahrnující projekty IMAGE2000 a CLC2000). Výchozím materiálem pro aktualizaci CLC a tvorbu změnové databáze byl soubor satelitních snímků Evropy nasnímaných pomocí družice LANDSAT s prostorovým rozlišením 25 m (projekt IMAGE 2000) (Büttner & kol., 2002; Cenia, 2012).

3.3 Neurčitost

Neurčitost se dotýká všech aspektů našeho života. Naše znalosti nejsou nikdy kompletní (Tanert & kol., 2007) a jak dosavadní výzkum ukazuje, není možné dosáhnout absolutního poznání a podat přesný obraz okolního světa. Informace jsou vždy neúplné a časově proměnlivé a naše poznání se pohybuje v mezích určité tolerance (Dudek, 1999). A protože nelze odstranit neurčitost z jakéhokoli vědeckého výzkumu, je důležitým a zároveň obtížným úkolem identifikace jejich jednotlivých typů a nalezení způsobu změření množství neurčitosti každého typu (Klir, 2008).

3.3.1 Vývoj postavení neurčitosti ve vědě

Klir (2008) shrnuje vývoj postoje k neurčitosti v rámci vědy. Před 20. stoletím byla neurčitost z vědy striktně vyřazována. Zlom přišel počátkem 20. století díky statistickým metodám používaným při výzkumu statistické mechaniky, kdy se neurčitost v některých oblastech vědeckého výzkumu ukázala jako nutná.

Dalším krokem bylo v roce 1927 odvození Heisenbergova principu neurčitosti, který prakticky ukázal, že se věda při měření nemůže neurčitosti vyhnout. V r. 1931 Kurt Gödel ve své práci uvedl i neurčitost v oblasti matematiky, z čehož vyplývalo, že se jí nedá vyhnout ani s pomocí matematiky (Klir, 2008).

Obecné studium neurčitosti začalo v 2. polovině 20. století s rozvojem počítačové technologie. Bylo zjištěno, že i pomocí počítačů je možné zpracovat jen omezené množství informací – bylo nutno provádět zjednodušení a hledat pouze přibližná řešení problémů.

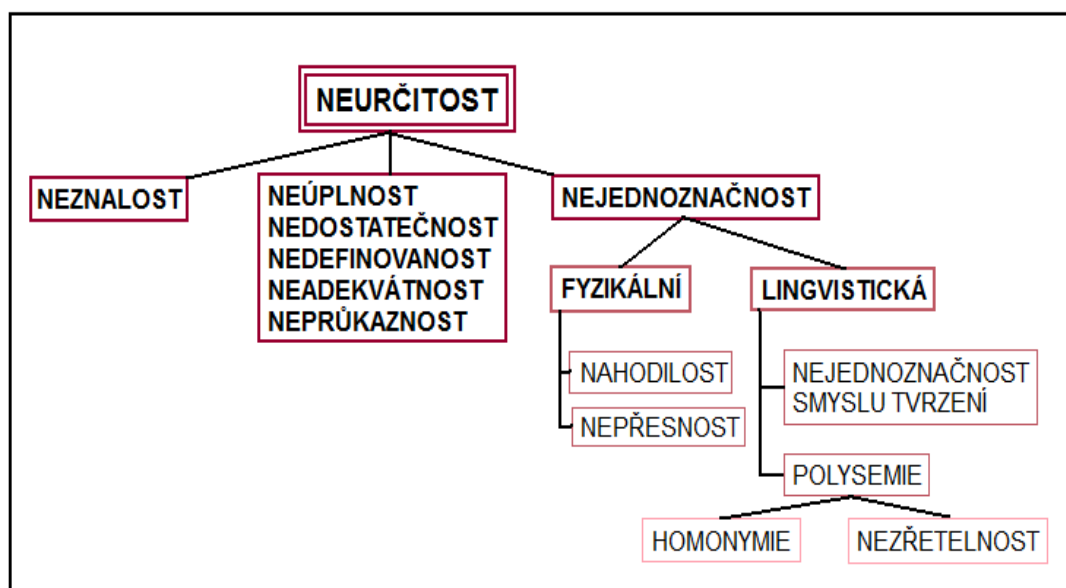
Pohled na neurčitost ve vědě se během minulého století zásadně změnil a neurčitost již není považována za nutné zlo, nýbrž za užitečnou součást vědy.

Mnoho autorů mělo v posledních letech snahu neurčitost nějakým způsobem klasifikovat a popsat, ale žádné všeobecně používané dělení není zavedené. V této práci je proto uvedeno pár způsobů dělení podle jednotlivých autorů.

3.3.2 Obecné dělení neurčitosti

Bokr (1993) i Bokr & Jáneš (2005) dělí neurčitost následovně (*viz obr. 3*):

- **Neznalost** – pod tímto pojmem autoři rozumí absenci vědomostí o entitě či pokud neznáme událost (změnu stavu jinak neměnného objektu), která se odehrála (odehrává či v budoucnu bude odehrávat) na objektu.
- **Neúplnost** nastává, pokud nejsou sesbírány všechny možnosti vědomosti při shromažďování informací o objektu. Principiálně nikdy není komplex znalostí kompletní.
- **Nedostatečnost** je tehdy, nejsou-li shromážděny všechny nezbytné znalosti o událostech.
- **Nedefinovanost** vzniká kvůli absenci exaktního vymezení některých znalostí, které jsou vymezeny například pouze z analogie.
- **Neadekvátnost** znamená, že není s dostupnými znalostmi možné dosáhnout cíl modelování.
- **Neprůkaznost** je zapříčiněná dočasným přerušением sbírání znalostí.
- **Fyzikální nejednoznačnost**
 - Dána existencí více alternativ nastání události, z nichž se každá může uskutečnit nahodile,
 - nebo nepřesností měřících a pozorovacích přístrojů. Všechna měření a pozorování jsou vždy zatížena chybou, která může být vyvolána chybami počítače, chybami při přenosu údajů, omezenou přesností reprezentace i zpracování čísel, omezenou přesností měřících a observačních přístrojů.
- **Lingvistická nejednoznačnost**
 - Nejednoznačnost smyslu nějakého tvrzení je vyvolána konečností slovní zásoby i konečným počtem výroků.
 - Mnohovýznamnost (polysemie) může vést k homonymii (totožnosti formy slov) či nezřetelnosti (přibližné denotativní shodě – jedná-li se o předměty blízkého významu).



Obr 3: Rozdělení neurčitosti dle Bokra (1993).

3.3.3 Dělení neurčitosti v ekologii

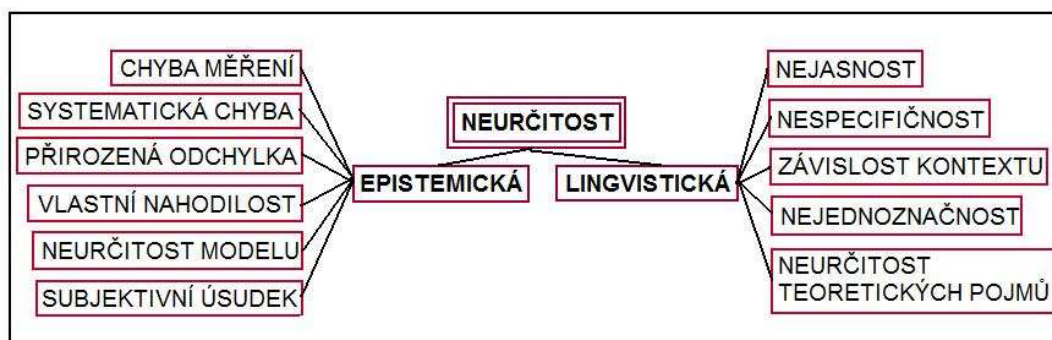
Dle Regan & kol. (2002) neurčitost doprovází veškeré naše snahy o zjištění pravdivých přírodních zákonitostí. V rámci ekologie neurčitost dělí do dvou kategorií (viz obr. 4):

- **nejistota epistemická,**
- **nejistota lingvistická.**

Epistemická nejistota je spojená se znalostí stavu systému, zahrnuje nejistoty vzniklé nepřesností měřících přístrojů, z nedostatečných dat, z extrapolace a interpolace či proměnlivosti v čase a prostoru.

Oproti tomu lingvistická neurčitost vzniká díky faktu, že náš jazyk (včetně velké části vědecké terminologie) je nejednoznačný, často významově závislý na kontextu či vykazující teoretické nejasnosti.

V minulosti již proběhlo několik pokusů mnoha autorů o klasifikování různých typů neurčitosti. Ale ještě nedošlo ke komplexnímu vyhodnocení vlivu celého spektra nejistot během rozhodovacích procesů. Přestože je v ochraně přírody lingvistická nejistota běžná, protože v ní má velký význam politika i rozhodování, je často ignorovaná a uvažuje se pouze nejistota epistemická.



Obr. 4: Rozdělení neurčitosti dle Regan & kol. (2002).

Epistemická nejistota (Regan & kol., 2002)

Epistemickou neurčitost je možné rozčlenit na 6 různých typů, z nichž každý vzniká jiným způsobem:

- **Chyba měření** vzniká z nepřesností měřících přístrojů a pozorovacích technik a zahrnuje chybu obslužného pracovníka i přístroje. Tato nejistota se projevuje jako náhodná změna v kvantitativním měření. Opakovaná měření budou kolísat kolem statistického průměru. V případě absence dalších typů nejistot bude vztah mezi skutečností a naměřenými hodnotami záviset na počtu měření, jejich rozdílných hodnotách, přesnosti měřícího vybavení a zkušenostech terénního pracovníka.

- **Systematická chyba** se vyskytuje v důsledku špatného nastavení měřícího zařízení nebo ovlivněním při sběru dat (dochází k systematickému nadhodnocení či podhodnocení). Na rozdíl od chyby měření tato chyba není náhodná a průměr naměřených hodnot konverguje k jiné, než ke skutečné hodnotě (nekolísá kolem skutečné hodnoty). Systematická chyba může vzniknout jak vědomým rozhodnutím vědce vyloučit/zahrnout data, která neměla být vyloučena/zahrnuta, tak i neúmyslným omylem - např. špatnou kalibrací přístrojů či nesprávným zaznamenáváním měření.

Systematickou chybu je velmi těžké rozpoznat a lze ji odstranit pouze pokud známe její velikost a směr vychýlení (podhodnocení / nadhodnocení).

- **Přirozená odchylka** se vyskytuje v proměnlivých (časově a prostorově) systémech a lze jen těžko předpovídat. Některé třídy nejistoty rozlišují mezi časovými a prostorovými složkami přirozené odchylky.

Přírodní odchylka není sama o sobě zdrojem epistemické neurčitosti, je to pouze skutečná hodnota parametru příslušných změn jako důsledek změn nezávislých proměnných. Je ale často považována za zdroj neurčitosti, protože skutečnou kvantitativní hodnotu je velmi obtížné změřit či předpovědět napříč celou řadou časových a prostorových hodnot.

Důležitá aplikace tohoto pojmu v ekologii je přirozené kolísání populací. Obměna jedinců v populaci může záviset na mnoha faktorech – např. potravní dostupnost, počasí, počty predátorů, rozšíření nemocí, atd.

- **Vlastní nahodilost** se v systému nevyskytuje kvůli našemu omezenému porozumění řídicím procesům a zákonitostem, ale proto, že systém není možné zjednodušit na systém neměnný (deterministický). Přestože je mnoho systémů považováno za nahodilé, je těžké nalézt názorný příklad tohoto druhu neurčitosti.

- **Neurčitost modelu** je výsledkem naší reprezentace fyzikálního a biologického systému. Modely mohou být založeny na diagramech, grafech, matematických reprezentacích, počítačových simulacích, atd. Regan & kol. (2002) se zaměřuje na matematické a počítačové modely, které jsou hojně využívány v přírodních vědách pro predikce budoucích událostí či zjišťování chování systému za specifických podmínek. Neurčitost modelu vzniká dvěma hlavními způsoby. Prvním z nich je fakt, že jsou do modelu zahrnuty pouze parametry a procesy, které se považují za významné pro modelované skutečnosti. Druhou cestou vzniku nejistoty je způsob konstrukce použitý k reprezentaci pozorovaných procesů. Často se jedná o přírodní procesy (reprodukce, vymírání, dostupnost zdrojů, odpočinek, atd.), které nejsou přirozeně matematické povahy a přesto mají dopad na populační početnosti a mohou být reprezentovány v mnoha matematických formách.

Do této kategorie dále patří nejistota vyplývající z prokládání naměřených dat přímkou (včetně interpolace a extrapolace) či z aproximace sestaveného modelu.

Neurčitost modelu je velmi obtížné kvantifikovat a v podstatě nemožné eliminovat. Jediným spolehlivým způsobem zjištění, jak je model přesný a schopný správně predikovat, je provedení validačních studií (což je bohužel často nemožné).

- **Subjektivní úsudek**. Neurčitost vyplývající ze subjektivního úsudku je výsledkem interpretace dat, obzvláště v případě, kdy jsou data vzácná a náchylná k chybám. Prvky subjektivního hodnocení se vyskytují v mnoha odhadech parametrů biologických věd. Standardní způsob vypořádání se s touto nejistotou je přiřadit stupeň důvěryhodnosti události formou subjektivní pravděpodobnosti (například by mohl odborník ohodnotit, že s pravděpodobností 0,7 je velikost populace větší než 500 jedinců). Avšak ani stupně důvěryhodnosti nejsou přesné.

Lingvistická nejistota

Tento typ nejistoty Regan & kol. (2002) dělí do pěti klasifikačních tříd: všechny tyto neurčitosti mohou mít vliv na biologické aplikace.

- **Nejasnost** vzniká kvůli tomu, že náš jazyk umožňuje používat tzv. hraniční případy. Například slovo ohrožený je nejasné z důvodu, že existují druhy, které nejsou ohrožené ani neohrožené – je to hraniční případ.

- **Závislost kontextu** je nejistota vyplývající z neschopnosti specifikovat kontext, ve kterém je potřeba problém chápat. Například řekne-li se, že populace nespecifikovaného taxonu je malá. Bez upřesnění souvislostí lze jen hádat, jak velkou populaci si pod pojmem malá představit.

- **Nejednoznačnost** vyplývá z mnohoznačnosti slov – každé slovo má více významů a není jasné, který význam byl zamýšlen.

- **Nespecifičnost** vyvstává z nechtěné obecnosti, pokud je prohlášení méně specifikované, než bychom potřebovali.

- **Neurčitost teoretických pojmů** vyplývá z problému, že budoucí používání teoretických pojmů není pořádně zakořeněno v minulosti. Což znamená, že některé teoretické pojmy, které zatím nejsou jednoznačné, se jednoznačnými mohou stát.

Nejistota má mnoho rozdílných zdrojů a v různých aplikacích se nejistoty z různých zdrojů hromadí, včetně nejistot epistemických a lingvistických.

Komplexní a systematické vyhodnocení nejistot v biologických vědách má značný význam a zaslouží si, dle Regan & kol. (2002), velké množství dalšího šetření, protože mnoho metod využívaných běžně v ekologii podceňuje jejich skutečný rozsah.

3.3.4 Neurčitosti v modelování distribuce druhů

Přesné mapy rozšíření organismů by byly vítaným zdrojem při navrhování národních parků, hodnocení lidského působení na biodiverzitu či při testování různých ekologických a biogeografických hypotéz. Ovšem ani perfektní znalosti biologie druhů nemohou zaručit, že statická mapa bude správně reflektovat dynamické vlastnosti distribuce druhů. Navíc je téměř nemožné přesně zmapovat velké a nepřístupné přírodní oblasti. Naše znalosti o prostorové distribuci většiny druhů tak mají, jak uvádí Seoane & kol. (2004b), značné mezery. Všeobecným řešením tohoto problému je predikční modelování výskytu druhů, které vzniklo už koncem 70. let minulého století, ovšem za posledních deset let prodělalo obrovský rozvoj (Zimmermann & kol., 2010).

Modely druhové distribuce - SDMs (Species Distribution Models, označované též jako modely niky či vhodných stanovišť, Dormann & kol. (2008)) kombinují body známého výskytu druhu s prostorově spojitými environmentálními veličinami s cílem odvození ekologických požadavků druhu pomocí nějakého statistického algoritmu. Geografická distribuce je pak předpovídána pomocí mapování oblastí, v kterých se všechny environmentální požadavky druhů střetávají (Barry & Elith, 2004; Hernandez & kol., 2006; Dormann & kol., 2008).

A právě proto, že předpovídaná distribuce je obecně určena třemi elementy - environmentálními vrstvami, druhovými daty a algoritmem/modelovací metodou (Fernandez & kol., 2009), Moudrý & Šimová (2012) označují tyto elementy za hlavní zdroje nejistot při tvorbě modelů distribuce druhů.

Distribuce pak může být dále ovlivněna i zvoleným časovým a prostorovým měřítkem použitým při sběru všech druhů dat zahrnutých do modelu (Elith & kol., 2002; Guisan & Thuiller, 2005; McPherson & kol., 2006).

3.3.4.1 Neurčitost nezávisle proměnných

Environmentální proměnné mohou být získány různými způsoby. Dříve běžné metody získávání dat z terénního průzkumu a analogových map postupně nahrazuje satelitní snímkování, které může pokrýt rozsáhlá území s nižší finanční náročností (Mack & kol., 1997).

Proměnné se nejčastěji zpracovávají do vrstev GIS (geografických informačních systémů), které jsou běžně vytvářeny interpolací bodových měření, interpretací leteckých a satelitních snímků nebo modelováním fyzikálních procesů (Elith & kol., 2002). Neurčitost existující již v základních datech (např. kvůli chybám

měření) se potom zvyšuje spolu s každým klasifikováním, interpolováním (Phillips & Marks, 1996) či sumarizováním těchto dat.

Při predikcích výskytu ptáků zřejmě nese velký předpovědní potenciál vegetace. Seoane & kol. (2004a) uvádějí, že tento poznatek potvrzuje i několik probíhajících projektů, které používají vegetační typy pro zmapování potenciálního výskytu ptačích druhů nebo které používají vegetační proměnné k vytvoření predikčních modelů. Dostupné vegetační mapy ale nemusí adekvátně reprezentovat vegetační proměnné relevantní pro ptačí druhy, jejichž výskyt chceme předpovědět, nebo nemusí mít adekvátní rozlišení.

Mapy vegetace, land cover (krajinného pokryvu) i land use (využití půdy) jsou v dnešní době vytvářeny vládními organizacemi v různém rozlišení s různým účelem, ale často jsou zaměřené především na zemědělství, takže informace o vegetačním krytu v přírodních a polopřírodních oblastech nemusí být dostatečně podrobné pro tvorbu modelů a nejlepší potenciální prediktory nemusí být nikdy zmapovány (Seoane & kol., 2004b). Vzdušné senzory, letecké a satelitní snímkování jsou nyní nejvíce používaným nástrojem pro tvorbu nových krajinných map (např. CORINE projekt v Evropě či MIOMBO v Africe), pro zlepšení přesnosti tematických map či finálního rozlišení (Seoane & kol., 2004a), které závisí na prostorovém a spektrálním rozlišení snímacích senzorů satelitů. Například SPOT družice je schopná spolehlivě zaznamenat oblasti s rozlohou větší jak 3 ha. Proto mohou být satelitní data použita pouze tehdy, pokud je v nich možné zaznamenat stanovištní charakteristiky důležité pro výzkum (Mack & kol., 1997).

Obecně se v lokálním měřítku považuje vegetace jako lepší prediktor výskytu, protože má přímější spojení s reprodukčními potřebami druhů než topografie nebo klima (Seoane & kol., 2004b). Ovšem jak dodávají Dormann & kol. (2008), při modelování v ekologii se právě často používají klimatická a topografická data a z nich odvozené environmentální proměnné, které jsou sice snadno dostupné, ale jejich ekologický význam je sporný. Naproti tomu vegetační mapy s vysokým rozlišením jsou velmi nákladné na tvorbu a aktualizaci (Seoane & kol., 2004a), a informace o biologických interakcích, kvalitě půdy či vody, intenzitě vlivu člověka či o historickém vývoji jsou velmi vzácné (Dormann & kol., 2008).

3.3.4.2 Neurčitost závisle proměnné

Kvalita druhových dat může mít hluboké důsledky v modelování distribuce, která je náchylná nejen ke zkreslení v distribuci dat či velikosti vzorku (Araújo & Guisan, 2006).

Nejistoty mohou být způsobeny například chybou měření. Zřídka je zaznamenána celá populace, někteří jedinci mohou být neúmyslně zaznamenáni víckrát, rozdílní pozorovatelé mohou učinit i za identických okolností rozdílná pozorování a poloha jedinců nemusí být zaznamenána přesně (Elith & kol., 2002).

V druhových datech se může vyskytovat i chyba systematická - lokality mohou mít geografické zkreslení, například vzorkování mohlo být zaměřeno na okolí lidských sídel, řek či přístupových cest – čili snadno dostupných míst (Fernandez & kol., 2009) nebo pouze na určitý typ vegetace či krajiny. Takovéto zkreslení geografického prostoru se může promítnout do zkreslení environmentálního prostoru využívaného druhem, díky kterému může dojít ke špatnému posouzení vztahu mezi prediktorem a závislou proměnnou (Araújo & Guisan, 2006). Skrytě žijící druhy mohou být v určitých oblastech konstantně přehlíženy a tím podhodnocovány (Elith & kol., 2002).

Šťastný & kol. (2006) uvádí podobné nejistoty při sběru atlasových dat. V kvadrátech, ve kterých není uveden žádný výskyt daného druhu, mohl být tento druh přehlédnutý. Prázdný kvadrát tak nemusí znamenat nutně absenci. Může se jednat o přehlédnutí, protože je druh obtížně zjistitelný nebo je jeho výskyt běžný, ale mapování prováděl méně zkušený pozorovatel.

Poloha je kritickým atributem prostorových dat. Pozice je typicky definována minimálně dvojrozměrným souřadným systémem, ale měření polohy nemůže být absolutně precizní.

V současné době mnoho dat využívaných k modelování distribuce druhů pochází z muzeí či přírodovědných sbírek, kde je nasbíráno téměř 2,5 bilionu exemplářů od různých pozorovatelů a za použití rozdílných technik (Fernandez & kol., 2009). Takováto data jsou často nekompletní a zkreslená oproti skutečné prostorové distribuci druhů. Mnoho lokalit bylo zaznamenáno pouze textovým popisem či pomocí geografických orientačních bodů bez udaných přesných souřadnic, což komplikuje jejich použití pro GIS analýzy. Je možné je sice zpětně georeferencovat pomocí interpretace textového popisu do souřadnic, ale může to zavést velkou chybu (Araújo & Guisan, 2006). Výrazné a zřetelně viditelné chyby je pak možné odstranit pomocí datových filtrů (Feeley & Silman, 2010).

Pokud je sběr dat prováděn pomocí telemetrie (tj. přenosem radiových vln skrz atmosféru mezi dvěma body), je z hlediska chybného zaměření výhodné použít GPS (Global Positional System) technologie, která umožňuje zaměřit data s velkou prostorovou i časovou přesností (Sand & kol. 2005).

Frair & kol. (2004) ovšem poukazuje na dva typy chyb, které jsou vlastní záznamům o výskytu živočichů zachyceným pomocí GPS telemetrie. Jedná se jednak o chybné zaměření polohy. Špatné určení polohy může vést k chybné klasifikaci využívaného stanoviště v závislosti na velikosti chyby zaměření a heterogenitě krajiny. Ovšem v roce 1990 zavedená úmyslná degradace GPS signálu byla zastavena v roce 2000 a od té doby jsou chyby zaměření menší než 31m v 95% času. S GPS lze dosáhnout přesnosti i 1,5 m, ruční přijímače používané při mapování v terénu zvládají zaměřit polohu s odchylkou 7-10 m, ale tato přesnost se může ještě výrazně snížit například zpomalením signálu při průchodu atmosférou, chybou satelitních hodin, vychýlením dráhy družice či v důsledku mnohacestného odrazu signálu (tzv. multipath) např. v lese (Čábelka, 2008). Druhým problémem mohou být zcela chybějící data kvůli neúspěšné lokalizaci. Tento druh chyby se vyskytuje systematicky. Jak říkají Frair & kol. (2004), pro telemetrické přijímače se prokázalo, že je 3,8 krát nižší pravděpodobnost získání polohy pod vysokým lesním zápojem než v bezlesém území.

Svoji roli hraje i zvolená velikost vzorku – čili celkový počet záznamů o výskytu druhu. Jak uvádí McPherson & kol. (2004), v jejich pokusu zvolená velikost vzorku významně ovlivnila výsledný distribuční model a s rostoucí velikostí vzorku vzrůstala i predikční schopnost modelu. To potvrzuje i Hernandez & kol. (2006), který říká, že s rostoucí velikostí vzorku by měla stoupat i přesnost modelu až k dosažení maximální přesnosti, čili asymptoty. Maximální potenciální přesnost a velikost vzorku, pro který je asymptota dosaženo, závisí na studované oblasti i druzích, kvalitě a prostorovém rozlišení environmentálních proměnných, použitých datech o výskytu druhu i na samotné metodě pro tvorbu modelu.

U prezenčně-absenčních dat je nutné hodnotit velikost vzorku vzhledem k nejméně časté třídě než k celkovému počtu všech záznamů. Jak uvádí Barry & Elith (2006), například u vzácných druhů může náhodný vzorek z celkových 500 pozorování obsahovat pouze 5 záznamů o prezenci druhu. Přestože je vzorek relativně velký, nebude evidentně adekvátní pro tvorbu modelu, protože s 5 záznamy prezence není možné model pořádně specifikovat.

Další otázka vyvstávající při tvorbě modelu distribuce druhu je, jaká druhová data je pro modelování výhodnější použít. Data je možné získat různými metodami průzkumů, transektů, mapování či nesystematického vzorkování, a to pomocí sčítání, zaznamenávání prezence či prezence-absence (Dormann & kol., 2008).

V SDM modelech mohou být použita data pouze prezenční, tzn., že nemáme žádné záznamy o lokalitách, ve kterých se druh nenachází (Pearce & Boyce, 2006), nebo prezenčně-absenční, kdy je zaznamenáván jak výskyt, tak i nevýskyt druhu na dané lokalitě.

Pearce & Boyce (2006) či Ward (2007) uvádí, že prezenčně-absenční data jsou sice běžně používána pro odhad distribuce druhů díky častému využití GLM (Generalized Linear Models – zobecněných lineárních modelů, mezi které patří i logistická regrese) modelů, avšak ne vždy je možné soubor takovýchto dat pořídít a jediná dostupná data jsou pouze zaznamenané prezence (například historické záznamy výskytu druhů, náhodná pozorování výskytu či údaje z radiové telemetrie) (Barry & Elith, 2006). Tito autoři se také zabývali ohodnocením a rozvojem metod pro modelování distribuce druhu vycházející pouze z prezenčních dat.

Gormley & kol. (2011) uvádí, že s oběma typy dat je možné dobře předpovídat současnou i potenciální distribuci druhů. Ovšem podle studie Brotonse & kol. (2004) sice existuje mnoho přístupů a metod pro modelování distribuce druhů a jde jen těžko stanovit jednu univerzálně použitelnou metodu, avšak pokud jsou pro daný druh dostupná absenční data, měly by být ve většině případů použity metody, které je využívají. Výsledky modelů založených na prezenčně/absenčních datech dávají o trochu lepší výsledky, hlavně například u tolerantních druhů s širokým rozšířením.

Přes značně významnou roli dat o druhové distribuci je geografie života na zemi velmi špatně zdokumentována. I u nejlépe prozkoumaných druhů jsou znalosti jejich geografického rozmístění řádově i 4x hrubšího měřítka než většina ostatních důležitých environmentálních proměnných (Jetz & kol., 2012).

3.3.4.3 Specifikace modelu

Neurčitost modelu je dána už tím, že použité parametry obvykle kompletně nepopisují daný ekologický proces. Navíc všechna rozhodnutí z průběhu sestavování modelu (například jaký vybrat model a jaké prediktory do počátečních hypotéz zahrnout, jak zahrnout interakce či se vypořádat s prostorovou autokorelací,

jak definovat vztahy a funkce mezi parametry, atd.) se promítnou do výsledků modelu (Elith & kol., 2002).

Araújo & Guisan (2006) poukazují na to, že s použitím rozdílných technik modelování vznikají i rozdílné výsledky, přestože jsou třeba modely kalibrovány pomocí stejných závisle i nezávisle proměnných. Modelové výstupy jsou řízeny i výběrem prediktorových proměnných vstupujících do modelu a nastavením vztahu mezi nezávisle a závisle proměnnou. Výběr prediktorů je ústředním krokem v modelování. Žádný smysluplný model není možné sestavit bez znalosti ekologie druhu, jeho populační dynamiky a citlivosti k disturbancím. Navíc je otázkou, kolik prediktorů by se do modelu mělo zahrnout. V tomto rozhodování mohou při statistickém vyhodnocování modelů pomoci informační kriteria, například Akaikeho informační kriterium (AIC) srovnávající modely s různým počtem zahrnutých parametrů (Soukup, 2010).

Nejasnosti mohou vzniknout i kvůli rozdílným názorům na to, co model ve skutečnosti reprezentuje – zda fundamentální nebo realizovanou niku. Původní (Hutchinsonova) definice fundamentální niku určuje jako n -rozměrný superobjem, v kterém je konkrétní druh, za absence kompetice s jinými druhy, schopen trvale přežít. Realizovaná nika je pak částí fundamentální niky, do které je druh vytěsněn díky konkurenci ostatních druhů. Nika je často vytvářena na základě dat o výskytu druhu zaznamenaných v mapě v kvadrátové síti, kde velikost kvadrátů může být velká (i 50 km). Druhy, které soutěží o stejné zdroje, se mohou v přírodě objevovat v rámci jednoho kvadrátu, v kterém se pohybují a vzájemně vyhýbají. Čili druhy, jejichž modelované realizované niky se překrývají, se ve skutečnosti mohou v daném prostoru vyskytovat společně. Proto se naskytuje otázka, zda je při modelování vhodné dělit niku na fundamentální a realizovanou a nepovažovat ji za jeden soubor biotických a abiotických faktorů umožňující druhu uspokojit jeho minimální životní požadavky (Araújo & Guisan, 2006).

3.3.4.4 Časové a prostorové měřítko

Volba měřítka je při tvorbě SDM modelů klíčová a nesprávná volba může také přinést zavádějící výsledky (Guisan & Thuiller, 2005).

Naše porozumění fungování ekosystémů a jejich reakcím na jakékoli změny je z hlediska časového a prostorového měřítka velmi komplikované minimálně ze tří důvodů. Zaprvé pozorované vzorce distribuce, které jsou výchozí pro jakékoli analýzy, jsou ovlivněny měřítkem analýzy a se změnou měřítka se mění i vzorce. Dále mohou být vzorce pozorované v určitém měřítku pouze výsledkem mnoha

procesů probíhajících na úrovni jiných měřítek. A zatřetí měřítka, která využívají ekologové, klimatologové, politikové či ochránci přírody jsou tradičně velmi odlišná a často brání efektivní interdisciplinární komunikaci (McPherson & kol., 2006).

Porozumění procesům určujícím pozorované vzorce distribuce je nezbytné pro vyhnutí se nesouladu mezi měřítkem použitým pro modelování a tím, v kterém se objevují klíčové procesy. Protože vzorce pozorované v jednom měřítku nemusí být v jiném měřítku zaznamenané (Guisan & Thuiller, 2005).

Elith & kol. (2002) uvádí, že časová nejistota může postihnout druhová data i prediktorové proměnné. Druhová data jsou obvykle sbírána po dobu několika let, ale pro modelování jsou používána jako aktuální výskyt druhu. Environmentální proměnné mohou být získány například ze starších snímků. Stáří dat pak může způsobit problémy, pokud jsou dohromady smíchány informace o výskytu a vrstvy environmentálních proměnných, které vznikly v jiném časovém období. S tím souhlasí Roubicek & kol. (2010), jejichž výzkum potvrzuje, že pokud jsou záznamy o výskytu druhu ze stejného časového období jako environmentální (v jejich případě klimatická) data, je tím dosaženo výrazně lepších výsledků v predikcích distribučních modelů.

Prostorové měřítko je nejlépe vyjádřeno pomocí rozlišení (velikosti zrn) a rozsahu studované oblasti, které na sobě nejsou závislé, protože studium velkých oblastí ještě neznamená použití hrubého rozlišení (Guisan & Thuiller, 2005).

Rozsah obvykle reflektuje účely analýzy (Elith & Leatwick, 2009). Použití špatného rozsahu (například politických místo přírodních hranic) může poskytovat protichůdné odpovědi na stejné ekologické otázky. Volba rozsahu závisí na znalosti environmentálních gradientů vyskytujících se ve studované oblasti a to tak, aby byly zahrnuty kompletní gradienty. Proto je nejprve nutné kvantifikovat základní rozsah tolerance organismu k prediktorům (Guisan & Thuiller, 2005).

Stejně je to i s rozlišením. Výběr správného rozlišení závisí na velikosti domovského okrsku druhu a způsobu, jak druh využívá zdroje v krajině (Guisan & Thuiller, 2005). Mezidruhové vztahy mohou být zaznamenány pouze v rozlišení, ve kterém druhy soutěží o stejné zdroje. Stejně environmentální parametry posbírané v různých rozlišeních mohou mít pro druhy rozdílný význam (Guisan & Thuiller, 2005).

Dále data nasbíraná v relativně hrubém prostorovém měřítku mohou mít nedostatek dostatečně užitečných informací pro ekologické procesy vyskytující se v jemnějším měřítku. Struktury vyskytující se v jemnějších měřítkách tak mohou být

zahrnovány do převažujících celků, což může zapříčinit zkreslení neobvyklých tříd – například vzácných vegetačních typů (Elith & kol., 2002). Druhová diverzita krajiny tak závisí na prostorovém měřítku, v jakém byla společenstva vzorkována (Hammer & Hill, 2000). S tím souhlasí i Oja & kol. (2005) a doplňuje, že prostorové měřítko, v jakém je měřená fragmentace krajiny, může ovlivnit výsledný efekt fragmentace na hnízdní úspěšnost ptáků a že závislost mezi vybranou krajinnou metrikou a druhovou distribucí je značná.

Problém při tvorbě SDM se pak může vyskytnout při použití vstupních dat s rozdílným rozlišením. Optimálně by druhová i environmentální data měla být nasbírána ve stejném rozlišení (Elith & kol., 2002).

Se vzrůstajícím zájmem států o zdokumentování jejich přírodního dědictví se rozrostla i dostupnost atlasů popisujících distribuci druhů v různých zemích i kontinentech. Přesto prostorové rozlišení těchto dat bývá hrubé vzhledem k měřítkům, jaká jsou používána například pro analýzy ochrany přírody. Management volně žijících živočichů či environmentální hodnocení obvykle pracují s měřítkem v řádu několika stovek hektarů. Oproti tomu jsou druhové distribuce v atlasech běžně mapovány s rozlišením od 100 km² v dobře zmapovaných regionech do více jak 10 000 km² ve špatně prozkoumaných částech světa.

McPherson & kol. (2006) popsaly metody pro odvození jemnozrnných dat (potřebných pro ekologické výzkumy) od dat hrubých měřítek, protože se již místo abiotických faktorů velkých měřítek považují biotické interakce s vysokým rozlišením za hlavní procesy ovlivňující druhovou distribuci.

Snahou o objasnění, zda měřítko, v kterém byla data nasbírána, ovlivňuje výsledek modelu distribuce druhu se zabývali například autoři Venier & kol. (2004) a Seoane & kol. (2004a).

Jak uvádí Seoane & kol. (2004a), ptačí druhy jsou svým výskytem vázány k určitým typům vegetace, avšak prediktivní modely rozšíření ptáků, které jsou založené právě na proměnných odvozených od vegetace, mívají někdy pouze omezený úspěch. Proto bylo navrženo, že přesnost existujících map používaných k odvozování prediktorů výskytu je částečně zodpovědná za omezený úspěch modelů rozšíření ptáků.

Byly provedeny následující dva pokusy, které porovnávají predikční schopnosti distribučních modelů výskytu ptáků odvozených od různých sad map vegetačního krytu.

- Pokus č. 1 (Seoane & kol., 2004a):

Tento pokus proběhl ve Španělsku v západní Andalusii ve dvou kvadrátech s plochou 4900 km². Cílem bylo porovnat predikční schopnosti distribučních modelů odvozených od již existujících obecných map land use/land cover, které se lišily svým rozlišením a přesností.

Byly použity tři nezávislé vegetační mapy- s nejhrubším měřítkem CORINE Land Cover 1990 od EEA v rastrovém formátu s rozlišením 250 m. Dále s jemnějším rozlišením byly použity regionální mapy zobrazující vegetační strukturu Andalusie, z nich jedna byla mapa vegetačního krytu z r. 1995 vytvořená společností SINAMBA, převedená z vektorového formátu na rastr s rozlišením 50 m, a druhá byla odvozená od satelitních snímků z let 1999 a 2000 s rozlišením rastru 30 m. Následně byla vytvořena ještě kombinace zlepšením přesnosti mapy SINAMBA satelitními snímky a další tři odvozené mapy od map SINAMBA a mapy satelitních snímků s uměle vytvořeným hrubším měřítkem.

Bylo provedeno vzorkování prezence/absence druhu v 1144 bodech, kde se v každém bodě zaznamenával výskyt/nevýskyt druhu během patnáctiminutového intervalu. Následně bylo vybráno 857 bodů z přírodních či polopřírodních oblastí a 54 ptačích druhů, které se vyskytovaly na více jak 5% z celkového počtu zahrnutých bodů. Jako potenciální prediktor výskytu byl použit stejný soubor krajinných a vegetačních proměnných změřených všech mapách.

Přestože se výchozí vegetační mapy liší ve své přesnosti (např. CORINE má hrubé měřítko, omezený počet tříd vegetačního krytu, dobře nereflexuje rozdíly ve vegetační struktuře a je o 10 let starší než druhová data), výsledky ukazují, že všechny produkující podobně přesné predikční modely. Seoane & kol. (2004a) podotýkají, že některé všeobecné mapy vegetačního krytu či krajinné struktury jsou dostatečně přesné pro odvození modelů předpovídajících výskyt ptačích druhů.

- Pokus č. 2 (Venier & kol 2004):

Cílem tohoto pokusu bylo vyšetřit schopnost, jak lze na základě regionálního klimatu a satelitních dat land cover s různým měřítkem předpovědět výskyt 10 druhů ptáků v oblasti Velkých jezer v Severní Americe, čili oblasti o ploše asi 800 000 km², v rozlišení odpovídajícímu atlasovým datům s kvadráty 5 – 11 km².

Data vegetačního krytu odvozená ze satelitních snímků jednak pomocí metody AVHRR (advanced very high resolution radiometer) s rozlišením asi 1 km² a jednak pomocí MSS(Landsat multi-spektral scanner) s rozlišením 200m². Dále byla použita ještě klimatická data (teplota, srážky).

Druhová data byla přebrána z atlasů hnízdních rozšíření ptáků a zahrnovala prezenci/absenci druhu v jednotlivých kvadrátech. Bylo vybráno 10 lesních pěvců, kteří využívají rozmanitá lesní stanoviště ve výsledných 1302 čtvercích.

Pomocí logistické regrese byly vytvořeny modely pro všech 10 druhů ptáků, kde jako vysvětlující proměnné byly použity: krajinný pokryv AVHRR; krajinný pokryv MSS; klimatická data; kombinace klima + AVHRR a klima+ MSS.

Výsledky této studie ukazují, že jak krajinný pokryv vytvořený v různém rozlišení ze satelitních snímků, tak klima byly dobrými a srovnatelnými prediktory výskytu všech 10 druhů ptáků. Nejlepších výsledků bylo dosaženo s kombinací informací krajinného pokryvu a klimatických podmínek.

Ve výše uvedených pokusech nebyl nalezen větší rozdíl mezi predikční silou modelů. Jak uvádí Moudrý & Šimová (2012), bylo to pravděpodobně z důvodu použití stále ještě jemného měřítka s ohledem na velikost zkoumaného území.

3.4 Statistické metody

Statistické metody, pomocí nichž lze odhadovat hodnotu nějaké náhodné veličiny (závisle proměnné) na základě jiných známých veličin (nezávisle proměnných) se označují jako regresní analýza.

Jak uvádí Řeháková (2000), cílem regresních analýz je nalezení nejlepšího a smysluplného modelu, který dokáže popsat vztah mezi závislou proměnnou a nezávisle proměnnými. Podle Dudka (1999) nám právě modely na základě nalezení příčiny prozatímního vývoje umožňují za pomoci regresní analýzy (či korelační analýzy, nejedná-li se o jednostrannou, ale vzájemnou závislost) předpovídat budoucí vývoj a tím snížit neurčitost znalostí o budoucnu.

Regresních modelů existuje celá řada, výběr modelu závisí na typu a rozložení dat. Pokud jsou hodnoty závisle proměnné kvalitativní, tedy nespojitě a nabývající pouze dvou hodnot 0 a 1 (například prezence či absence druhu v kvadrátu), pro vysvětlení takovéto proměnné se využívá logistická regrese.

3.4.1 Logistická regrese

Logistická regrese je tedy metoda, která umí vysvětlit chování diskrétní veličiny s alternativním rozdělením.

Alternativní rozdělení je speciální případ binomického rozdělení. Veličina X má binomické rozdělení, pokud nabývá hodnot pouze $0, 1, 2, \dots, n$ s pravděpodobnostmi

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, 2, \dots, n$$
$$n \in \mathbb{N}; \quad p \in (0, 1)$$

Střední hodnota $\mu = np$ a rozptyl $\sigma^2 = np(1-p) = npq$. Pokud je $n = 1$, jedná se o alternativní rozdělení (Anděl, 2011).

Vysvětlovaná veličina je tedy obvykle binární - nabývající pouze dvou hodnot (i když existuje logistická regrese, pomocí které jsou výpočty možné i pro více jak dvě hodnoty) (Řeháková, 2000; Zvára, 2008).

Předpokladem je, že závisle proměnná Y_i může nabývat hodnoty 0 a 1.

- pokud $Y_i = 1$, jev J nastal;
- pokud $Y_i = 0$, jev J nenastal

Podle Řehákové (2000) je potřeba zjistit, zda je možné roztřídit sledované případy do dvou kategorií závisle proměnné (čili do kategorií 0 a 1) na základě skupiny nezávisle proměnných s normálním rozdělením. Snahou tedy je predikovat pravděpodobnost, s jakou případ spadá do jedné kategorie závisle proměnné (protože jak uvádí Hendl (2009), pomocí pravděpodobnosti lze modelovat nejrůznější jevy). Pokud známe $P(Y_i = 1)$, známe i $P(Y_i = 0)$:

$$P(Y_i = 0) = 1 - P(Y_i = 1)$$

Střední hodnota μ_i vysvětlované závisle proměnné Y_i se bude rovnat pravděpodobnosti výskytu sledovaného jevu (pravděpodobnosti jedničky, $\mu_i = P(Y_i = 1)$). Rozptyl Y_i bude záviset na této střední hodnotě a bude roven $\text{var } Y_i = \mu_i(1 - \mu_i)$. Zároveň bude střední hodnota (čili pravděpodobnost jedničky) vyjádřena jako funkce nezávisle proměnných x_i (Zvára, 2008). Ale protože je pravděpodobnost jevu nějaké číslo z intervalu $\langle 0; 1 \rangle$, není možné modelovat $P(Y_i = 1)$ regresní rovnicí:

$$P(Y_i = 1) = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i,$$

protože kořeny této rovnice mohou být jakákoli reálná čísla (Řeháková, 2000). Je tedy nezbytné zaměnit pravděpodobnost za jiný tvar závislosti, který bude interpretovatelný touto rovnicí.

Nejprve je možné zaměnit pravděpodobnost jevu jeho šancí. Tzv. „šance jevu“ (neboli „odds“), že $Y_i = 1$ (šance, že jev nastal) je určena podílem pravděpodobností, že jev nastal a že nenastal.

$$\text{šance}(Y_i = 1) = \frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \frac{\mu_i}{1 - \mu_i}$$

Minimální hodnota šance je nula, proto ještě šanci převedeme na její přirozený logaritmus. Zlogaritmovaná šance se nazývá „logit“, proto se model popsany níže uvedenou rovnicí nazývá binární logistický regresní model (Hendl, 2009).

$$\text{Logit}(Y_i) = \ln \frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \ln \frac{\mu_i}{1 - \mu_i}$$

Protože hodnoty logitu mohou nabývat nekonečných hodnot, využijeme ho jako závisle proměnnou a získáme regresní rovnici, kde je logit pravděpodobnosti roven lineární funkci neznámých parametrů (Řeháková, 2000; Zvára, 2008).

$$\text{Logit}(Y_i) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k.$$

Převod logitu zpět na šanci je možný pomocí exponenciální funkce

$$\text{šance}(Y_i = 1) = e^{\text{Logit}(Y)} = e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} = e^{\beta_0} \cdot e^{\beta_1 x_1} \cdot \dots \cdot e^{\beta_k x_k}$$

a na pravděpodobnost:

$$P(Y_i = 1) = \frac{\text{šance}(Y_i = 1)}{1 + \text{šance}(Y_i = 1)} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}$$

čímž je zaručena platnost, že hodnoty $\mu_i \in (0, 1)$.

- Odhad parametrů:

Podle Rychnovského (2008) a Zváry (2008) je cílem najít co nejlepší odhad parametrů modelu, čili vektoru β . Při odhadu modelu logistické regrese se většinou nepoužívá *metoda nejmenších čtverců* jako u lineární regrese, ale vychází se z *metody maximální věrohodnosti L* (označení L = likelihood). Ta spočívá v konstrukci věrohodnostní funkce udávající pravděpodobnost, s níž nastanou v daném modelu všechny pozorované hodnoty. Vyhovující je model s maximální pravděpodobností.

Souhrnně lze zapsat pravděpodobnosti dvou možných hodnot $Y_i = 1$ a $Y_i = 0$ jako

$$P(Y_i = j) = \mu_i^j (1 - \mu_i)^{1-j} \quad \text{pro } j = 0, 1.$$

Podle předpokladu jsou pozorované hodnoty nezávislé, lze proto definovat věrohodnostní funkci $L(\beta)$ pomocí součinu podmíněných pravděpodobností pro jednotlivá pozorování.

$$L(\beta) = \prod_{i=1}^n \mu_i^{y_i} (1 - \mu_i)^{1-y_i}$$

Tato věrohodnostní funkce je kvůli nalezení jejího maxima zlogaritmována. Logaritmus neovlivní polohu extrému a výsledná funkce se bude snadněji derivovat. Vzniká tak

$$l(\mu) = \ln \prod_{i=1}^n \mu_i^{y_i} (1 - \mu_i)^{1-y_i} = \sum_{i=1}^n (y_i \ln \mu_i + (1 - y_i) \ln(1 - \mu_i)) = \sum_{i=1}^n y_i \ln \left(\frac{\mu_i}{1 - \mu_i} \right) + \sum_{i=1}^n \ln(1 - \mu_i)$$

Pro získání hledaného maxima vzhledem k vektoru parametrů β je vhodné nahlížet na funkci μ_i jako na funkci proměnných β a x . Platí:

$$l(\beta) = \sum_{i=1}^n y_i \eta_i(\beta) + \sum_{i=1}^n \ln(1 - \mu_i(\beta)),$$

$$\text{kde } \eta_i = \text{Logit}(\beta) = \ln \left(\frac{\mu_i}{1 - \mu_i} \right)$$

a také platí rovnice

$$\frac{\partial}{\partial \eta_i} \ln(1 - \mu_i) = -\frac{\partial}{\partial \eta_i} \ln(1 + e^{\eta_i}) = -\frac{e^{\eta_i}}{1 + e^{\eta_i}} = -\mu_i$$

Z těchto dvou rovnic získáme vztah pro parciální derivace logaritmické věrohodnostní funkce:

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^n \frac{\partial l}{\partial \beta} \frac{\partial \eta_i}{\partial \beta} = \sum_{i=1}^n (y_i - \mu_i(\beta)) x_i .$$

Jednotlivé parciální derivace funkce $L(\beta)$ budou podle parametrů $\beta_0, \beta_1, \dots, \beta_k$ položeny jako rovny nule. Tak se získá nelineární soustava tzv. věrohodnostních rovnic:

$$\sum_{i=1}^n (y_i - \mu_i) = 0$$

a

$$\sum_{i=1}^n x_{ij} (y_i - \mu_i) = 0, \quad \text{pro } j = 1, 2, \dots, k; \text{ kde } x_{ij} \text{ je } j\text{-tá složka vektoru } x_i.$$

Tyto rovnice se obvykle řeší numericky s využitím nějakého statistického softwaru. Řešením získáme maximálně věrohodný odhad β' vektoru β (Rychnovský, 2008; Zvára, 2008).

- Interpretace parametrů

Nejjednodušší model logistické regrese je model s jednou nezávislou proměnnou

$$\eta_i = \beta_0 + \beta_1 x_1, \quad \text{kde } \eta_i = \text{Logit}(y_i) = \ln \frac{\mu_i}{1 - \mu_i}$$

V práci použité nezávisle proměnné jsou spojité veličiny. Šance se pro zvolenou hodnotu x rovná:

$$\text{šance}(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} .$$

Pokud u sledovaného výsledku porovnáme šance dvou hodnot nezávisle proměnných, které se liší o jedničku, získáme:

$$\frac{\omega(x+1)}{\omega(x)} = \frac{e^{\beta_0 + \beta_1(x+1)}}{e^{\beta_0 + \beta_1 x}} = e^{\beta_1}.$$

Z toho vyplývá, že parametr β_1 dává informaci o změně logaritmu poměru šancí vtažené k jednotkovému přírůstku nezávisle proměnné x . Testovaná nulová hypotéza $H_0: \beta_1 = 0$ by pak znamenala stejné šance pro obě hodnoty nezávisle proměnné, neboli nezávislost šancí (čili pravděpodobnosti $P(Y_i = 1)$) na nezávisle proměnné x (Zvára, 2008).

4. METODIKA

V rámci této diplomové práce byl proveden pokus, jehož cílem bylo zhodnotit, zda vysvětlující proměnné podrobnějšího měřítka lépe vysvětlují distribuci druhu. Jinými slovy byly z výše popsaných map s rozdílným rozlišením (CORINE Land Cover, DMÚ a LPIS) získány stejné prediktory výskytu kalouse ušatého, které byly následně statisticky zpracovány. Realizace samotného pokusu proto vyžadovala jednak přípravu dat do formátu vhodného pro statistické zpracování a jednak vyhodnocení vytvořených statistických modelů.

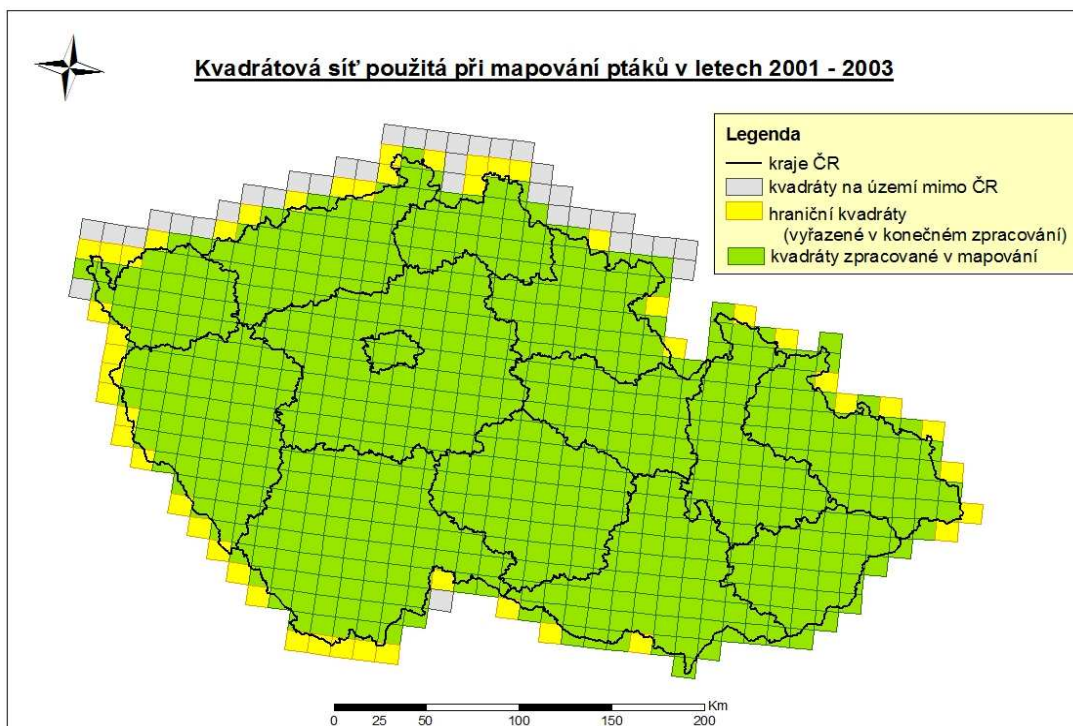
Dle zadání bylo původně zamýšleno v pokusu použít geodatabázi ARC ČR 500, ale nakonec byla zaměněna za CORINE Land Cover, protože z ní nebylo možné získat vrstvu zemědělské půdy (případně tuto vrstvu doplnit z jiné pro tuto práci dostupné databáze podobného měřítka jako v případě DMÚ a LPIS).

4.1 Atlasové údaje

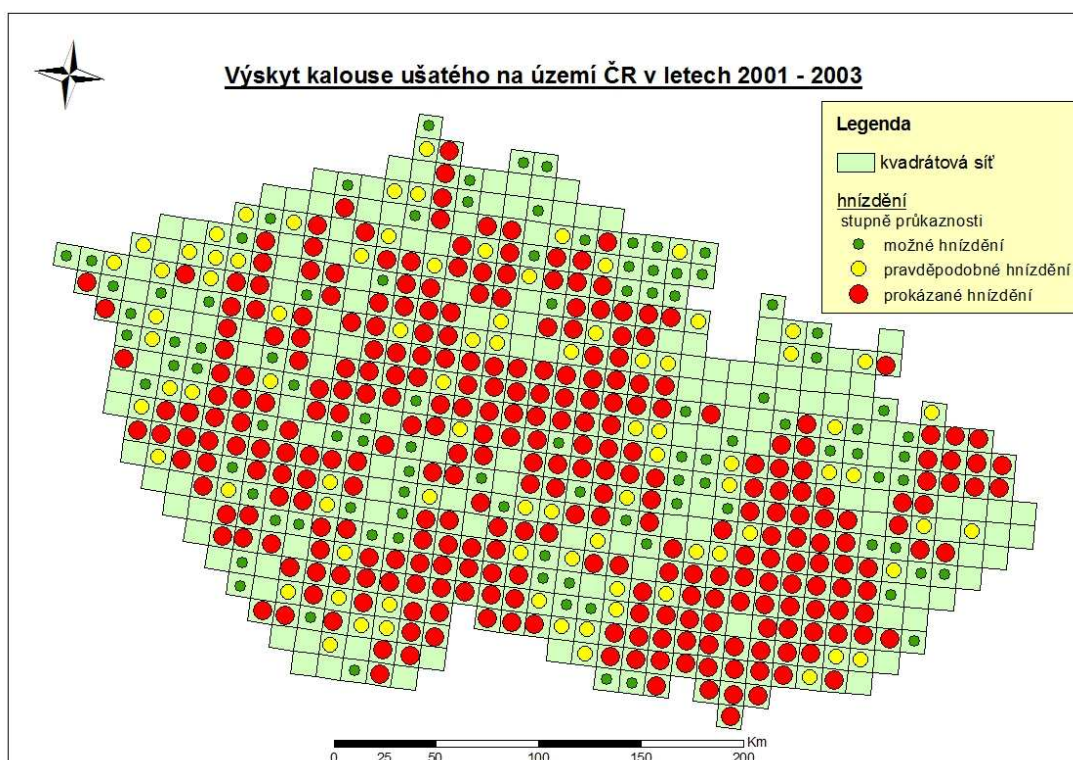
Pro použití druhových dat bylo nutné převést atlasové údaje o výskytu kalouse ušatého do digitální podoby. Byl k tomu využit program ArcGIS 9.3 a již v digitální formě zpracovaná vrstva kvadrátové sítě zahrnující území České republiky.

Kvadrátová síť je vytvořena pro území celé Evropy s účelem zaznamenávání výskytu a jeho změn u rostlinných a živočišných druhů. Metoda síťového mapování používá lichoběžníkové kvadráty s označením KFME (Kartierung der Flora Mitteleuropas), které jsou vyznačeny zeměpisnými souřadnicemi 6 minut zeměpisné šířky a 10 minut zeměpisné délky. V České republice jsou ale k dispozici 2 GIS vrstvy obsahující kvadrátovou síť našeho území a splňující parametry sítě KFME. Ovšem nejsou totožné, při jejich překrytí se hranice mohou lišit až do vzdálenosti 100 m. Tato nesrovnalost vznikla díky rozdílnému elipsoidu, na kterém byla základní síť KFME nadefinovaná. Podle poledníků a rovnoběžek Krasovského elipsoidu vznikla vrstva *q01_cr* a podle souřadnic Besselova elipsoidu vrstva *sitmap_cz* (Janitor, 2003; Anděra, 2005).

Pro tuto práci byla použita vrstva *q01_cr*. Součástí vrstvy původně bylo 714 kvadrátů, z nichž 36 nezasahovalo vůbec území ČR a u dalších 50-ti nepokrývalo území republiky ani polovinu plochy, proto byly podle popisu Šťastného & kol. (2006) (viz kap. 3.2.1) a jejich zpracování v Atlasu hnízdního rozšíření ptáků vyřazeny (viz obr. 5).



Obr. 5: Kvadrátová síť použitá v mapování v letech 2001 – 2003 upravená dle Šťastného & kol. (2006)



Obr. 6: Zaznamenané hnízdění kalouse ušaté v jednotlivých kvadrátech dle Šťastného & kol. (2006), v kvadrátech bez znázorněného bodu nebyl zaznamenán žádný výskyt.

Následně byla vytvořena nová vektorová bodová vrstva hnízdění funkcí *FEATURE TO POINT*, kdy se vytvořil bod v centru každého kvadrátu. Podle v atlasu zaznamenaného výskytu byla ke každému bodu přiřazena v atributové tabulce průkaznost hnízdění (viz obr. 6). Nakonec byla vygenerována tabulka *vyskyt.dbf* obsahující pouze kódy čtverců a výskyt kalouse.

4.2 Geodatabáze

Při přípravě dat ze všech geodatabází bylo vycházeno především z navrženého předpokladu, že celková délka hranice lesů sousedících se zemědělskou půdou významně předpovídá výskyt druhu. Vždy bylo nutné získat dvě polygonové vrstvy – a to možná místa výskytu hnízd (oblasti lesů) a dále oblasti s předpokládaným četným výskytem hraboše polního čili lokality s výskytem potravy (oblasti zemědělské půdy). Následně byly změřeny: délka společné hranice obou typů polygonových vrstev připadající na každý kvadrát ČR, rozlohy vrstvy lesů i zemědělské půdy připadající na každý kvadrát a také samostatné délky hranic ohraničujících jak lesní, tak zemědělskou půdu v každém zahrnutém kvadrátu. Všechna tato měření byla použita jako prediktory modelů pro výskyt druhu.

4.2.1 CORINE Land Cover

Nejsložitější struktura vrstev je v databázi CORINE Land Cover (viz Příloha 2). Dělí se na pět hlavních tříd, ze kterých by z hlediska výskytu bylo možné použít některé kategorie třídy 1 (urbanizovaná území), 2 (zemědělské plochy) a 3 (lesy a polopřírodní oblasti). Ale aby výsledky byly srovnatelné s ostatními geodatabázemi, ze kterých byly k dispozici pouze lesy a zemědělská půda, byly použity pouze kategorie tříd 2 a 3. Při výběru jednotlivých kategorií bylo vycházeno z informací o oblastech hnízdění a lovu kalouse ušatého (viz kapitola 3.1.1). Vybrané kategorie jsou znázorněny v tab. 1.

Oblast hnízdění		Oblast lovu	
číslo třídy	název	číslo třídy	název
222	ovocné sady a keře	211	orná půda mimo zavlažovaných ploch
311	listnaté lesy	221	vinice
312	jehličnaté lesy	231	louky
313	smíšené lesy	241	roční kultury přidané ke stálým kulturám
324	přechodová stadia lesa a křoviny	242	komplexní systémy kultur a parcel
		243	převážně zemědělská území s příměsí přirozené vegetace
		321	přírodní pastviny

Tab. 1: kombinace tříd CLC vybraných pro oblasti hnízdění a lovu kalouse ušatého.

Snahou bylo třídy rozdělit tak, aby v oblastech lovu byla zahrnuta místa s velkými travnatými či zemědělsky využívanými plochami.

V programu ArcGIS 9.3 byly dle výše uvedeného výběru tříd vytvořeny ze zadaných databázových vrstev *clc_90.shp / clc_00.shp* (čili jak z dat CLC 90, tak následně i z aktualizované CLC 2000) dvě samostatné polygonové vrstvy – *hnizdeni_C90 / hnizdeni_C00* a *lov_C90 / lov_C00*. Vrstvy byly vytvořeny spojením výše uvedených kategorií. Hranice jejich polygonů byly převedeny na linie a funkcí *INTERSECT* postupně vytvořena vrstva obsahující pouze společnou hranici pro oba typy polygonových vrstev, která je rozdělena do linií podle kvadrátové mapy. V atributové tabulce této vrstvy byla vygenerována sumarizační tabulka obsahující součet délek všech společných hranic v rámci každého kvadrátu.

Následně byly vytvořeny zjednodušené polygonové vrstvy hnízdění a lovu funkcí *DISSOLVE*. Touto funkcí byly odstraněny vnitřní polygonové hranice, aby se při výpočtu prediktoru obvodu lesa/zemědělské půdy nezapočítávaly i hranice mezi dvěma sousedícími lesními/zemědělsky využívanými polygony. Nakonec byly vygenerovány 2 sumarizační tabulky obsahující rozlohu a obvod oblastí lesů a oblastí zemědělské půdy podle kvadrátů.

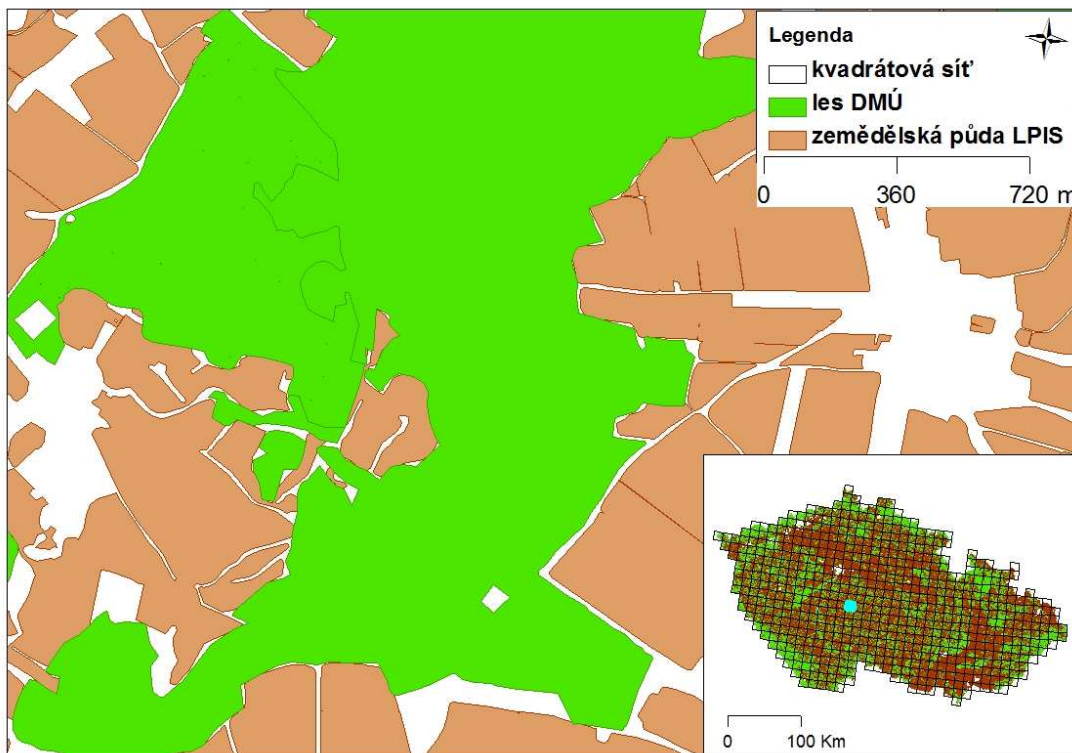
4.2.2 LPIS + DMÚ

Pro získání námi požadovaných prediktorů v jemnějším měřítku, než v jakém jsou tyto údaje nasbírány v CORINE, byly použity dvě databáze – DMÚ 25 a LPIS. Z DMÚ byla použita polygonová vrstva lesů, která byla doplněna zemědělskou půdou z registru LPIS.

Z DMÚ bylo nejprve nutné získat vektorovou vrstvu lesů pro celou ČR. Polygonové vrstvy lesů byly uloženy po jednotlivých okresech republiky, proto bylo nutné použít funkce *MERGE* na spojení lesů do jedné vrstvy *les_DMU*.

Dále byly použitím funkce *DISSOLVE* odstraněny vnitřní polygonové hranice obou vrstev *LPIS* i *les_DMU* a funkcí *INTERSECT* byly tyto vrstvy spojeny s kvadrátovou sítí. Poté byly, stejně jako v případě *CORINE*, dopočteny obvody a obsahy polygonů obou vrstev v jejich atributových tabulkách a vygenerovány sumarizační tabulky obsahující součty obvodů a ploch lesů / zemědělské půdy v rámci každého kvadrátu.

K oběma vrstvám *LPIS* i *les_DMU* byly vytvořeny 50 m široké buffery kolem polygonových hranic a funkcí *INTERSECT* udělán jejich průnik. Tento postup pro získání délky společné hranice byl zvolen proto, že při tvorbě byla každá vrstva vektorizována z jiného podkladu a polygonové hranice lesů přesně neseseděly na hranice zemědělské půdy (viz obr. 7). Z průniku bufferů vznikly úzké polygony, jejichž poloviční obvod připadající na každý kvadrát byl považován za společnou hranici lesa a zemědělské půdy.



Obr. 7: Detail části kvadrátu 6354 (označeného na mapce ČR modrou barvou) znázorňující mezeru mezi polygony jednotlivých geodatabází DMÚ25 (zelená barva) a LPIS (hnědá barva).

4.3 Statistické vyhodnocení

K vysvětlení distribuce kalouse ušatého byla tedy použita logistická regrese. Jako nezávisle proměnné byly uvažovány stejné vegetační prediktory odvozené ze všech geodatabází. Jednalo se o délku společné hranice mezi lesem a zemědělskou půdou v každém kvadrátu, dále rozloha lesa, rozloha zemědělské půdy, obvod lesních polygonů a obvod zemědělské půdy v každém kvadrátu. Seznam prediktorů je uvedený v *tabulce 2*. Jako závisle proměnná pak byl použit výskyt kalouse v daném kvadrátu (čili jeho prevalence či absence).

Pro statistické vyhodnocení byl použit program Excel 2003 a statistický software R 2.1.3.0.

V programu Excel byla z databázových tabulek *vyskyt.dbf* a všech tří výsledných sumarizačních tabulek z každé geodatabáze (tabulka s délkou společné hranice v kvadrátu, tabulka se součtem ploch a obvodů lesů v každém kvadrátu a poslední se sumou ploch a obvodů zemědělské půdy v kvadrátu) vytvořena pro CLC90, CLC00 i DMU_LPIS jedna nová tabulka obsahující sloupce:

- číslo (kód) kvadrátu;
- 5 sloupců s prediktory uvedenými v *tab. 2*
- výskyt druhu (označený buď 0 – absence nebo 1 – prevalence druhu).

Tabulky byly uloženy po názvy *clc90.txt*, *clc00.txt*, *dmu_lpis.txt*.

prediktor	
A	délka rozhraní lesa a zemědělské půdy
B	plocha lesa
C	obvod lesa
D	plocha zemědělské půdy
E	obvod zemědělské půdy

Tab. 2: Prediktory výskytu kalouse ušatého použité při tvorbě modelů.

Bylo vyhodnoceno celkem 15 modelů – pro každou geodatabázi bylo vytvořeno 5 modelů s jedním prediktorem A – E.

Do programu R byla postupně načtena textová tabulka každé databáze, ze které byla pro každý prediktor zvlášť spočtena logistická regrese příkazem *GLM* (s výskytem závislým pouze na jednom prediktoru a binomickou třídou). Následnými příkazy *SUMMARY (GLM)* a *CONFINT* byly získány odhady β_0' a β_1' parametrů β_0 a β_1 , jejich střední odchylka, významnost a interval spolehlivosti, v jakém se hodnota

parametru β' nachází s pravděpodobností 95%. Skript použitý pro načtení dat a výpočet je uvedený v Příloze 3.

Jak již bylo uvedeno výše, pokud by se parametr β_1 rovnal nule, příslušný faktor by neměl vliv na pravděpodobnost sledovaného jevu, protože by to znamenalo stejné šance pro obě hodnoty nezávisle proměnné. Proto je při známém odhadu β_1' parametru β_1 i jeho směrodatné odchylky σ_1' pro testování základní hypotézy $H_0: \beta_1' = 0$ vhodnou statistikou výraz

$$z = \frac{\beta_1'}{\sigma_1'}$$

který má při platné nulové hypotéze asymptoticky normální rozdělení $N(0,1)$ a porovnává odhad β_1 s jeho asymptotickou střední chybou. Toto testování se nazývá Waldův test (Zvára, 2008; Hendl, 2009) a je implementované ve funkci *SUMMARY* (*GLM*).

Jak uvádí Elith & kol. (2002) odhad intervalu spolehlivosti kolem predikovaných hodnot je jedním ze způsobů kvantifikování neurčitosti spojené s odhadem parametru ve výsledném modelu. Nejjednodušší a nejvíce používaný intervalový odhad pro namodelované hodnoty logistickou regresí je interval spolehlivosti Waldovy statistiky pro logit (v programu R získaný právě příkazem *CONFINT*).

Výsledný model získaný pomocí logistické regrese má tedy tvar

$$\ln\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i$$

Čili pravděpodobnost výskytu kalouse ušatého predikovaná konkrétním parametrem se získá:

$$P(Y_i = 1) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

Na základě této pravděpodobnostní funkce je možné dopočítat distribuční funkci $F(X)$, která představuje nejlepší popis pravděpodobnostního chování náhodné veličiny X . Distribuční funkce je pravděpodobnost, že náhodná veličina X bude nabývat určité hodnoty x nebo hodnoty menší $F(X) = P(X \leq x)$. Je definována pro všechna $x \in \mathbb{R}$ a leží v intervalu $(0,1)$ (Hendl, 2009).

Distribuční funkce pro každý prediktor byla vykreslena s pomocí programu MS Excel.

5. VÝSLEDKY

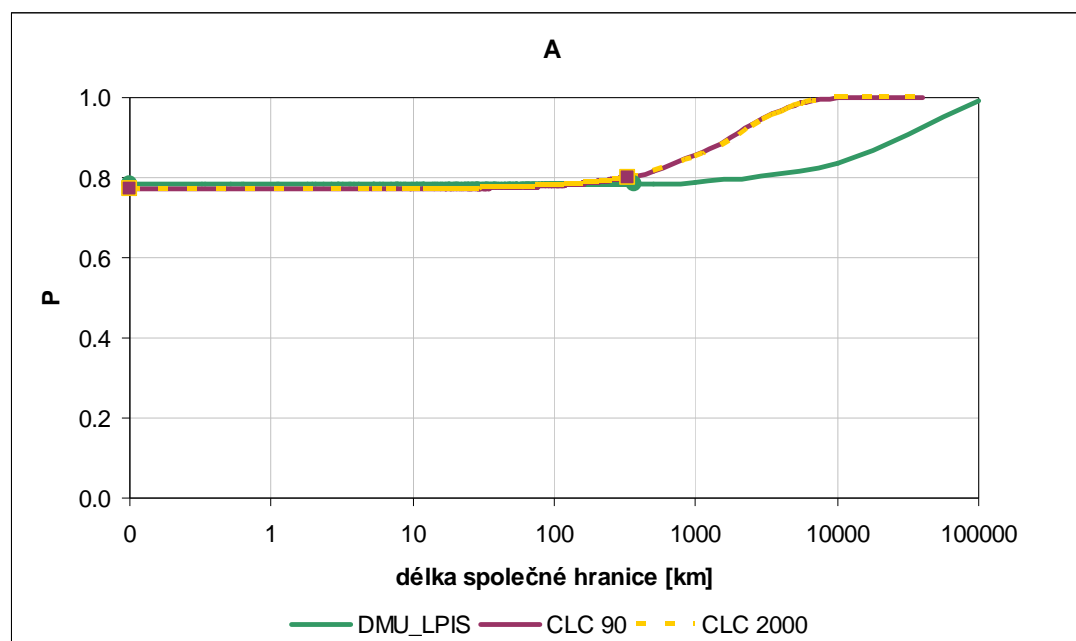
Podle výše uvedené metodiky bylo zpracováno celkem 15 modelů zahrnujících 5 prediktorů výskytu kalouse ušatého, získané postupně z třech různých map s rozdílným rozlišením: CLC 90 a CLC 2000 s rozlišením 1:100 000 a DMÚ 25 společně s LPIS s měřítkem 1:25 000 (respektive 1:10 000 u LPIS). Výsledná tabulka všech vypočítaných parametrů všech modelů je uvedena v Příloze 4.

5.1 Vizuální porovnání geodatabází

Pro každý prediktor A (délka společné hranice les/zemědělská půda), B (plocha lesů), C (obvod lesů), D (plocha zemědělské půdy) i E (obvod zemědělské půdy) byly vykresleny distribuční funkce porovnávající pravděpodobnost výskytu kalouse ušatého v závislosti na daném prediktoru, který byl získán z rozdílných geodatabází. Výsledky z obou variant geodatabáze CORINE Land Cover (CLC 90 i CLC 2000) jsou v některých kvadrátech shodné nebo se liší pouze nepatrně a v grafech se jejich křivky překrývají.

Skutečně vypočítaná data jsou na křivce ohraničena body, vně těchto bodů byla data extrapolována až k hodnotě pravděpodobnosti 1 (či 0 u křivky s klesajícím trendem – u prediktoru B a C).

U grafu A, C a E bylo pro lepší názornost použito logaritmické měřítko.



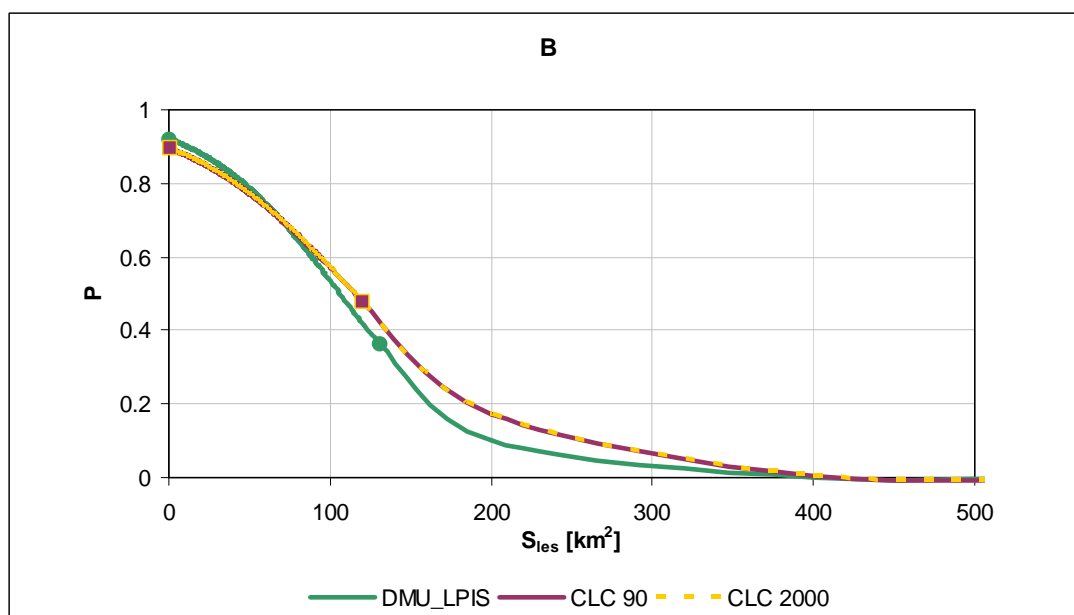
Obr. 8: Pravděpodobnost výskytu kalouse ušatého v závislosti na délce společného rozhraní lesní a zemědělské půdy v jednotlivých kvadrátech.

Pravděpodobnost výskytu kalouse podle délky společné hranice lesa a zemědělské půdy získané z CORINE 90 i 2000 se pro všechna naměřená data (od 0 km do 335 km délky na kvadrát) pohybuje v rozmezí 77-80%. Velice podobné hodnoty byly naměřeny i z DMU+LPIS, z kterých vyšly délky společných hranic v rozmezí od 0 do 371 km a pravděpodobnosti výskytu pro všechny hodnoty prediktoru se pohybují v rozmezí 78-79%. Už z toho je patrné, že pravděpodobnost není téměř ovlivněna hodnotou prediktoru A (viz obr. 8).

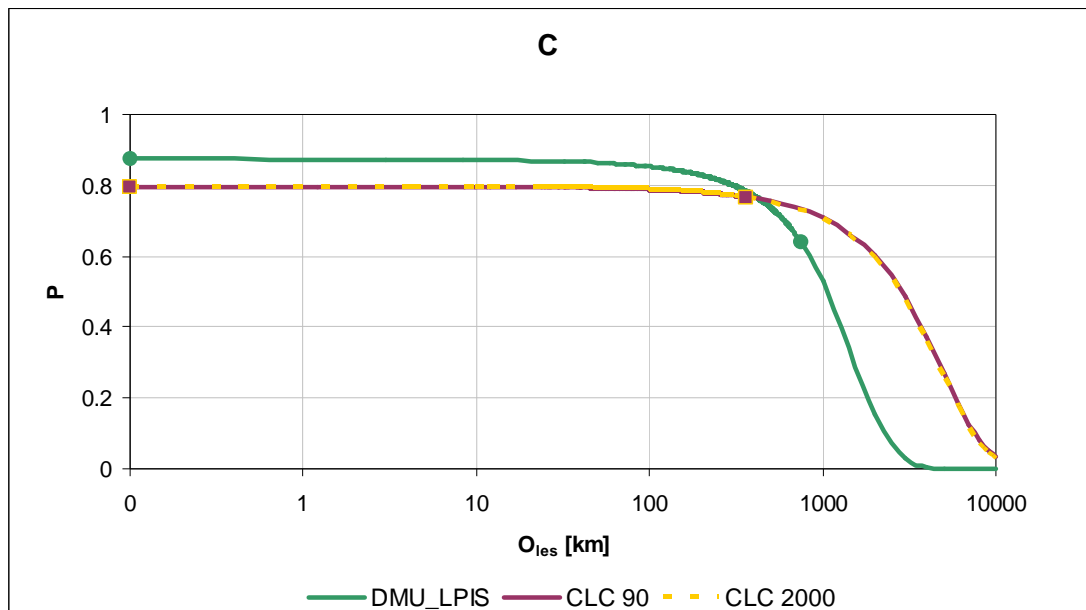
U modelů s prediktory týkajících se pouze lesů vyšla pravděpodobnost výskytu nepřímou úměrnou k rostoucí ploše (prediktor B – viz obr. 9) i obvodu (prediktor C – viz obr. 10) lesních polygonů v kvadrátu.

V modelech s prediktorem B jsou u všech geodatabází pravděpodobnosti výskytu kolem 90% při nulové rozloze lesů a mají velmi podobný průběh. Při rozloze okolo 120 km² je již pravděpodobnost výskytu necelých 48% u CLC, respektive 42% u DMÚ, a těchto 6% je i největší pravděpodobnostní rozdíl u naměřených dat.

Prediktor C (obvod lesů) získaný z DMÚ je oproti CORINE v každém kvadrátu přibližně dvojnásobný. Podrobnější měřítko DMÚ tak zřejmě lépe vystihuje členitý okraj lesů, který je v CLC více zjednodušen, čímž se skutečný obvod zmenšuje. Maximální délka v kvadrátu byla 745 km (DMÚ), při níž se druh vyskytuje s 64% pravděpodobností oproti polovičním 363 km z CLC, kde je pravděpodobnost kolem 77%. Při nulové hodnotě prediktoru je pravděpodobnost výskytu kolem 87% (DMÚ) a 80% (CLC).

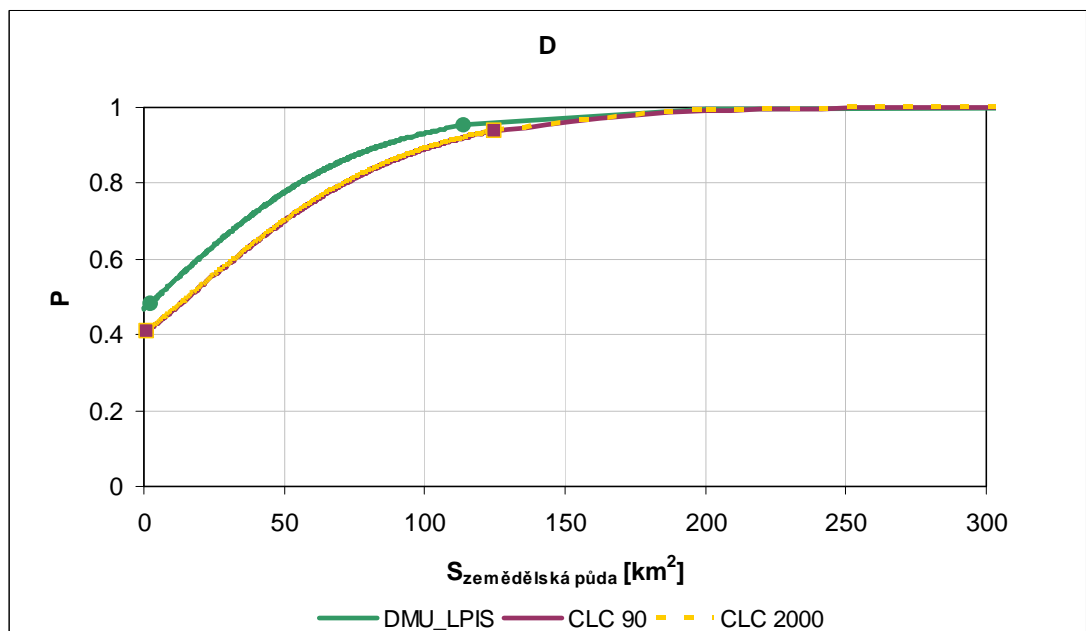


Obr. 9: Pravděpodobnost výskytu kalouse ušatého v závislosti na celkové rozloze lesních polygonů v jednotlivých kvadrátech.



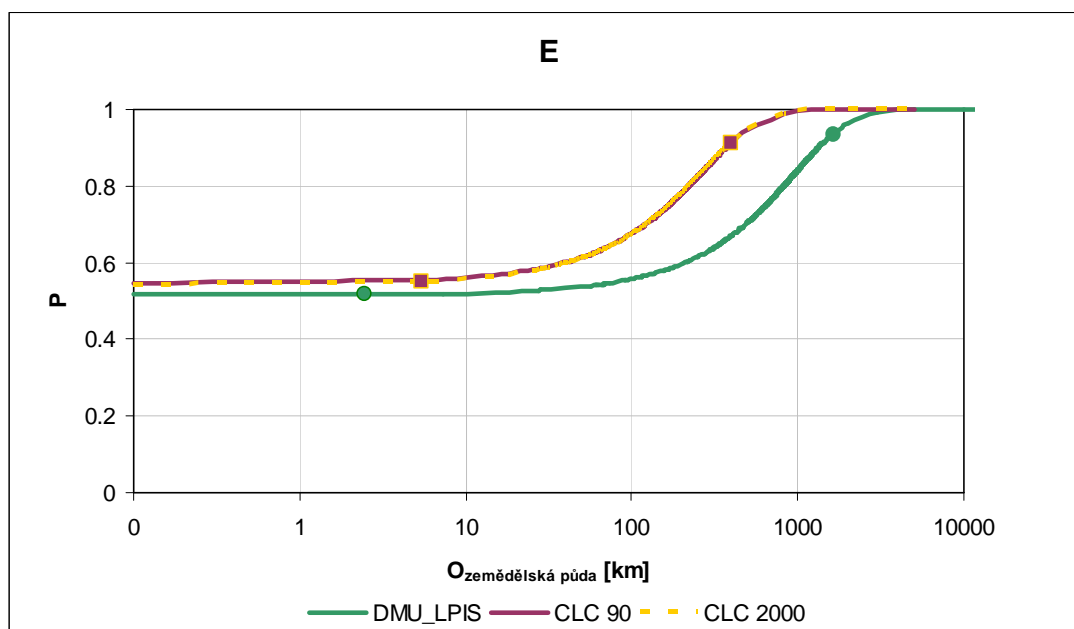
Obr. 10: Pravděpodobnost výskytu kalouse ušatého v závislosti na celkovém obvodu lesních polygonů v jednotlivých kvadrátech.

V modelech s prediktorem D byla zkoumána závislost výskytu kalouse na ploše zemědělské půdy v kvadrátu (viz obr. 11). Hodnoty prediktorů i pravděpodobnosti získané z LPIS i CORINE se příliš neliší, největší rozdíl je asi 7% při rozloze 50 km².



Obr. 11: Pravděpodobnost výskytu kalouse ušatého v závislosti na celkové rozloze zemědělské půdy v jednotlivých kvadrátech.

Hodnoty obvodů zemědělské půdy byly v LPIS v rozmezí od 2 do 1750 km, zatímco u CLC od 5 do 399 km, tudíž u LPIS byly i více než 4-krát vyšší, díky čemuž jsou u tohoto prediktoru i největší pravděpodobnostní rozdíly ze všech. Při obvodu 390 km je rozdíl 25% (viz obr. 12).



Obr. 12: Pravděpodobnost výskytu kalouse ušatého v závislosti na celkovém obvodu zemědělské půdy v jednotlivých kvadrátech.

5.2 Statistická významnost odhadnutých parametrů

Kromě hodnot parametru logistické regrese β_1 byla Waldovým testem vypočítána statistická významnost tohoto parametru, neboli pravděpodobnost zamítnutí nulové hypotézy, která platí. Zamítnutí platné nulové hypotézy se nazývá chyba prvního druhu (α) a pravděpodobnost chyby prvního druhu je obecně stanovena na maximální velikost $P(\alpha) = 0,05 = 5\%$ (Soukup, 2010). Vypočítaná hladina významnosti (v R uvedená ve sloupci $Pr(>/z/)$) musí být tedy menší než stanovená hladina významnosti α , aby bylo možné hypotézu H_0 zamítnout. Čím nižší je vypočítaná hladina významnosti, tím vyšší je statistická významnost (Soukup, 2010). V následující tabulce (tab. 3) jsou znázorněny odhady parametrů jednotlivých geodatabází spolu s hladinou významnosti α , s kterou je zamítnuta nulová hypotéza, že daný prediktor nemá vliv na pravděpodobnost výskytu kalouse ušatého.

	DMU_LPIS			CLC 90			CLC 2000		
	odhad β_1	vypočtená význam.	α	odhad β_1	vypočtená význam.	α	odhad β_1	vypočtená význam.	α
A	0,0000334	0,975		0,0005572	0,701		0,0005292	0,716	
B	-0,0231670	2,33E-09	***	-0,0187020	4,26E-06	***	-0,0186660	4,19E-06	***
C	-0,0018236	0,00743	**	-0,0004740	0,728		-0,0004871	0,721	
D	0,0274700	1,17E-11	***	0,0250210	1,87E-12	***	0,0249830	2,11E-12	***
E	0,0016115	2,28E-07	***	0,0054540	0,000119	***	0,0055320	9,93E-05	***

Tab. 3: Odhady parametru β_1 získané logistickou regresí pro každou geodatabázi a každý prediktor A – E. Hladina významnosti je označena hvězdičkami, *** značí $\alpha = 0,001$; ** značí $\alpha = 0,01$; žádná hvězdička znamená $\alpha > 0,1$.

Pro prediktor A, čili společnou hranici lesní a zemědělské půdy, je vypočítaná statistická významnost velice vysoká (přes 0,7; u DMU+LPIS dokonce 0.975), takže v modelech s tímto prediktorem nelze zamítnout hypotézu $H_0: \beta_1 = 0$. V modelu s prediktorem C (obvodem lesních polygonů) se parametr β_1 jako statisticky významný jeví pouze pro kombinaci dat z DMÚ25 a LPIS (respektive z DMÚ, protože se jedná o les). Pro ostatní prediktory – plochu lesů a plochu a obvod zemědělské půdy můžeme podle β_1 nulovou hypotézu zamítnout na hladině významnosti 0,1%.

5.3 Porovnání rozsahu intervalů spolehlivosti a jejich překryvů

Intervaly spolehlivosti uvedené v tabulce 4 zahrnují s pravděpodobností 95% hodnotu odhadovaného parametru β_1 .

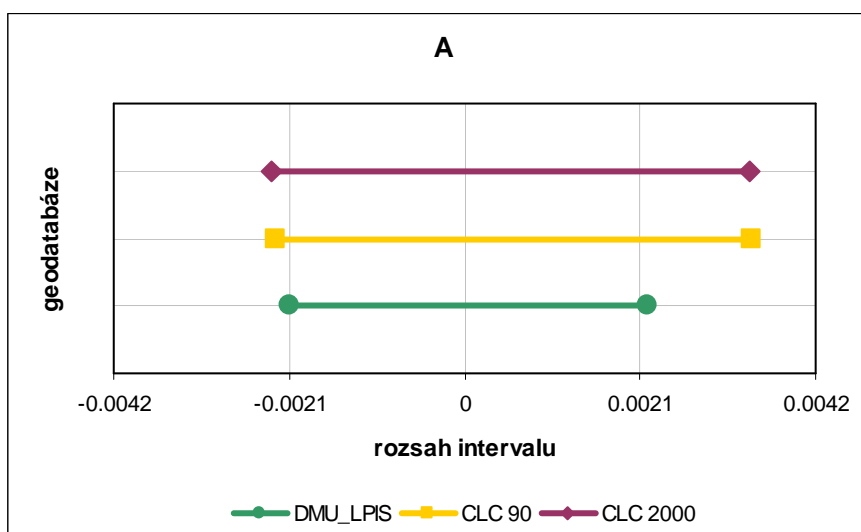
	DMU_LPIS			CLC 90			CLC 2000		
	interval spolehlivosti		rozsah	interval spolehlivosti		rozsah	interval spolehlivosti		rozsah
A	-0,0021	0,0022	0,0043	-0,0023	0,0034	0,0057	-0,0023	0,0034	0,0057
B	-0,0309	-0,0157	0,0152	-0,0268	-0,0108	0,0160	-0,0267	-0,0108	0,0159
C	-0,0032	-0,0005	0,0027	-0,0031	0,0022	0,0054	-0,0032	0,0022	0,0054
D	0,0196	0,0355	0,0159	0,0182	0,0321	0,0139	0,0181	0,0321	0,0140
E	0,0010	0,0022	0,0012	0,0027	0,0083	0,0056	0,0028	0,0083	0,0056

Tab. 4: Vypočítané intervaly spolehlivosti pro parametr β_1 jednotlivých modelů.

V intervalu spolehlivosti je skryta i informace o statistické významnosti (pokud je v intervalu obsažena hodnota nulové hypotézy – zde nula, není regresní koeficient statisticky významný (Soukup, 2010)).

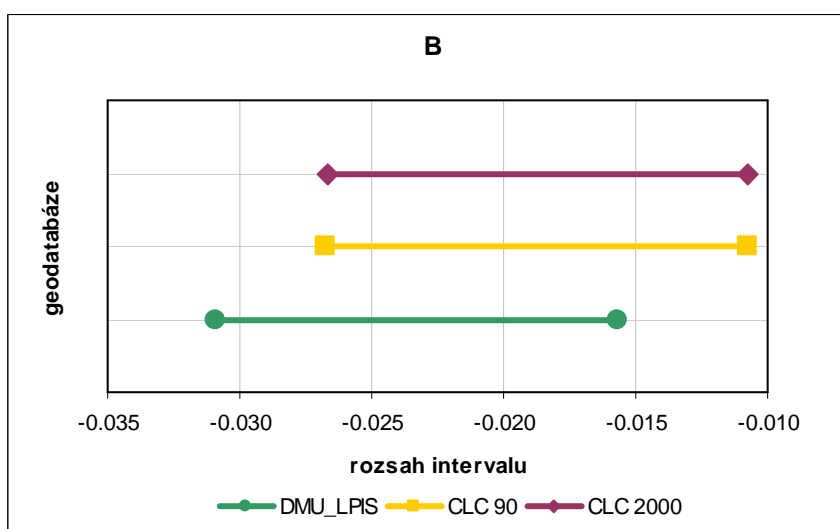
Překryv intervalů spolehlivosti byl graficky znázorněn (viz obr. 13 -17) pro srovnání rozdílů mezi geodatabázemi. Čím více se intervaly překrývají, tím menší jsou rozdíly mezi jednotlivými modely. Rozsah intervalu zase značí přesnost - čím užší interval, tím by měl být model přesnější.

Jak je patrné z obr. 13, intervaly spolehlivosti jednotlivých bází se velmi podobají, přičemž celý interval DMU+LPIS je obsažený v intervalu CORINE. Navíc všechny tři intervaly obsahují nulovou hodnotu, z čehož je patrné, že modely s prediktorem A nemají statisticky významné koeficienty regrese.



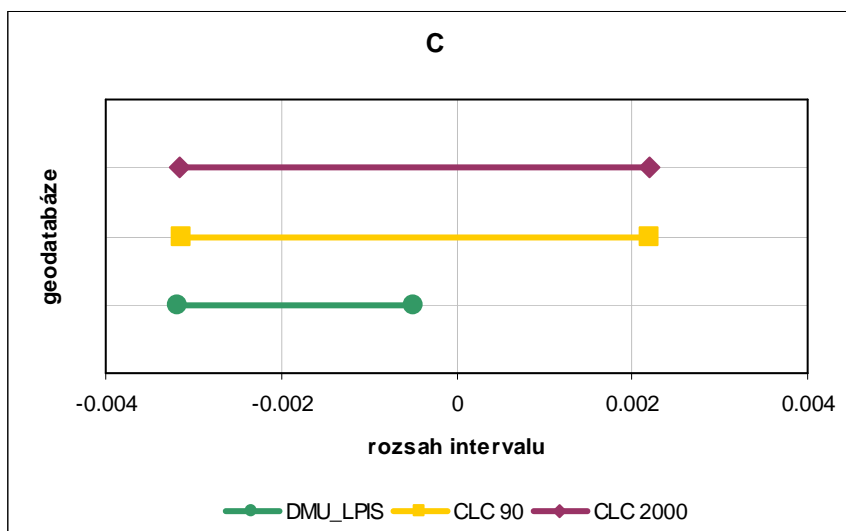
Obr. 13: Znázornění intervalů spolehlivosti pro modely s prediktorem A

Intervaly spolehlivosti pro modely s prediktorem B (plocha lesů) se z velké části překrývají a mají přibližně stejný rozsah (viz obr. 14).



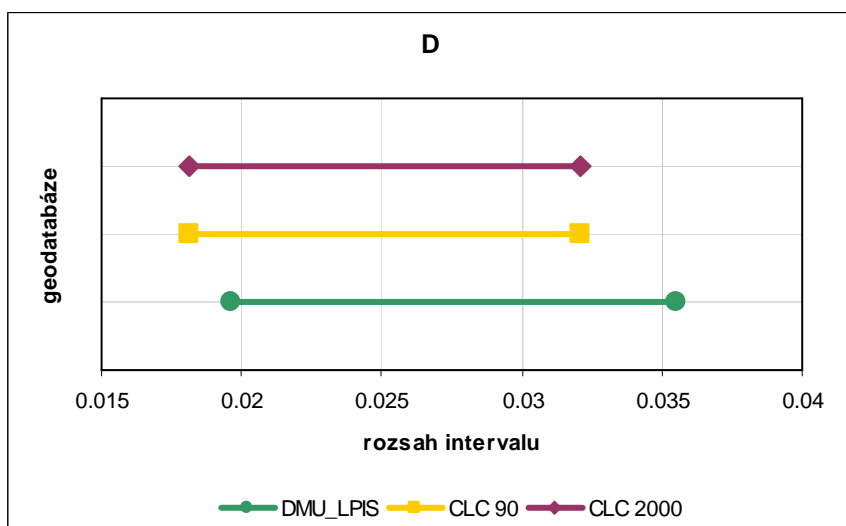
Obr. 14: Znázornění intervalů spolehlivosti pro modely s prediktorem B

U modelů s parametrem C (obvod lesů) vyšel statisticky významný regresní koeficient β_1 pouze z geodatabáze DMÚ, což je názorně vidět z *obr. 15*, kdy interval spolehlivosti pro DMU_LPIS jako jediný neobsahuje nulovou hodnotu. Navíc je jeho rozsah zhruba poloviční, což značí přesnější model.



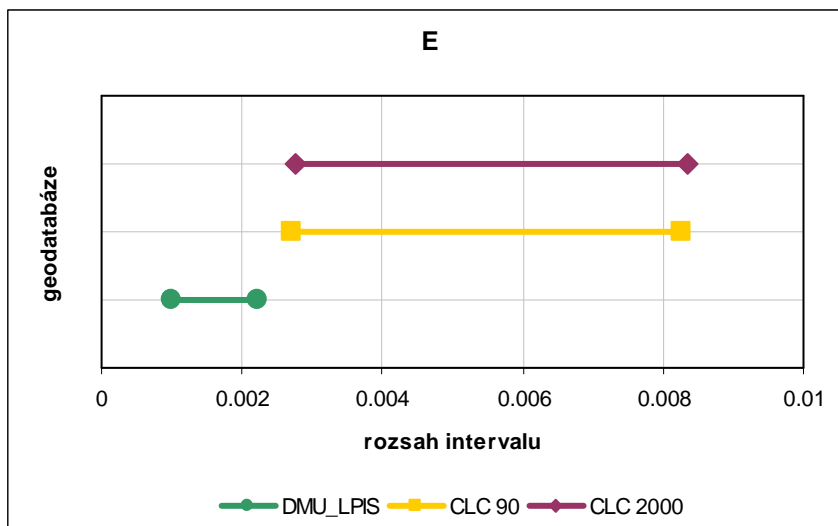
Obr. 15: Znárodnění intervalů spolehlivosti pro modely s prediktorem C

Modely s prediktorem plochy zemědělské půdy (D) nevykazují velké rozdíly, jejich intervaly spolehlivosti se z velké části překrývají a mají téměř stejný rozsah (viz *obr. 16*).



Obr. 16: Znárodnění intervalů spolehlivosti pro modely s prediktorem D

Největší rozdíly jsou i podle intervalů spolehlivosti patrné pro modely s prediktorem E (obvod zemědělské půdy), kdy se interval DMU_LPIS výrazně liší svým rozsahem (je zhruba čtvrtinový v porovnání s CORINE). Navíc nemá s geodatabázemi CLC ani částečný překryv (viz obr. 17). Dle šíře intervalu by tedy měl být model DMÚ/LPIS považován za vhodnější.



Obr. 17: Znárodnění intervalů spolehlivosti pro modely s prediktorem E

Důvodem pro použití pouze základních modelů s jedním prediktorem výskytu byla častá korelace mezi jednotlivými veličinami. Mezi každými dvěma prediktory byl spočítán korelační koeficient r_{xy} (jeho hodnoty jsou uvedeny v tabulce 5). Kritická hodnota korelačního koeficientu pro 628 měření (hodnoty z 628 kvadrátů) je na hladině významnosti 0,1% rovna $r_{628} = 0,13101$. Pokud je tedy hodnota koeficientu $|r_{xy}| > r_{628}$, je zamítnuta nulová hypotéza o nezávislosti veličiny a prediktory jsou považovány za vzájemně korelované.

	B (plocha lesa)		C (obvod lesa)		D (plocha zem.púd.)		E (obvod zem.púd.)	
	CLC90	DMU/LPIS	CLC90	DMU/LPIS	CLC90	DMU/LPIS	CLC90	DMU/LPIS
A	0,41434	0,09814	0,97035	0,50564	0,10761	0,17424	0,87540	0,48062
B			0,41903	0,61189	-0,56387	-0,68791	0,14662	-0,43031
C					-0,03240	-0,35152	0,82557	0,11075
D							0,52306	0,76433

Tab. 5: Vypočítané hodnoty korelačního koeficientu pro různé kombinace dvou prediktorů. Tučně vyznačené hodnoty jsou menší než hodnota r_{628} .

6. DISKUZE

Ačkoli práce vychází z atlasových údajů, které jsou jedny z nejlepších v Evropě, druhová data jsou díky kvadrátům o ploše 133 km² možná příliš hrubého měřítka. Pro environmentální data dostupná pro pokus by bylo vhodnější použití podrobnějších druhových dat, protože zahrnuté atlasové výskyty jsou ze všech použitých dat nejhrubšího rozlišení. Podrobnější druhová data pro území celé republiky však nejsou k dispozici.

Původním záměrem proto bylo naopak použít ještě hrubší měřítko prediktorových proměnných, jako je např. v geodatabázi ARC ČR 500 (1:500 000), ovšem to nakonec nebylo vzhledem ke zvoleným prediktorům možné, protože pro tuto práci nebyla v takovém měřítku k dispozici data zemědělské půdy (v ARC ČR je dostupná pouze vrstva lesů).

Použití obou geodatabází CORINE bylo v tomto pokusu celkem zbytečné. Jejich měřítka se neliší a změny kvůli aktualizaci mezi lety 1990 a 2000 téměř neovlivnily hodnoty odvozených prediktorů (nebo jen velmi nepatrně) a výsledky logistické regrese i hodnoty regresních koeficientů jsou skoro totožné. Proto je celkové srovnání výsledků prakticky jen mezi dvěma typy dat – CORINE a DMÚ+LPIS. Do jednotlivých výsledků se ovšem nepromítne jen rozdílné měřítka, všechny geodatabáze vznikly za jiným účelem a liší se svojí strukturou. Například výběr tříd z CLC do oblastí pouze lovu či pouze hnízdění je subjektivní a nemusí být odpovídající. Často může docházet k tomu, že kalous ušatý v uvedené třídě zároveň hnízdí i loví, jelikož zde má vhodné podmínky pro obě tyto činnosti (vyskytují se zde křoviska či skupiny stromů vhodné k zahnízdění a zároveň travnaté či zemědělské plochy obývané hrabošem - např. louky zahrnují rozptýlené keře a stromy až na 20% své rozlohy (Bossard & kol., 2000)). K tomu samozřejmě může částečně docházet i v případě LPIS.

Původně hlavní uvažovaný prediktor výskytu kalouse ušatého stanovený dle popisu jeho prostředí od Hagemeiera & Blaira (1997), tedy celková délka hranice mezi lesem a zemědělskou půdou, se v pokusu jeví jako absolutně statisticky neprůkazný. Při jeho výpočtech především z DMÚ a LPIS, kdy se počítal pomocí polovičního obvodu vytvořených bufferů kolem linií polygonů, mohlo dojít ke zkreslení hodnot, avšak oproti datům z CLC se hodnoty příliš nelišily. Proto se tento prediktor jeví jako zcela nevhodný pro predikci výskytu kalouse.

Pro všechna geodata je ještě velice podobných výsledků dosaženo u prediktorů B a D, tedy u plochy lesa a zemědělské půdy. Tyto veličiny jsou

navzájem záporně korelované - s přibývajícím lesem ubývá plocha zemědělské půdy a naopak. Rozdíly pravděpodobností výskytu mezi geodatabázemi jsou nejvýše asi 6-7%, intervaly spolehlivosti mají u obou prediktorů podobný rozsah i překryv mezi jednotlivými geodaty.

Největší rozdíly v dosažených výsledcích v závislosti na použitých geodatech jsou viditelné u prediktorů C a E. U obvodů lesních polygonů (pred. C) podrobnější měřítko DMÚ pravděpodobně lépe vystihuje skutečnou hranici lesů a dochází tak k menšímu zkreslení než u CLC. Také pouze s použitím dat z DMÚ je tento prediktor statisticky průkazný. Ovšem u obvodu zemědělské půdy v kvadrátu (pred. E) může být čtyřnásobná hodnota nezávisle proměnné získaná z LPIS (oproti hodnotám z CORINE) zapříčiněna nejen lepším vystižením skutečné hranice zemědělské půdy, ale i strukturou dat LPIS, ze které nebylo možné při výpočtech odstranit všechny vnitřní polygonové hranice nacházející se uvnitř větších ploch zemědělské půdy, protože sousedící farmářské bloky nemají vždy společnou polygonovou hranici, a proto jich velké množství zůstalo neodstraněno. Tím se výrazně zvýšila hodnota obvodu. Na druhou stranu možná právě díky tomu může LPIS lépe předpovídat výskyt druhu – sousedící polygony zemědělské půdy, které nemají společnou polygonovou hranici, jsou od sebe pravděpodobně ve skutečnosti odděleny např. remízky, keřovými pásy, křovisky kolem potoků, atd. Všechna tato vegetace bývá využívána kalousem k hnízdění a čím více jí je mezi zemědělskou půdou, tím se zvyšuje šance na výskyt druhu. U CLC však tato vegetace kolem polí není díky hrubšímu rozlišení zaznamenána. Dle rozsahu intervalu spolehlivosti by měla být považována za nevhodnější právě data z LPIS, avšak i pro CLC jsou regresní koeficienty statisticky průkazné na stejné hladině významnosti.

Elith & kol. (2002) i Seoanne & kol. (2004a) sice říkají, že data nasbíraná v relativně hrubém prostorovém měřítku mohou mít nedostatek adekvátních informací relevantních pro ptačí druhy, ale protože v tomto pokusu nejsou jako prediktory využívány žádné vzácné vegetační typy, i obecnější mapa vegetačního krytu CORINE Land Cover se zdá být dostačující pro zaznamenání stanovištních charakteristik důležitých pro výskyt kalouse ušatého. Podobných výsledků se pro několik jiných ptačích druhů dobrali např. i Seoanne & kol. (2004a) či Sklenářová (2011). Mack & kol. (1997) taktéž zhodnocují, že data hrubších měřítek ze satelitního snímkování jsou dostatečné pro hrubší odhady vztahů mezi druhem a jeho prostředím. Sklenářová (2011) s využitím stejného síťového mapování na území ČR uvádí, že vzhledem k podrobnosti a kvalitě vysvětlovaných dat mohou být k modelování použity databáze v měřítkách 1:100 000 (CORINE Land Cover)

i 1:10 000 (DIBAVOD - Digitální báze vodohospodářských dat) s dosažením takřka stejných výsledků. Avšak geodatabáze s měřítkem 1:500 000 (ARC ČR) se již jevila jako nedostatečně přesná pro tvorbu modelů distribuce ptačích druhů.

Přesnost modelů může být samozřejmě ovlivněna i chybami v druhových datech. Do modelů byla zahrnuta hnízdění všech stupňů průkaznosti. Avšak ze všech uvedených výskytů je pouze 63% prokázané hnízdění, 17% hnízdění je pravděpodobné a 20% možné (Šťastný & kol., 2006). Pro další zhodnocení přesnosti vytvořených modelů by bylo možné sestavit všechny modely znovu za použití pouze prokázaného hnízdění.

Ideální způsob pro zjištění správnosti modelů a jejich schopnosti přesné predikce by byla validace (Regan & kol., 2002). Na tu by ovšem bylo potřeba získat další data síťového mapování a environmentálních prediktorů pro jiné území.

7. ZÁVĚR

Na základě dosažených výsledků bylo zjištěno, že všechny použité geodatabáze produkují obdobně přesné modely. Nepatrně lepšího výsledku bylo dosaženo s kombinací dat DMÚ a LPIS, kde je oproti databázi CLC statisticky průkazný i prediktor obvod lesů a o trochu přesnější model s prediktorem obvodu zemědělské půdy.

Pro takovouto studii se však jeví jako dostačující geodata CORINE Land Cover v měřítku 1:100 000, protože získání dat podrobnějších měřítek je, vzhledem k uspokojivým výsledkům z CORINE, zbytečně ekonomicky náročné.

Pro modelování výskytu kalouse ušatého se pak jeví jako nejvhodnější prediktor celková plocha zemědělské půdy v kvadrátu, případně celková plocha lesa v kvadrátu. Tyto prediktory jsou zároveň nejméně ovlivněny měřítkem použitých geodat. Měřítko mapy nejvíce ovlivňuje hodnoty prediktorů obvodu lesů a zemědělské půdy pravděpodobně díky skutečné členitosti obvodové linie, která je u hrubších dat trochu zkreslena.

Tuto metodiku je možné použít pro tvorbu modelů distribuce jiných ptačích druhů v závislosti na jejich prediktorech výskytu. Není však možné závěry této práce o dostatečně vhodné geodatabázi zobecnit pro všechny ptačí druhy.

8. POUŽITÁ LITERATURA

- ANDĚL J., 2011: Základy matematické statistiky. Matfyzpress, Praha. 358 s.
- ANDĚRA M., 2005: Mapování výskytu savců v ČR. Online: <http://www.biolib.cz/cz/article/id1/>, cit: 18.2.2013.
- ARAÚJO M. B. & GUISAN A., 2006: Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33: 1677-1688.
- BARRY S. & ELITH J., 2006. Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43 (3): 413–423.
- BEJČEK V. & ŠŤASTNÝ K., 1999: Encyklopedie ptáků. Rebo production, Čestlice. 288 s.
- BĚLINA A., 2010: Kalous ušatý. Online: http://fotoales.cz/lexikon_ptaku.php, cit: 11.4.2013.
- BOHDAL J., 1984: Kalous ušatý. Online: <http://www.naturfoto.cz/kalous-usaty-fotografie-147.html>, cit: 11.4.2013.
- BOKR J., 1993: Selektivní informace, porucha a neurčitost. ZČU, Plzeň. 19 s.
- BOKR J. & JÁNEŠ V., 2005: Neurčitost v popisu technologických procesů. *Automatizace* 48 (1): 20 – 24.
- BOSSARD M., FERANEC J. & OTAHEL J., 2000: CORINE Land Cover technical guide – Addendum 2000. Technical report no 40. European Environment Agency. 105 s.
- BROTONS L., THUILLER W., ARAÚJO M. B. & HIRZEL A. H., 2004: Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27: 437-448.
- BŘOUŠEK L., LAŽA L., 2006: 55 let vojenské geografie v Dobrušce. Vojenský geografický a hydrometeorologický úřad, Dobruška. 24 s.
- BÜTTNER G., FERANEC J. & JAFFRAIN G., 2002: CORINE Land Cover update 2000. Technical guidelines. European Environment Agency, Copenhagen. 56 s.
- CENIA, 2012: CORINE Land Cover 2006. Online: <http://www1.cenia.cz/www/projekty-cenia/ukoncene-projekty>, cit.: 19.10.2012.
- COMMISSION OF EUROPEAN COMMUNITIES, 1994: CORINE Land Cover. European Environment Agency, 163 s.
- ČÁBELKA M., 2008: Úvod do GPS. CITT Praha-Akademie kosmických technologií. 74 s.

- DORMANN C. F., PURSCHKE O., GARCIA MARQUEZ J. R., LAUTENBACH S. & SCHRÖDER B., 2008: Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology* 89 (12): 3371 – 3386.
- DUCKHAM M., 2002: Uncertainty and geographic information: computational and critical convergence. Department of Computer Science, University of Keele, Staffordshire. 13 s.
- DUDEK I., 1999: Poznání a neurčitost, základní racionální přístupy a praktické metody. *Electronic journal for philosophy* 99.
- ELITH J., BURGMAN M.A. & REGAN H.M., 2002: Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* 157: 313–329.
- ELITH J. & LEATWICK J. R., 2009: Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution and Systematics* 40: 677-697.
- FEELEY K. J. & SILMAN M. R., 2010: Modelling the responses of Andean and Amazonian plant species to climate change: the effect of georeferencing errors and the importance of data filtering. *Journal of Biogeography* 37: 733-740.
- FERNANDEZ M. A., BLUM S. D., REICHLER S., GUO Q., HOLZMAN B. & HAMILTON H., 2009: Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodiversity Informatics* 6: 36-52.
- FRAIR J. L., NIELSEN S. E., MERRILL E. H., LELE S. R., BOYCE M. S., MUNRO R. H. M., STENHOUSE G. B. & BEYER H.L., 2004: Removing GPS collar bias in habitat selection studies. *Journal of Applied Ecology* 41: 201 – 212.
- GORMLEY A. M., FORSYTH D. M., GRIFFIOEN P., LINDEMAN M., RAMSEY D. S. L., SCROGGIE M. P., WOODFORD L., 2011: Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology* 48: 25-34.
- GUISAN A. & THUILLER W., 2005: Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009.
- HAGEMEIER W.J.M. & BLAIR M.J., 1997 (eds): *The EBCC Atlas of European Breeding Birds: Their Distribution and Abundance*. Poyser, London. 908 s.
- HAMER K. C. & HILL J.K., 2000: Scale – dependent effects of habitat disturbance on species richness in tropical forest. *Conservation Biology* 14 (5): 1435 – 1440.

- HENDL J., 2009: Přehled statistických metod. Analýza a metaanalýza dat. Portál, Praha. 696 s.
- HERNANDEZ P. A., GRAHAM C. H., MASTER L. L. & ALBERT D. L., 2006: The effect of sample size and species characteristics on performance of species distribution modeling methods. *Ecography* 29: 773 – 785.
- JAGER H. I. & KING A.W., 2004: Spatial uncertainty and ecological models. *Ecosystems* 7 (8): 841 – 847.
- JANITOR, 2003: Zpráva o analýze vrstev kvadrátů síťového mapování. Bulletin no. 2/2013.
- JETZ W., McPHERSON J.M. & GURALNICK R.P., 2012: Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology and Evolution* 27 (3): 151-159.
- KLIR G. J., 2008: Patří neurčitost do vědy? Z nutného zla užitečná vlastnost. *Vesmír* 87 (10): 696 - 699.
- KOŽELUH M., 2003: Definice tříd CLC. Zpracováno dle BOSSARD & kol., 2000: CORINE Land Cover technical guide – Addendum 2000. 71 s.
- LANGR J., 2001: Geografická data. T-Mapy. Online: http://www.tmapy.cz/public/tmapy/cz/aktualne/Propagacni_materialy/clanky/clanky_old/geograficka_data.html, cit.:10.9.2012
- MACK E. L., FIRBANK L. G., BELLAMY P. E., HINSLEY S. A. & VEITCH N., 1997: The comparison of remotely sensed and ground-based habitat area data using species-area models. *Journal of Applied Ecology* 34: 1222-1228.
- MARTÍNEK D., 2007: Neurčitost, heterogenní M&S, bezpečnost. Online: <http://perchta.fit.vutbr.cz/DavidMartinek/6>, cit. 20.3.2012
- MARTIŠKO J., 1999: Ochrana dravců a sov v zemědělsky využívané krajině. EkoCentrum Brno, Brno. 226 s.
- McPHERSON J. M., JEZT W. & ROGERS D. J., 2004: The effect of species' range size on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41: 811-823.
- McPHERSON J. M., JEZT W. & ROGERS D. J., 2006: Using coarse-grained occurrence data to predict species distributions at finer spatial resolution – possibilities and limitations. *Ecological Modelling* 192: 499-522.
- MINISTERSTVO ZEMĚDĚLSTVÍ, 2011: Registr půdy – LPIS. Online: <http://eagri.cz/public/web/mze/farmar/LPIS/>, cit.: 16.2.2013
- MLÍKOVSKÝ J., 1998: Potravní ekologie našich dravců a sov. Metodika ČSOP č. 11. ZO ČSOP, Vlašim. 103 s.

- MOUDRÝ V. & ŠÍMOVÁ P., 2012: Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science*. Vol. 26, No 11: 2083 – 2095.
- OJA T., ALAMETS K. & PÄRNAMETS H., 2005: Modelling bird habitat suitability based on landscape parameters at different scales. *Ecological Indicators* 5: 314 – 321.
- PEARCE J. L. & BOYCE M. S., 2006: Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* 43: 405-412.
- PHILLIPS D. L. & MARKS D. G., 1996: Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. *Ecological Modelling* 91: 213-229.
- REGAN H. M., COLYVAN M. & BURGMAN M. A., 2002: A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12 (2): 618 – 628.
- RONDININI C., WILSON K. A., BOITANI L., GRANTHAM H. & POSSINGHAM P., 2006: Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters* 9: 1136 – 1145.
- ROUBICEK A. J., VANDERWAL J., BEAUMONT L.J., PITMAN A. J., WILSON P. & HUGHES L., 2010: Does the choice of climate baseline matter in ecological niche modelling? *Ecological Modelling* 221: 2280–2286.
- RÖSNER S., 2005: Waldohreule in freier Wildbahn. Online: http://commons.wikimedia.org/wiki/File:Waldohreule_in_freier_Wildbahn.jpg, cit: 22.4.2013
- RYCHNOVSKÝ M, 2008: Postupná výstavba modelů ohodnocení kreditního rizika. Bakalářská práce, Katedra pravděpodobnosti a matematické statistiky, MFF UK. 46 s.
- ŘEHÁKOVÁ B., 2000: Nebojte se logistické regrese. *Sociologický časopis*, vol.36, no. 4: 475 – 492.
- SAND H., ZIMMERMANN B., WABAKKEN P., ANDREN H. & PEDERSEN H. C., 2005: Using GPS technology and GIS cluster analysis to estimate kill rates in wolf – ungulate ecosystems. *Wildlife Society Bulletin* 33 (3): 914 – 925
- SEOANE J., BUSTAMANTE J. & DÍAZ-DELGADO R, 2004a: Are existing vegetation maps adequate to predict bird distribution? *Ecological Modelling* 175: 137 – 149.
- SEOANE J., BUSTAMANTE J. & DÍAZ-DELGADO R, 2004b: Competing roles for landscape, vegetation, topography and climate in predictive models of bird distribution. *Ecological Modelling* 171: 209 - 222.

- SHAO M. & LIU N., 2006: The diet of the Long-eared Owls, *Asio otus*, in the desert of northwest China. *Journal of Arid Environments* 65: 673-676.
- SITEWELL s.r.o., 2004: SITEWELL LPIS 2, implementace systému. Případová studie projektu. 34 s.
- SKLENÁŘOVÁ L., 2011: Spolehlivost použitých geodat pro fitování biotopových efektů na prostorovou distribuci vybraných ptačích druhů. Diplomová práce, ČZU. 59 s.
- SOUKUP P., 2010: Nesprávná užívání statistické významnosti a jejich možná řešení. *Data a výzkum – SDA Info* 2010 4 (2): 77-104.
- ŠOLTÉSOVÁ J., 2007: Kalous ušatý – *Asio otus*. Online: <http://www.priroda.cz/lexikon.php?detail=854>, cit. 28.2.2012
- ŠŤASTNÝ K., BEJČEK V. & HUDEC K., 2006: Atlas hnízdního rozšíření ptáků v České republice. Adventum, Praha. 463 s.
- TICHÝ B., 2012: Datové standardy VTOPÚ Dobruška. Online: http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_1998/Sbornik/Tichy/stand98.htm, cit. 26.2.2012
- VENIER L. A., PEARCE J., Mc. KEE J., MCKENNEY D.W. & NIEMI G.J., 2004: Climate and satellite – derived land cover for predicting breeding bird distribution in the Great Lake Basin. *Journal of Biogeography* (31): 315 – 331.
- WARD G., 2007: Statistics in ecological modeling; presence-only data and boosted mars. A dissertation, Stanford University. 128 s.
- ZÁVALSKÝ O., 2004: Naši dravci a sovy a jejich praktická ochrana. *Metodika ČSOP* č. 29. ZO ČSOP, Nový Jičín. 84 s.
- ZIMMERMANN N. E., EDWARDS Jr. T. C., GRAHAM C. H., PEARMAN P. B. & SVENNING J. CH., 2010: New trends in species distribution modelling. *Ecography* 33: 985 – 989.

9. PŘÍLOHY

Příloha 1

Mezinárodní kódy pro zápis terénních pozorování (Šťastný & kol., 2006)

stupeň	kategorie	
A	0	Druh pozorovaný v době hnízdění
B	1	Druh pozorovaný v době hnízdění ve vhodném hnízdním prostředí
	2	Pozorování zpívajícího samce (samců) anebo zaslechnutí hlasů souvisejících s hnízděním v hnízdním období.
C	3	Pár pozorovaný ve vhodném hnízdním prostředí v době hnízdění.
	4	Stálý okrsek předpokládaný na základě pozorovaného teritoriálního chování (např. zahánění soků, zpěv apod.) na tomtéž stanovišti nejméně dvakrát v odstupu 1 týdne.
	5	Pozorování toku a imponování nebo páření.
	6	Hledání pravděpodobných hnízdišť.
	7	Vzrušené chování a varování starých ptáků nejspíše v blízkosti hnízda či mláďat.
	8	Přítomnost hnízdních nažin u chycených starých ptáků.
	9	Staří ptáci pozorování při stavbě hnízda nebo dlabání hnízdní dutiny.
D	10	Odpoutávání pozornosti od hnízda nebo mláďat předstíráním zranění.
	11	Nález použitého hnízda (obydleného či opuštěného během pozorování) či zbytků vaječných skořápek.
	12	Nález čerstvě vylétaných mláďat (u krmivých ptáků) nebo mláďat v prachovém peří (u nekrmivých).
	13	Pozorování starých ptáků přelétajících na hnízdiště či opouštějících je za okolností, které nasvědčují přítomnosti obsazeného hnízda (včetně vysoko umístěných hnízd nebo hnízdních dutin, do nichž není vidět) či pozorování starých ptáků vysezujících snůšky.
	14	Pozorování starých ptáků při odnášení trusu od hnízda nebo při přinášení potravy mláďatům.
	15	Nález hnízda s vejci.
	16	Nález hnízda s mláďaty (viděnými nebo slyšenými).

Příloha 2

Přehled tříd databáze CORINE Land Cover (Koželuh, 2003)

Třída 1: Urbanizovaná území (ARTIFICIAL AREAS)
Třída 1.1 Obytné plochy (Urban fabric)
Třída 1.2 Průmyslové a obchodní zóny, komunikační síť (Industrial, commercial and transport units)
Třída 1.3 Doly, skládky a staveniště (Mine, dump and construction sites)
Třída 1.4 Plochy umělé, nezemědělské zeleně (Artificial non-agricultural vegetated areas)
111 Městská souvislá zástavba (Continuous urban fabric) 112 Městská nesouvislá zástavba (Discontinuous urban fabric) 121 Průmyslové nebo obchodní zóny (Industrial or commercial units) 122 Silniční a železniční síť a přilehlé prostory (Road and rail networks and associated land) 123 Přístavní zóny (v ČR se nevyskytují) (Port areas) 124 Letiště (Airports) 131 Těžba hornin (Mineral extraction sites) 132 Skládky (Dump sites) 133 Staveniště (Construction sites) 141 Plochy městské zeleně (Green urban areas) 142 Zařízení pro sport a rekreaci (Sport and leisure facilities)
Třída 2: Zemědělské plochy (AGRICULTURAL AREAS)
Třída 2.1 Orná půda (Arable land)
Třída 2.2 Stálé kultury (Permanent crops)
Třída 2.3 Pastviny (Pastures)
Třída 2.4 Různorodé zemědělské plochy (Heterogeneous agricultural areas)
211 Orná půda mimo zavlažovaných ploch (Non-irrigated arable land) 212 Plochy stále zavlažované (v ČR se nevyskytují) (Permanently irrigated land) 213 Rýžová pole (v ČR se nevyskytují) (Rice fields) 221 Vinice (Vineyards) 222 Ovocné sady a keře (Fruit trees and berry plantations) 223 Olivové porosty (v ČR se nevyskytují) (Olive groves) 231 Louky (Pastures) 241 Roční kultury přidáné ke stálým kulturám (Annual crops associated with permanent crops) 242 Komplexní systémy kultur a parcel (Complex cultivation patterns) 243 Převážně zemědělská území s příměsí přirozené vegetace (Land principally occupied by agriculture, with significant areas of natural vegetation) 244 Území zemědělskolesnická (v ČR nepravděpodobné) (Agro-forestry areas)
Třída 3: Lesy a polopřirodní oblasti (Forest and Semi-natural Areas)
Třída 3.1 Lesy (Forests)
Třída 3.2 Plochy s křovinnou a travnatou vegetací (Shrubs and/or herbaceous vegetation associations)
Třída 3.3 Otevřené plochy s malým zastoupením vegetace nebo bez vegetace (Open spaces with little or no vegetation)

- 311 Listnaté lesy (Broad-leaved forest)
- 312 Jehličnaté lesy (Coniferous forest)
- 313 Smíšené lesy (Mixed forest)
- 321 Přírodní pastviny (Natural grassland)
- 322 Slatiny a vřesoviště, křovinaté formace (Moors and heathland)
- 323 Sklerofylní vegetace (v ČR se nevyskytují) (Sclerophyllous vegetation)
- 324 Přechodová stadia lesa a křoviny (Transitional woodland shrub)
- 331 Pláže, duny, písky (v ČR silně nepravděpodobné) (Beaches, dunes and sand plains)
- 332 Holé skály (Bare rock)
- 333 Oblasti s řídkou vegetací (Sparsely vegetated areas)
- 334 Vypálené oblasti (v ČR silně nepravděpodobné) (Burnt areas)
- 335 Ledovce a věčný sníh (v ČR se nevyskytují) (Glaciers and perpetual snow)

Třída 4: Humidní území (WETLANDS)

Třída 4.1 Vnitrozemská humidní území (Inland wetlands)

Třída 4.2 Přímořská humidní území (Coastal wetland)

- 411 Vnitrozemské bažiny (Inland marshes)
- 412 Rašeliniště (Peatbogs)
- 421 Přímořské bažiny (v ČR se nevyskytují) (Salt marshes)
- 422 Slané bažiny (v ČR se nevyskytují) (Salines)
- 423 Příbřežní zóny (v ČR se nevyskytují) (Intertidal flats)

Třída 5: Vodní plochy (WATER BODIES)

Třída 5.1 Pevninské vody (Inland waters)

Třída 5.2 Mořské vody (Marine waters)

- 511 Vodní toky a cesty (Water courses)
- 512 Vodní plochy (Water bodies)
- 521 Laguny (v ČR se nevyskytují) (Coastal lagoons)
- 522 Ústí řek (v ČR se nevyskytují) (Estuaries)
- 523 Moře a oceány (v ČR se nevyskytují) (Sea and ocean)

Příloha 3

Skript pro výpočet pěti modelů logistické regrese z připravených dat jedné geodatabáze.

```
data = read.table('CLC90.txt', dec = '.')
data
data = data.frame(data)
names(data) = c("kvadrat", "spol_hran", "plocha_les", "obvod_les", "plocha_pole", "obvod_pole", "vyskyt")
data

#generování matice pro jednotlivé prediktory a výpočty regrese

#prediktor A
počet=628
pm1=c(rep(0,počet))
pm2=c(rep(0,počet))
for (i in 1:počet)
{
  pm1[i]=data[i,2]
  pm2[i]=data[i,7]
}
pm1
pm2
prediktor_A=matrix(c(pm1,pm2),ncol=2,byrow=F)
prediktor_A

#pojmenování sloupců
prediktor_A = data.frame(prediktor_A)
names(prediktor_A) = c("spol_hran", "vyskyt")

#regrese
vysledek_A = glm(vyskyt ~ spol_hran, family = binomial, data = prediktor_A)
summary(vysledek_A)
confint(vysledek_A)

#prediktor B
počet=628
pm1=c(rep(0,počet))
pm2=c(rep(0,počet))
for (i in 1:počet)
{
  pm1[i]=data[i,3]
  pm2[i]=data[i,7]
}
pm1
pm2
prediktor_B=matrix(c(pm1,pm2),ncol=2,byrow=F)
prediktor_B

#pojmenování sloupců
prediktor_B = data.frame(prediktor_B)
names(prediktor_B) = c("plocha_les", "vyskyt")

#regrese
vysledek_B = glm(vyskyt ~ plocha_les, family = binomial, data = prediktor_B)
summary(vysledek_B)
confint(vysledek_B)

#prediktor C
počet=628
pm1=c(rep(0,počet))
```



```

pm2=c(rep(0,počet))
for (i in 1:počet)
{
pm1[i]=data[i,4]
pm2[i]=data[i,7]
}
pm1
pm2
prediktor_C=matrix(c(pm1,pm2),ncol=2,byrow=F)
prediktor_C
#pojmenování sloupců
prediktor_C = data.frame (prediktor_C)
names (prediktor_C)= c("obvod_les", "vyskyt")
#regrese
vysledek_C = glm (vyskyt ~ obvod_les, family = binomial, data = prediktor_C)
summary (vysledek_C)
confint (vysledek_C)

#prediktor D
počet=628
pm1=c(rep(0,počet))
pm2=c(rep(0,počet))
for (i in 1:počet)
{
pm1[i]=data[i,5]
pm2[i]=data[i,7]
}
pm1
pm2
prediktor_D=matrix(c(pm1,pm2),ncol=2,byrow=F)
prediktor_D
#pojmenování sloupců
prediktor_D = data.frame (prediktor_D)
names (prediktor_D)= c("plocha_pole", "vyskyt")
#regrese
vysledek_D = glm (vyskyt ~ plocha_pole, family = binomial, data = prediktor_D)
summary (vysledek_D)
confint (vysledek_D)

#prediktor E
počet=628
pm1=c(rep(0,počet))
pm2=c(rep(0,počet))
for (i in 1:počet)
{
pm1[i]=data[i,6]
pm2[i]=data[i,7]
}
pm1
pm2
prediktor_E=matrix(c(pm1,pm2),ncol=2,byrow=F)
prediktor_E
#pojmenování sloupců
prediktor_E = data.frame (prediktor_E)
names (prediktor_E)= c("obvod_pole", "vyskyt")
#regrese
vysledek_E = glm (vyskyt ~ obvod_pole, family = binomial, data = prediktor_E)
summary (vysledek_E)
confint (vysledek_E)

```

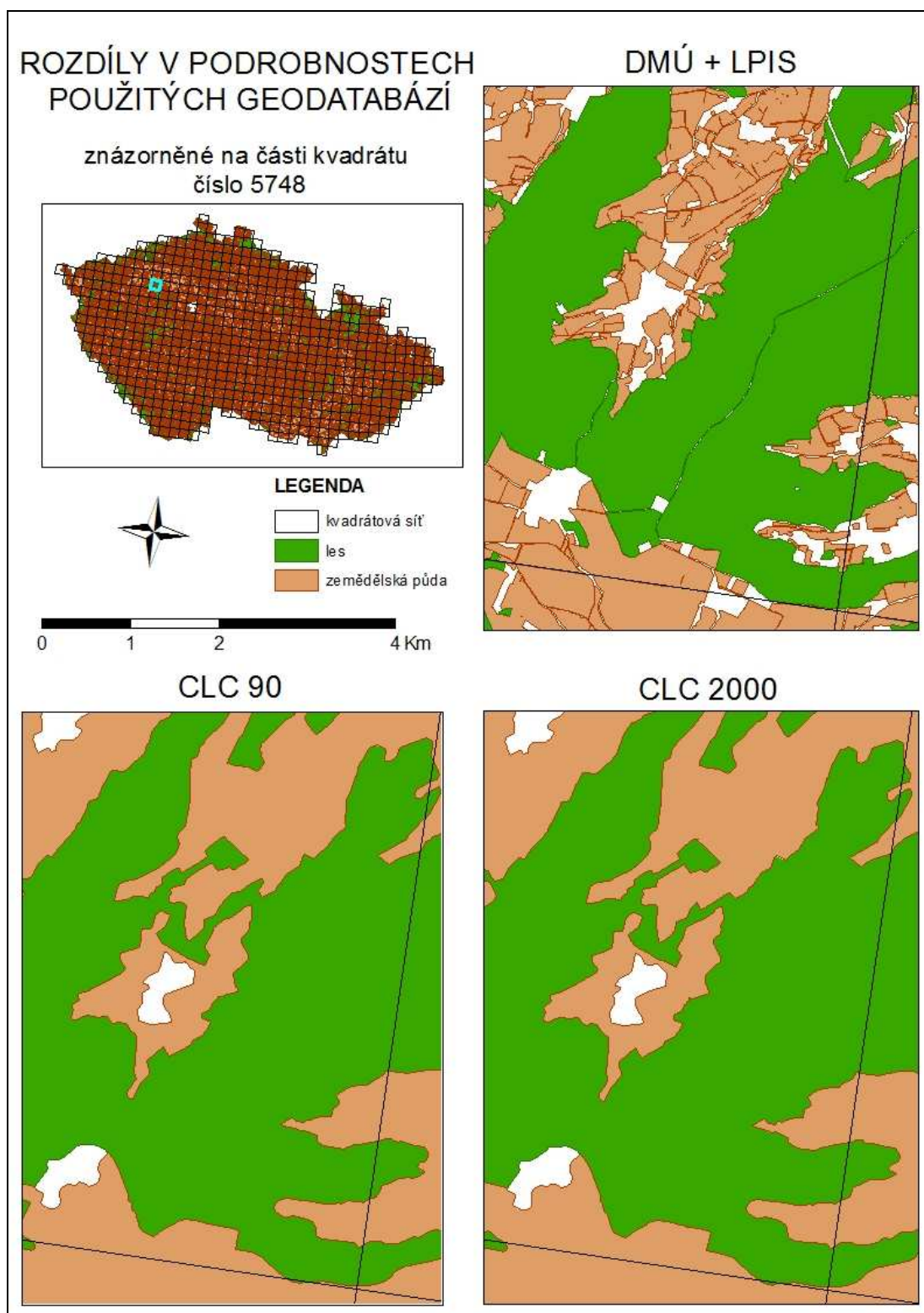
Příloha 4

Výsledná tabulka vypočítaných parametrů logistické regrese a intervalů spolehlivosti.

geodatabáze	model	prediktor	odhadnuté parametry	střední chyba odhadu	interval spolehlivosti	rozsah intervalu	vypočítaná významnost	hladina významnosti		
DMU+LPIS	A	(intercept)	β_0	1.281000	0.185700	0.924063	1.653056	0.000000	*** 0.001	
		spol_hran	β_1	0.000034	0.001092	-0.002104	0.002185	0.004289	0.975000	
	B	(intercept)	β_0	2.457218	0.232227	2.014283	2.925862	0.911579	<2e-16	*** 0.001
		plocha_les	β_1	-0.023167	0.003878	-0.030894	-0.015660	0.015234	2.33E-09	*** 0.001
	C	(intercept)	β_0	1.939781	0.269585	1.423198	2.481445	1.058247	6.23E-13	*** 0.001
		obvod_les	β_1	-0.001824	0.000681	-0.003170	-0.000496	0.002675	0.00743	** 0.01
	D	(intercept)	β_0	-0.126570	0.216240	-0.550640	0.298658	0.849298	0.558	
		plocha_pole	β_1	0.027470	0.004050	0.019648	0.035547	0.015899	1.17E-11	*** 0.001
	E	(intercept)	β_0	0.067729	0.243894	-0.409360	0.548794	0.958154	0.781	
		obvod_pole	β_1	0.001612	0.000311	0.001007	0.002230	0.001223	2.28E-07	*** 0.001
CLC 90	A	(intercept)	β_0	1.211423	0.215682	0.794951	1.641590	0.846639	1.95E-08	*** 0.001
		spol_hran	β_1	0.000557	0.001451	-0.002272	0.003424	0.005696	0.701	
	B	(intercept)	β_0	2.153700	0.222533	1.727789	2.601304	0.873515	<2e-16	*** 0.001
		plocha_les	β_1	-0.018702	0.004067	-0.026753	-0.010778	0.015975	4.26E-06	*** 0.001
	C	(intercept)	β_0	1.362389	0.241398	0.897079	1.844716	0.947637	1.66E-08	*** 0.001
		obvod_les	β_1	-0.000474	0.001363	-0.003145	0.002207	0.005352	0.728	
	D	(intercept)	β_0	-0.401061	0.245730	-0.884890	0.080554	0.965444	0.103	
		plocha_pole	β_1	0.025021	0.003552	0.018151	0.032100	0.013949	1.87E-12	*** 0.001
	E	(intercept)	β_0	0.188976	0.292473	-0.382245	0.766918	1.149163	0.518196	
		obvod_pole	β_1	0.005454	0.001417	0.002694	0.008260	0.005566	0.000119	*** 0.001
CLC 2000	A	(intercept)	β_0	1.215273	0.215740	0.798675	1.645537	0.846861	1.77E-08	*** 0.001
		spol_hran	β_1	0.000529	0.001453	-0.002305	0.003401	0.005706	0.716	
	B	(intercept)	β_0	2.153400	0.222347	1.727829	2.600609	0.872780	<2e-16	*** 0.001
		plocha_les	β_1	-0.018666	0.004056	-0.026695	-0.010763	0.015932	4.19E-06	*** 0.001
	C	(intercept)	β_0	1.364429	0.241616	0.898672	1.847164	0.948492	1.63E-08	*** 0.001
		obvod_les	β_1	-0.000487	0.001366	-0.003163	0.002199	0.005362	0.721	
	D	(intercept)	β_0	-0.394462	0.245377	-0.877513	0.086534	0.964047	0.108	
		plocha_pole	β_1	0.024983	0.003555	0.018106	0.032067	0.013961	2.11E-12	*** 0.001
	E	(intercept)	β_0	0.172826	0.293249	-0.400045	0.752174	1.152219	0.556	
		obvod_pole	β_1	0.005532	0.001421	0.002764	0.008346	0.005581	9.93E-05	*** 0.001

Příloha 5

Porovnání měřítek mapových podkladů.



Příloha 6

Kalous ušatý (Asio otus) (Rösner, 2005)

