

UNIVERZITA PALACKÉ HO V OLMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA GEOINFORMATIKY



Zbyněk JANOŠKA

**FAKTOROVÁ ANALÝZA VÝSKYTU VYBRANÝCH  
BOTANICKÝCH A ZOOLOGICKÝCH TAXONŮ**

**Bakalářská práce**

Vedoucí práce: RNDr. Vilém PECHANEC, Ph.D  
Olomouc 2009

Prohlašuji, že jsem zadanou bakalářskou práci řešil sám a že jsem uvedl veškerou použitou literaturu.

Olomouc, 27. května 2009

.....

Děkuji vedoucímu práce RNDr. Vilému Pechancovi za pomoc, odborné vedení a poskytnutá data. Za pomoc při řešení a řadu cenných podnětů děkuji Mgr. Pavlu Tučkovi a pracovníkům správy CHKO Bílé Karpaty.

## Obsah

<b>1. ÚVOD.....</b>	<b>8</b>
<b>2. CÍLE PRÁCE.....</b>	<b>9</b>
<b>3.PROSTŘEDÍ A METODY ZPRACOVÁNÍ.....</b>	<b>10</b>
3.1 Prostředí zpracování.....	10
3.1.1 ArcGIS Desktop 9.2.....	10
3.1.2 Geostatistical Analyst.....	11
3.1.2.1 Deterministické metody.....	12
3.1.2.2 Geostatistické metody.....	13
3.1.2.3 Výběr vhodné interpolační metody.....	14
3.1.2.4 Postup při interpolaci dat.....	15
3.1.3 R.....	23
3.2 Metody zpracování.....	26
3.2.1 $\chi^2$ -test.....	26
3.2.2 ANOVA.....	27
3.2.3 Logistická regrese.....	29
3.2.4 Geografický střed.....	31
3.2.5 Shluková analýza.....	31
<b>4. CHARAKTERISTIKA SLEDOVANÝCH JEVŮ.....</b>	<b>34</b>
4.1 Charakteristika území.....	34
4.2 Motýli.....	35
4.3 Cévnaté rostliny.....	36
<b>5. STATISTICKÉ VYHODNOCENÍ DAT.....</b>	<b>39</b>
5.1 Zdrojová data.....	39

5.2 Zpracování dat.....	41
5.3 Sestavení modelu.....	42
5.3.1 $\chi^2$ -test.....	42
5.3.2 ANOVA.....	43
5.3.3 Logistická regrese.....	43
5.3.4 Seskupení dat.....	44
5.3.5 Modelování.....	46
<b>6. SIMULACE.....</b>	<b>48</b>
<b>7. DISKUZE.....</b>	<b>51</b>
<b>8. ZÁVĚR.....</b>	<b>53</b>
<b>9. POUŽITÉ ZDROJE.....</b>	<b>54</b>
<b>10. SUMMARY.....</b>	<b>56</b>
<b>PŘÍLOHY.....</b>	<b>57</b>

## 1. ÚVOD

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

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

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

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

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

## **2. CÍLE PRÁCE**

Cílem práce je analýza závislosti výskytu vybraných botanických a zoologických taxonů navzájem a závislosti na základních přírodních podmínkách.

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

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

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

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

### 3. PROSTŘEDÍ A METODY ZPRACOVÁNÍ

#### 3.1 Prostředí zpracování

Tématem práce je hodnocení závislosti proměnných, většina statistických operací tedy byla prováděna v prostředí programu R. Shluková analýzy byla provedena v trial verzi software QC.Expert. Pro prostorově orientované operace bylo použito programu ArcGIS Desktop 9.3, k geostatistickému modelování pak jeho extenze Geostatistical Analyst. Popisu Geostatistical Analyst je v práci věnován větší prostor, protože se jedná o nástroj, který je přístupný velkému množství uživatelů, nabízí širokou škálu funkcí, práce s ním však vyžaduje základní znalosti tvorby povrchů a geostatistiky. Tento základ, potřebný pro úspěšné osvojení extenze, je obsažen v kapitole 4.1.2.

##### 3.1.1 ArcGIS Desktop 9.3

ArcGIS je rodina produktů firmy ESRI. Jedná se o profesionální GIS software, který umožňuje prohlížení, tvorbu a správu prostorových dat. ArcGIS Desktop se skládá z řady integrovaných aplikací ArcMap, ArcCatalog, ArcToolbox a ArcGlobe. Většina funkcionality programu je přístupná přes ArcToolbox, pokročilé sady nástrojů jsou dostupné v různých extenzích. ArcMap slouží k prohlížení dat (atributová složka i geometrie) a tvorbě grafických výstupů. ArcCatalog je aplikací pro správu dat, tvorbu a správu geodatabáze a editaci metadat. Součástí ArcGIS je i Geoprocessing, jenž obsahuje funkční část programu a umožňuje automatizaci práce a tvorbu vlastních specifických nástrojů pro manipulaci s daty.

ArcGIS Desktop 9.3 je přístupný ve třech licenčních úrovních, lišících se funkcionalitou:

##### **ArcView**

- základní úroveň produktu ArcGIS, umožňující prohlížení dat, jejich editaci, analýzy nad daty a tvorbu grafických výstupů, zejména map. Dovoluje také pracovat s daty uloženými v systému relačních databází přístupných přes ArcSDE.

##### **ArcEditor**

- k funkcionalitě ArcView přidává rozšířené možnosti práce s geodatabázemi, včetně tvorby, editace a kompilace.



## ArcInfo

- nejvyšší verze produktu s rozšířenými možnostmi prostorových operací, tvorby map a navýšenou funkcionalitou Geoprocessingu<sup>[16]</sup>.

### 3.1.2 Geostatistical Analyst

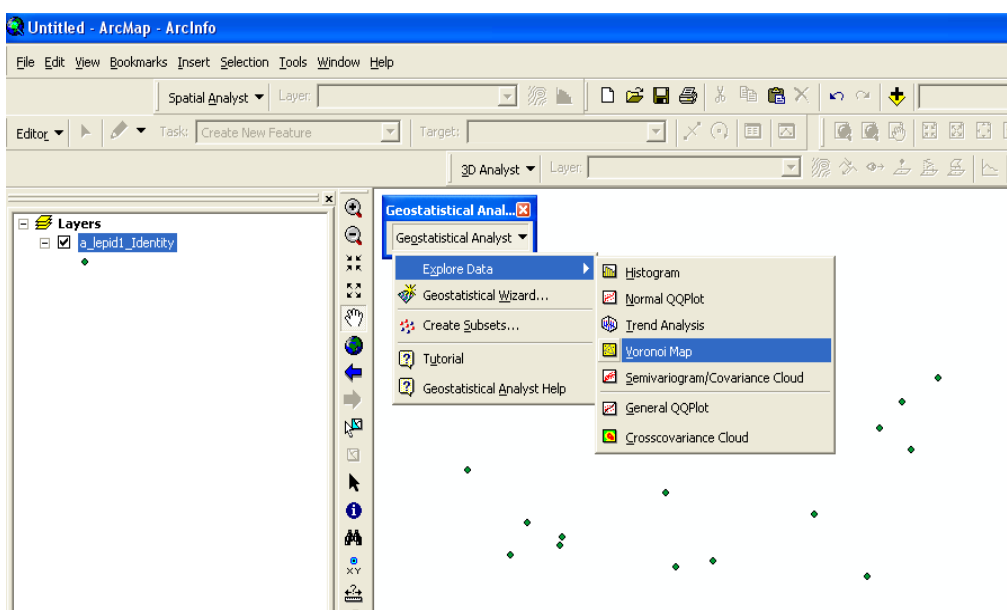
Jedná se o extenzi k programu ArcGIS 9.x, přístupná je přes menu Tools. Obsahuje množství nástrojů pro statistické vyšetření dat, identifikaci anomálií v datech, předpovědi, vyhodnocení přesnosti předpovědí, a tvorbu povrchů.

Právě tvorba povrchů je hlavním nástrojem geostatistického modelování. K tomu se využívá interpolačních metod. Interpolace je procedura odhadu neznámých hodnot ze známých hodnot v okolí<sup>[7]</sup>. Interpolační metody mohou být posuzovány podle mnoha kritérií, z nich nejdůležitější je to, zda zachovávají původní hodnoty, tedy zda si po interpolaci ponechávají vstupní body své hodnoty, nebo jestli je jim přidělena hodnota jiná. Podle tohoto principu dělíme interpolační metody na exaktní (zachovávají vstupní hodnoty) a aproximační (dochází k vyhlazení hodnot i v místech se známou hodnotou)<sup>[7]</sup>.

Nicméně Geostatistical Analyst používá jiné rozdělovací kritérium, a to sice, je-li při výpočtu zohledněn prvek náhodnosti.

Metody, které nepracují s prvkem náhodnosti, se nazývají deterministické, na rozdíl od metod geostatistických (stochastických), kdy je ve výpočtu tento prvek zohledněn.

Všechny metody dostupné v Geostatistical Analyst vychází z předpokladu, že jevy prostorově bližší jsou si více podobné, a tento poznatek (*Spatial Dependence*) kvantifikují.



Obr.č.1: uživatelské rozhraní Geostatistical Analyst

Ovládání je velmi jednoduché, po přidání panelu Geostatistical Analyst do nástrojové lišty je uživateli nabídnuto jen hlavní menu. Z něj je přístup do tří hlavních komponent extenze – Expore Data nabízí výběr z mnoha nástrojů pro základní analýzu dat (popsáno v kapitole 3.1.2.4), Geostatistical Wizard otevře dialogové okno pro samotnou interpolaci dat, a Create Subsets je nástroj pro vytvoření pracovního a kontrolního datasetu (viz. kapitola 3.1.2.3).

### 3.1.2.1 Deterministické metody

Tyto metody jsou založeny buď na měřítku, které je pro model použito, nebo na stupni zarovnání povrchu (Smoothing). Nepoužívají model náhodných prostorových procesů. Existují dvě skupiny deterministických metod.

- **Lokální** - počítají předpovídané hodnoty za použití jen nejbližších okolních bodů k předpovídané lokaci. Mezi lokální metody přístupné v Geostatistical Analyst patří IDW, lokální polynomická interpolace a RBF.
- **Globální** - počítají předpovídané hodnoty za použití celého vstupního datasetu, v Geostatistical Analyst je přístupná jediná metoda – globální polynomická interpolace<sup>[14]</sup>.
- **IDW (Inverse Distance Weight, metoda inverzních vzdáleností)** - základním předpokladem při tvorbě povrchu pomocí IDW je fakt, že měření získaná blíže předpovídané lokaci mají větší vliv na hodnotu jevu, než měření pořízená dále, a že míra vlivu je závislá pouze na vzájemné vzdálenosti bodů. Při výpočtu se využívá váženého lineárního průměru, kde váhou je reciproká hodnota vzdálenosti měření od lokálního odhadu s mocninou  $p$ <sup>[7]</sup>. Pro mocninu  $p$  se používají obecně hodnoty mezi 1 až 3, nejčastěji pak hodnota 2. Větší hodnota  $p$  znamená vyšší vliv nejbližších hodnot. Míra vlivu klesá se vzdáleností exponenciálně.

IDW patří mezi exaktní interpolační metody, kdy povrch prochází přímo vstupními daty a maximální a minimální hodnoty se mohou vyskytovat pouze ve vstupních bodech. Povrch vytvořený pomocí IDW je citlivý na výskyt clusterů a outlierů.

Příklad použití: modelování rozšíření invazivních druhů rostlin<sup>[21]</sup>

- **Lokální polynomická interpolace** - pokrývá celé zájmové území jednotlivými povrchy menších rozsahů, které se vzájemně částečně překrývají; každý z těchto

povrchů je modelován za použití jedinečné matematické funkce. Takový povrch je schopen zachytit více lokálních změn. Polynomická interpolace je vhodná u povrchů, kde v jednom směru převládá určitý trend (růst či pokles). Lokální polynomická interpolace je vhodná pro modelování povrchu, který se vyznačuje změnami malého prostorového rozsahu (*Short Range Variation*). Jedná se o aproximační metodu.

Příklad použití: modelování distribuce pylu v usazeninách<sup>[17]</sup>.

- **RBF (Radial Basis Functions, radiální funkce)** - Základním principem je vytvoření povrchu při co nejmenším zakřivení. Na rozdíl od lokálních a globálních polynomických funkcí vyžaduje, aby povrch procházel hodnotami vstupních dat, na rozdíl od IDW dokáže modelovat i hodnoty vyšší či nižší než jsou vstupní data. RBF je vhodné použít pro modelování hladkých modelů z velkého počtu vstupních dat. Dobrých výsledků dosahuje při modelování pozvolna se měnících povrchů, jako je například převýšení nebo některé hydrologické jevy.

Příklad použití: modelování desertifikace krasových hornin<sup>[20]</sup>.

- **Globální polynomická interpolace** - Globální polynomická interpolace vytvoří hladký povrch, který je definován matematickou funkcí. Takový povrch se mění pozvolna a zachycuje jen hrubý vzor povrchu.

Příklad použití: modelování zněčištění ovzduší nad průmyslovou oblastí<sup>[16]</sup>.

### 3.1.2.2 Geostatistické metody

Tyto metody předpokládají, že alespoň některé z rozdílů sledovaného jevu mohou být modelovány náhodnými procesy s prostorovou autokorelací. Rozdíl mezi deterministickými a geostatistickými metodami je především ten, že geostatistické metody počítají se statistickým modelem, který obsahuje pravděpodobnost. Výsledkem tedy není jen předpověď sledovaného jevu, ale také pravděpodobnost, se kterou je předpovídán.

- **Kriging** - je základní geostatistická metoda pro předpověď hodnoty sledovaného jevu v určitém místě. Všechny metody krigingu jsou závislé na předpokladu autokorelace (korelace proměnných závisí na vzdálenosti a/nebo směru mezi nimi). Stejně jako u IDW jsou měřené hodnoty nejbližší k počítané hodnotě brány jako nejvíce relevantní, nicméně zatímco IDW používá pro výpočet jen jednoduchého algoritmu založeného pouze na vzdálenosti, hodnoty pro kriging vycházejí

z variogramu, stanoveného na základě znalosti vstupních dat. Teoreticky by mělo jít o exaktní metodu, praktické implementace ale ukazují aproximační chování, tedy vyhlazení původních hodnot. V Geostatistical Analyst jsou přístupné následující typy krigingu:

- Ordinary Kriging
- Simple Kriging
- Universal Kriging
- Indicator Kriging
- Probability Kriging
- Disjunctive Kriging
- Cokriging - používá se k modelování proměnné několika typů. Je v podstatě generalizovanou formou krigingu pro případ více vstupních datasetů.

Dvě nejjednodušší a nejčastěji používané metody krigingu jsou následující:

Ordinary Kriging: je nejjednodušší metodou, k výpočtu je třeba znát průměrnou hodnotu sledovaného jevu v poli.

Simple Kriging: předpokládá, že prostorová proměnná má neznámý, ale konstantní aritmetický průměr.

### 3.1.2.3 Výběr vhodné interpolační metody

Výběr interpolační metody souvisí úzce s povahou zkoumaného jevu. Obecně se dá říci, že přírodní procesy je vhodné modelovat pomocí metod deterministických (například převýšení), kdežto sociální jevy pomocí metod geostatistických, protože jsou komplexnější ve své povaze i chování (například ale pro predikci výskytu geologických jevů je vhodné započítat prvek náhodnosti a použít tedy metody geostatistické).

Stejně doporučení platí i pro volbu mezi metodami exaktními a aproximačními (zohledněna například může být přesnost vstupních dat), je také důležité, jestli se jev v území mění spojitě (teplota) nebo skokově (geologické jevy) a zda-li ve výsledku připouštíme hodnoty vyšší či nižší než u vstupních dat.

Nejdůležitější způsob, jak ocenit přesnost výsledného modelovaného povrchu, je porovnat výsledné hodnoty s hodnotami naměřenými v terénu. Jedním z postupů je rozdělení originálního datasetu na dvě části, z nichž jedna je použita pro zjištění struktury dat a tvorbu povrchu, a druhá pak k ověření přesnosti výsledků. Create Subsets dialog box umožňuje vytvoření obou – jak pracovního, tak testovacího – datasetů<sup>[14]</sup>.

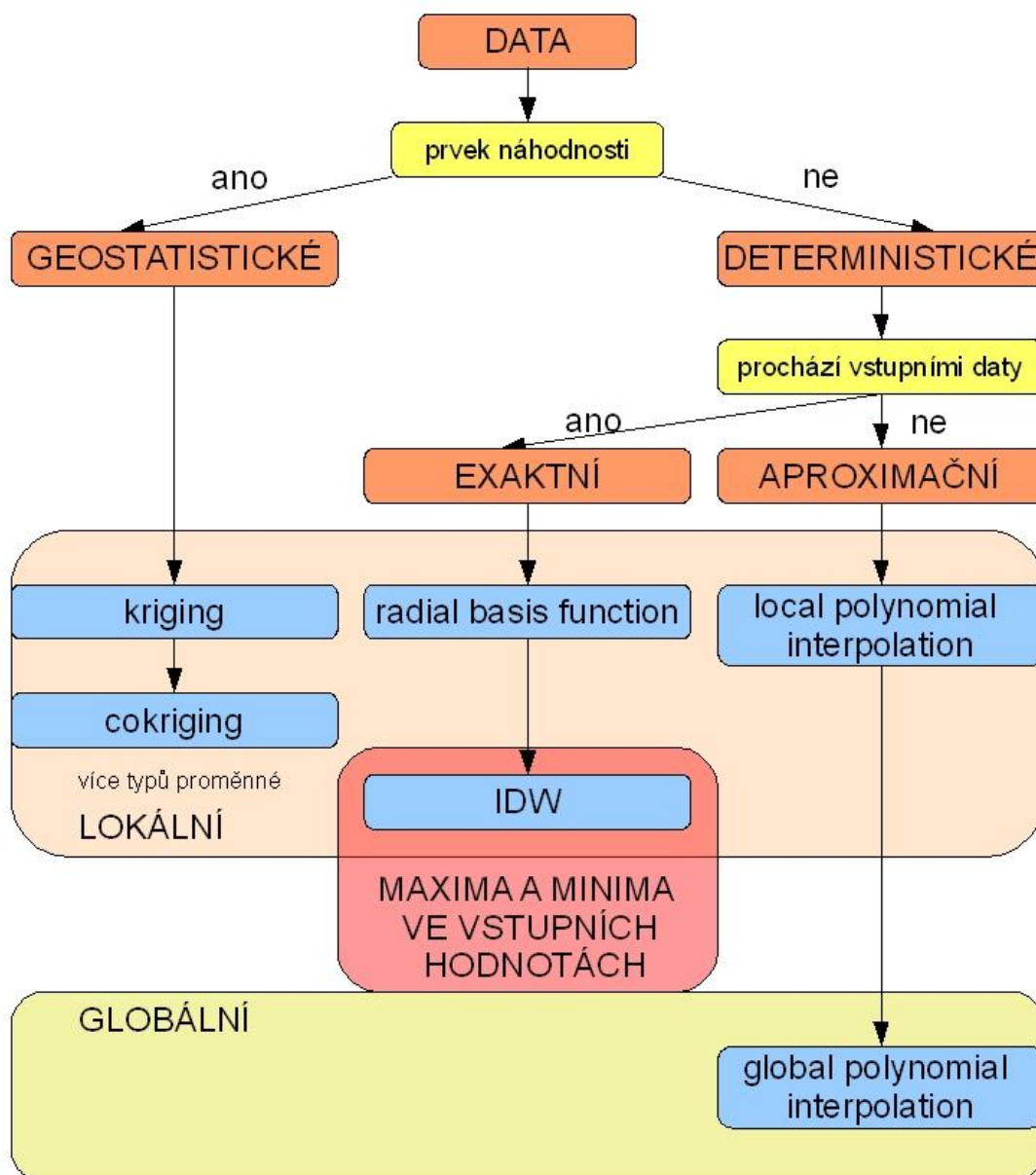
#### 3.1.2.4 Postup při interpolaci dat

Před samotnou interpolací je nejprve nutné provést sadu kroků, které odhalí strukturu dat a pomohou vybrat nejvhodnější interpolační metodu. Nástroje pro průzkumovou analýzu jsou přístupné v menu Explore Data. V průzkumové analýze se vyšetřují anomálie v datech, testuje se normálnost rozdělení a zjišťuje se struktura dat.

Dalším krokem je zvolení prohledávané oblasti. Pokud se jedná o lokální interpolaci (k výpočtu jsou použity pouze nejbližší body, nikoli celý dataset), umožňuje Geostatistical Analyst nastavit prohledávanou oblast (*Searching Neighborhood*). Je možné nastavit tvar (kruh, elipsa), parametry tvaru, jeho orientaci a rozdělení na sektory včetně jejich uspořádání. Je také možné nastavit minimální/maximální počet bodů, který bude do výpočtu zahrnut.

Třetím důležitým krokem je průzkum prostorové autokorelace. K tomu se používají tzv. strukturální funkce, z nichž nejběžnější je semivariogram. Semivariogram je graf, ukazující změnu hodnoty sledované proměnné v závislosti na vzdálenosti. Hodnota semivariogramu se počítá pro každý pár vstupních lokací. Na základě empirického semivariogramu je pak konstruován model, pomocí něhož je modelována distribuce jevu v prostoru. Geostatistical Analyst nabízí celou řadu modelů.

Ukázky dat pocházejí ze vzorového příkladu interpolace koncentrace ozónu<sup>[16]</sup>. Sledovaným jevem je právě distribuce koncentrace ozónu.



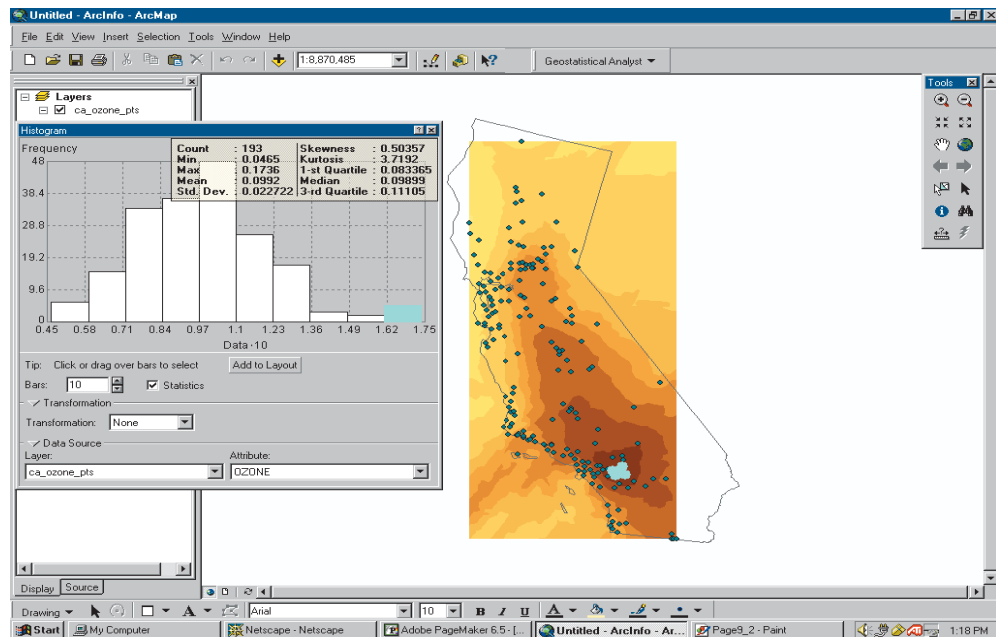
Obr. č.2: schéma interpolačních metod dostupných v Geostatistical Analyst

### Průzkumová analýza dat

- **Histogram** - poskytuje jednoproměnný<sup>1</sup> popis dat. Zobrazuje frekvenční rozdělení a počítá souhrnné statistiky (*summary statistics*). Souhrnné statistiky mohou popisovat míru umístění (*measure of location*) – ARITMETICKÝ PRŮMĚR, MEDIÁN, A HODNOTU PRVNÍHO A TŘETÍHO KVARTILU, míru rozsahu (*measure of spread*) – SMĚRODATNÁ ODCHYLKA, a míru tvaru (*measure of shape*) – ŠIKMOST A ŠPIČATOST. Dále je

<sup>1</sup> charakterizovaný nebo závisející na pouze jedné náhodné proměnné

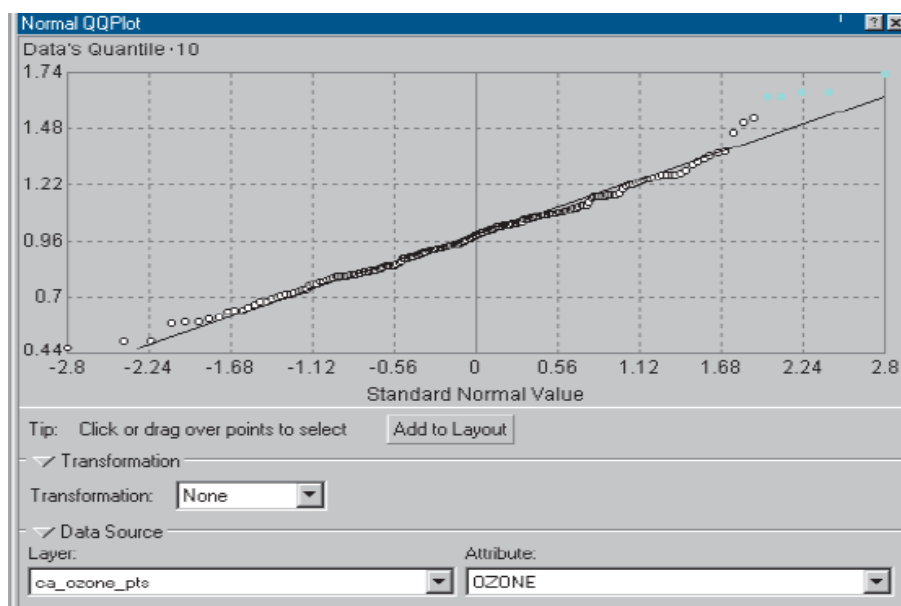
možné v menu histogram transformovat data (Box-Cox, LOGARITMICKÁ, ARCSINOVÁ TRANSFORMACE), pokud jejich rozdělení není normální. Ukazateli normality jsou hodnoty šikmosti a špičatosti blíží se nule.



Obr. č.3: histogram četnosti s vyznačením zvýšené koncentrace ozónu

Histogram vzorových dat ukazuje, že rozdělení sledovaného jevu je normální, území je tedy homogenní, jen poslední sloupeček hodnot ukazuje zvýšenou koncentraci ozónu (označen modře).

- **Voronoi Map** - Voronoi mapy jsou sestaveny ze sérií polygonů formovaných okolo vstupních bodů. Voronoi polygony jsou vytvořeny tak, že každé místo v daném polygonu má blíže ke středovému bodu než k jakémukoli jinému. Podle toho, zda spolu polygony sousedí svou hranicí se určuje, které z bodů patří do sousedství daného bodu. Pokud je definováno sousedství bodů, je možné provést množství statistických operací.
- **Normal QQPlot** – QQPlot je graf, ve kterém kvantily ze dvou různých distribucí jsou vyneseny relativně jeden k druhému. Normal QQPlot je vytvořen vynesením hodnot rozdělení zkoumaného datasetu relativně k normálnímu rozdělení. Pokud leží body grafu blízko vyznačené přímky, je to ukazatelem normálního rozdělení dat.



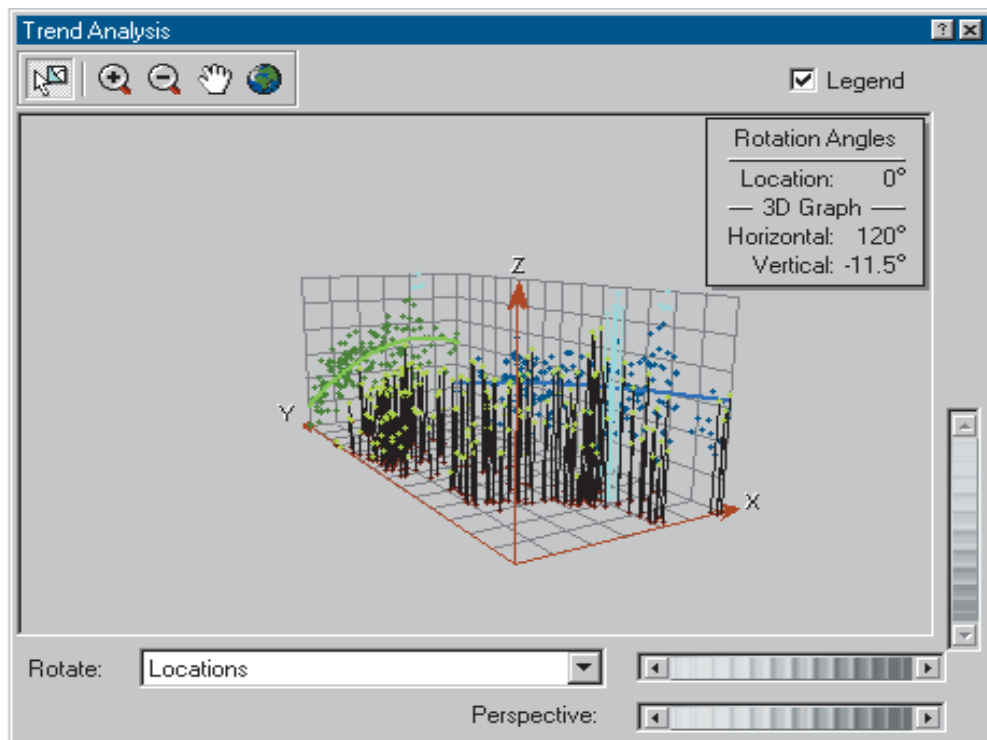
Obr. č.4: Normal QQPlot s vyznačením zvýšené koncentrace ozónu

Normal QQPlot sestrojený nad vzorovými daty ukazuje na normální rozdělení, jen hodnoty vysoké koncentrace ozónu (na obrázku vyznačeny modře) se odchyľují od přímky.

Obecně se dá říci, že data, která se výrazně odchyľují od zbytku datasetu (v tomto případě právě ony vyznačené koncentrace), jsou buďto chyby – a pak musí být z datasetu odstraněny, nebo význačné hodnoty, které mají často pro modelování klíčový význam.

- **General QQPlot** – používá se k porovnání podobnosti rozdělení dvou různých datasetů. Vytvoří se vnesením hodnot dvou datasetů relativně k sobě.
- **Trend Analyst** – pomáhá vysledovat globální trendy ve vstupních datasetech. Podává trojrozměrnou informaci o datech. V Geostatistical Analyst je trend zobrazen prostorově. Body jsou promítnuty do roviny a reprezentují hodnotu jevu v určité lokaci. Ve dvou směrech (sever-jih, východ-západ) je pak zobrazen trend jako polynom (až třetího řádu).

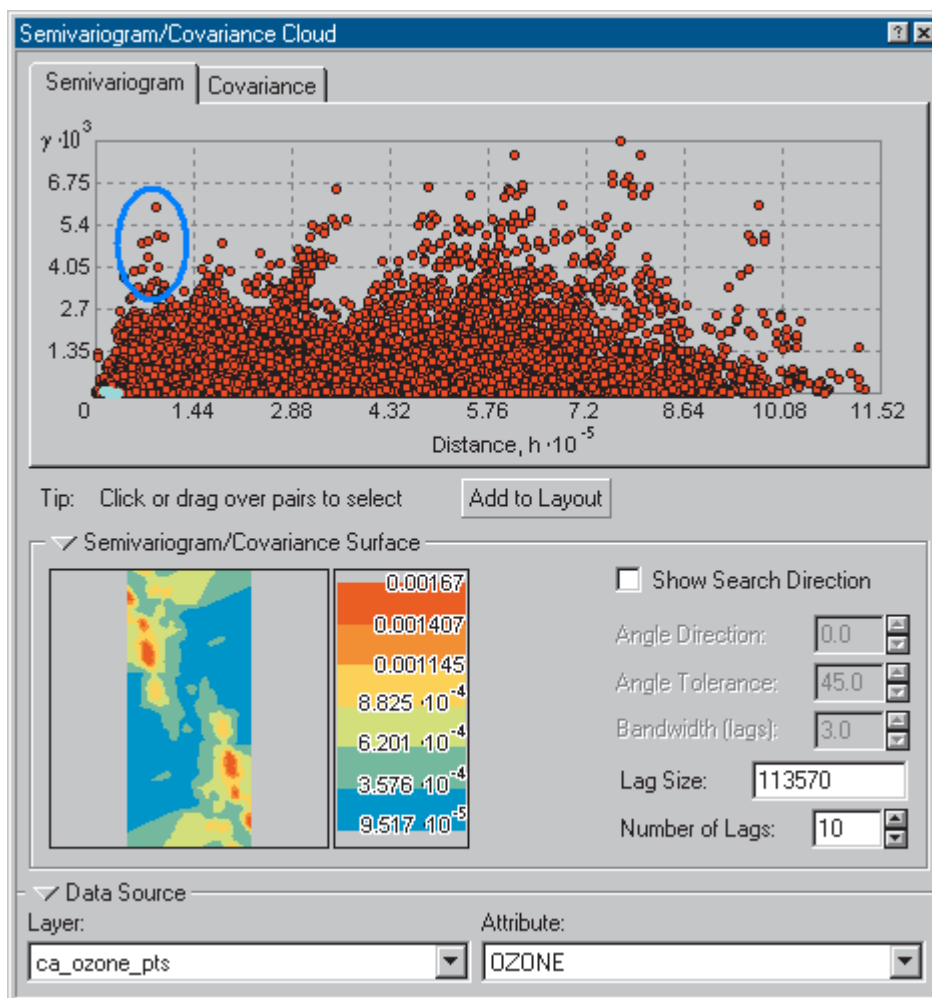




Obr. č.5: Zobrazení trendu

Pokud by byly jak zelená, tak modrá linie rovné, znamenalo by to, že se v datech nevyskytuje žádný trend, protože však zelená linie má tvar obráceného U, ukazuje to na trend, který může být vyjádřen pomocí polynomu druhého stupně.

- **Semivariogram/Covariance Cloud** – ukazuje empirický semivariogram a kovarianci pro všechny páry lokací v datasetu. Pomocí tohoto nástroje se dá studovat autokorelace mezi jednotlivými body. Autokorelace předpokládá, že body, které jsou si blíže, si jsou více podobné, tudíž by měly mít nižší hodnotu v semivariogramu. Se zvyšující se vzdáleností bodů od sebe roste i hodnota v semivariogramu.



Obr. č.6: Semivariogram/Covariance Cloud

Na obrázku je vidět, že některá data v blízkosti osy x mají vyšší hodnotu semivariogramu než by se dalo v tak malé vzdálenosti očekávat (vyznačena modrou elipsou). Tato data jsou vhodná k bližšímu prozkoumání (ve skutečnosti se jedná o vysoké koncentrace ozónu, tedy o stejná data, která se jevila jako odlišná na histogramu i na QQPlotu). Nutným krokem je úprava parametrů semivariogramu (LAG SIZE, NUMBER OF LAGS).

- **Crosscovariance Cloud** – ukazuje empirickou crossvarianci pro všechny páry lokací ve dvou datasetech. Může být použit k zjištění lokálních charakteristik prostorové závislosti mezi dvěma datasety a také pro zjištění prostorových posunů závislosti mezi dvěma datasety.

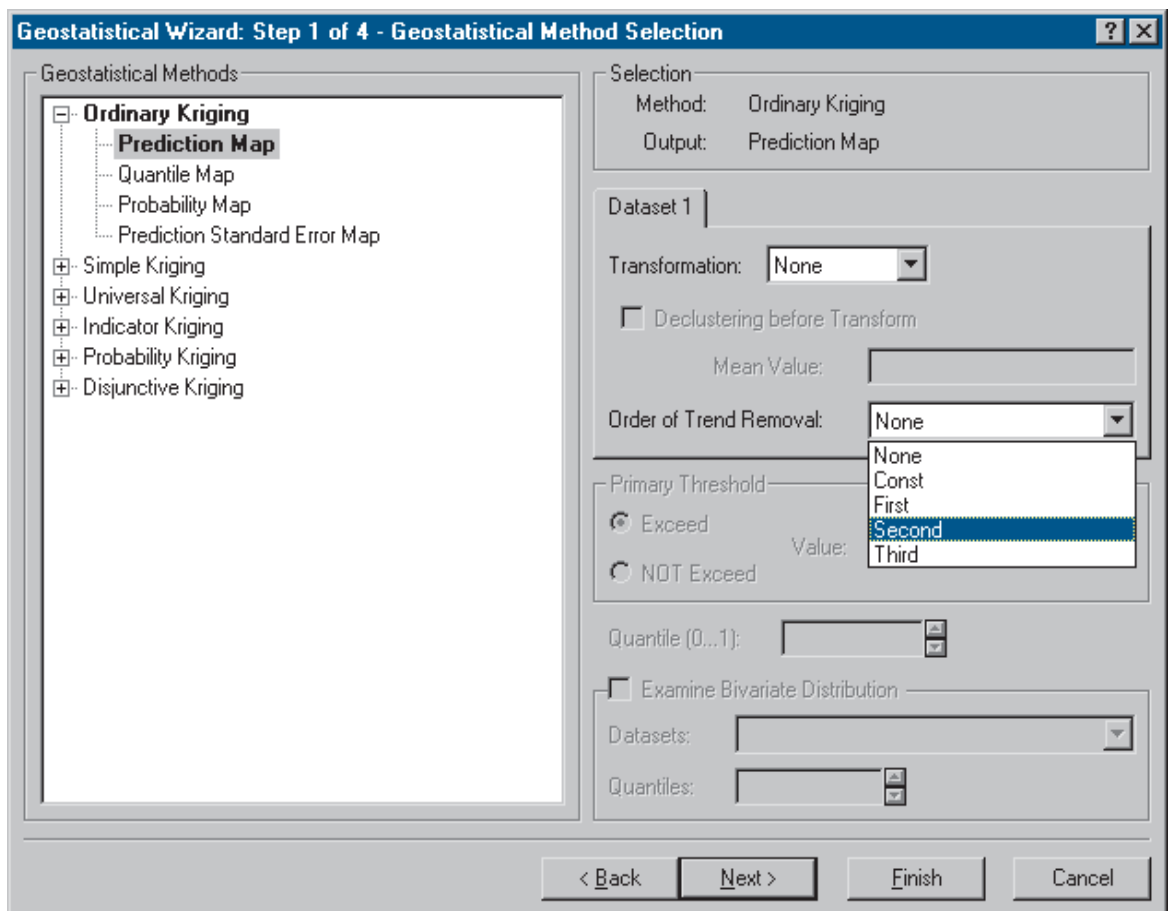
### Vlastní interpolace

Po základních analýzách je možné přistoupit k vlastní interpolaci. V prostředí

Geostatistical Analyst probíhá modelování přes menu Geostatistical Wizard, kde dochází k výběru interpolační metody i nastavení jejích parametrů.

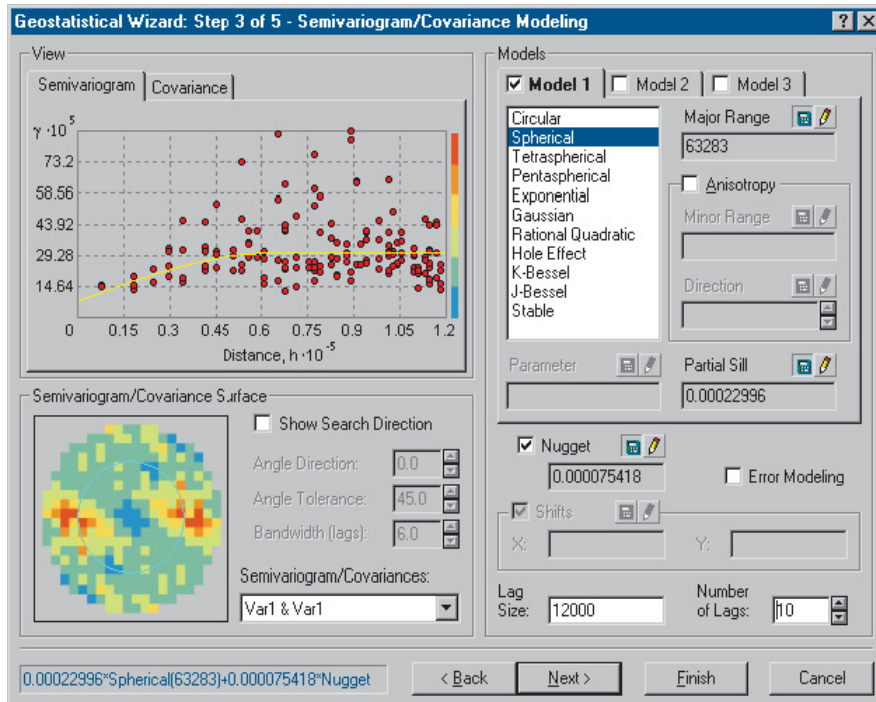
Pro interpolaci modelových dat bude použito metody ORDINARY KRIGING. Před samotnou interpolací je také nutné odstranit trend, pokud se v datech vyskytuje. Ten by totiž zkreslil výsledky.

Prvním krokem je vybrání vhodné interpolační metody (v tomto případě Ordinary Kriging) a typu mapy, která má být vytvořena (PREDICTION MAP jako mapa předpokládaných koncentrací). V tomto dialogovém okně (obr.7) je také možné nastavit ODSTRANĚNÍ TRENDU.



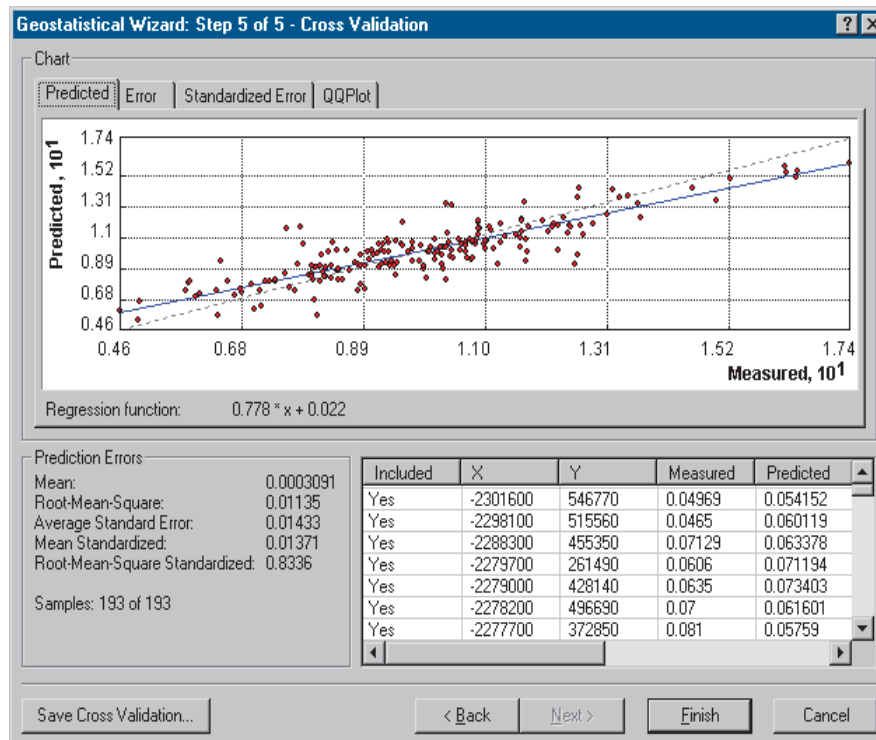
Obr. č.7: Dialogové okno pro volbu interpolační metody a odstranění trendu

V dalším kroku (obr.6) dochází k nastavení parametrů semivariogramu a volbě prohledávací oblasti. Volí se také model semivariogramu a je možné zároveň použít až tři různé modely na jednu datovou sadu.



Obr. č.8: Volba modelu semivariogramu, nastavení parametrů

Závěrečné dialogové okno (obr.9) umožňuje tzv. křížovou validaci (*Cross Validation*), která poskytuje informaci o tom, jak dobře model sedí na vstupní hodnoty. V tabulce jsou jak hodnoty původní, tak hodnoty predikované, dále střední kvadratická chyba (měla by se blížit 1) a průměrná chyba predikce (měla by se blížit 0)<sup>[7]</sup>.



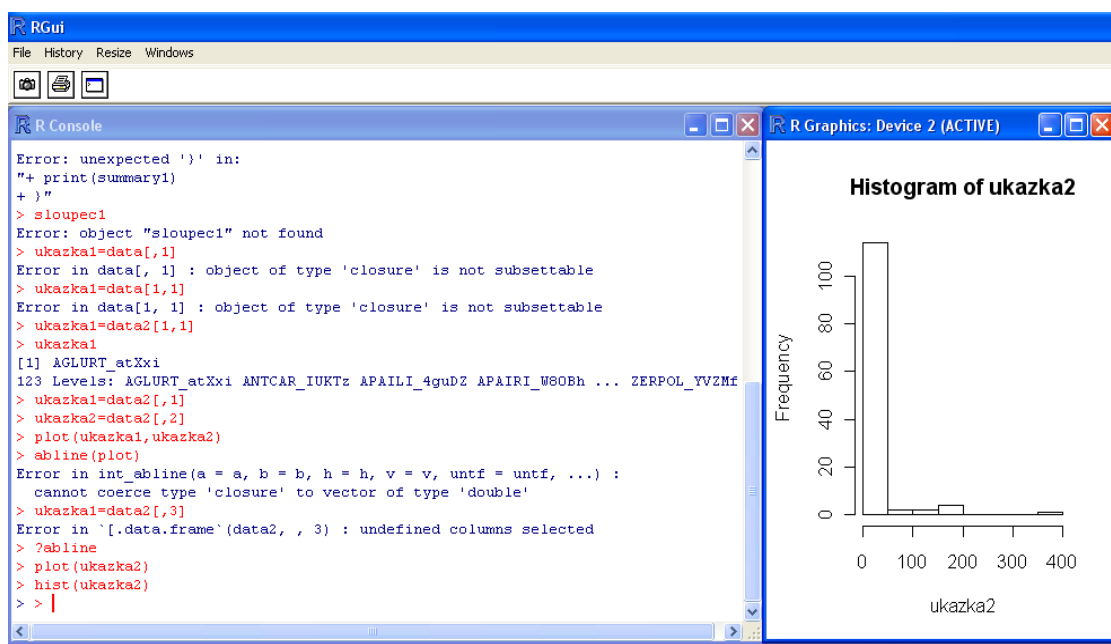
Obr. č.9: Cross Validation

Na závěr je možné vytvořit mapu předpokládaných středních chyb, která ukazuje pravděpodobnost, se kterou jsou hodnoty na různých místech předpovídány. Tyto mapy jsou vhodné pro porovnávání a hledání optimálního modelu pro daný jev a území.

### 3.1.3 R

R je integrované prostředí pro analýzu dat, statistické a matematické výpočty a produkci grafiky. Na rozdíl od jiných programů konstruovaných pro matematické výpočty, R obsahuje množství už definovaných funkcí pro složitější statistické operace, a uživatel je nemusí sám programovat. R je zároveň programovací jazyk (je implementací programovacího jazyka S). V prostředí R jsou jednotlivé příkazy spouštěny z příkazového řádku.

Současná verze R je 2.8.1, v práci bylo pracováno s touto základní verzí bez jakýchkoli extenzí nebo speciálních kompilací.

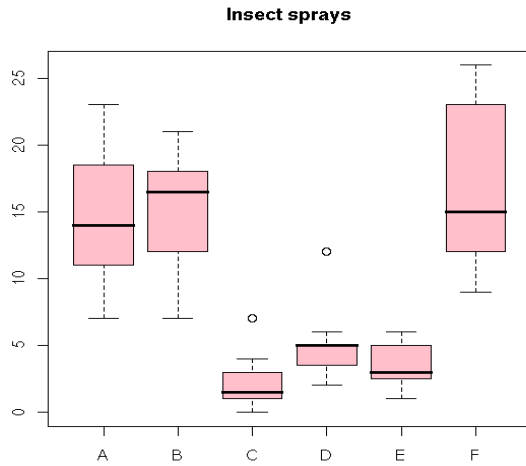


Obr. č.10: ukázka pracovního prostředí R

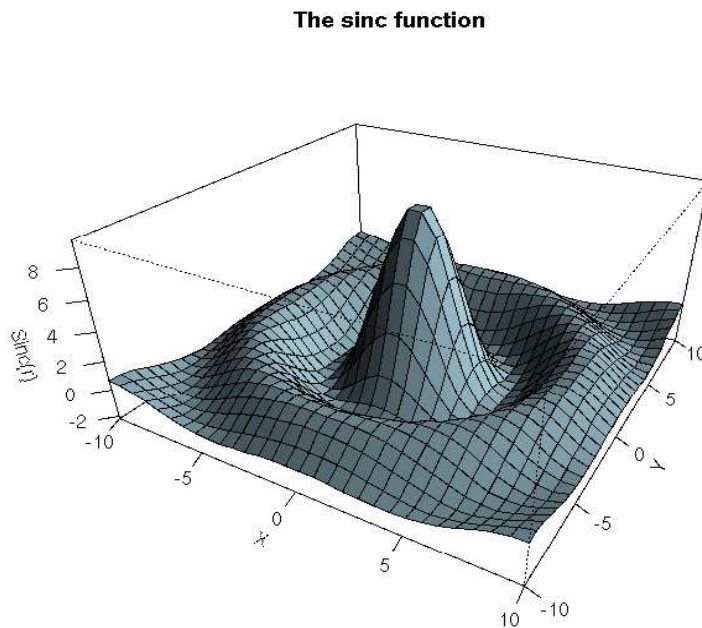
Protože mnoho funkcí je již definováno a zabudováno přímo v programu, snižuje se čas strávený programováním uživatele, nevýhodou je ale pomalejší rychlost některých složitějších operací, běžný uživatel však tento problém nepocítí. Další problémy plynou z toho, že R ukládá veškerá používaná data a všechny proměnné přímo do paměti, tento problém lze odstranit nepoužíváním platformy Windows, nebo propojením s databázemi,

ze kterých R bere jen aktuálně používaná data a neskladuje veškerá data v paměti.

R také nabízí nástroje pro produkci grafiky, tato je ale vhodná spíše pro studijní účely než pro prezentaci. Základními grafickými výstupy jsou grafy s velkým množstvím uživatelsky definovatelných parametrů.



Obr. č.11: grafické výstupy z R – Box Plot



Obr. č.12: grafické výstupy z R – vícerozměrná data

R je freeware s možností definice vlastních funkcí, stejně jako s možností prostředí rozšířit o mnoho předvytvořených knihoven. Některé z nich ale slouží jen pro nekomerční užití. Pro složitější úlohy je možné volat kód C, C++ nebo Fortran a také pomocí C kódu manipulovat s R objekty.

### 3.2 Metody zpracování

Pro zpracování dat bylo použito více metod statistické analýzy, které budou v následující kapitole probrány a vysvětleny. Jejich konkrétní použití je popsáno v příslušných podkapitolách kapitoly 6.3. Nejprve bylo nutné vyšetřit data obecně, zjistit, zda vůbec v celém datasetu existují nějaké závislosti. K tomu posloužil  $\chi^2$ -test prováděný v kontingenčních tabulkách. Dále se zjišťovaly podrobnější vazby mezi jednotlivými jevy pomocí metody ANOVA. Po prokázání závislostí mezi sledovanými jevy bylo potřeba tuto závislost vyjádřit pomocí modelu. Pro velký počet vstupních dat však každý jednotlivý model obsahoval více než 1000 parametrů. Bylo nutné model zjednodušit a to sloučením botanických i zoologických taxonů do skupin. Každá tato skupina pak vstupovala do výpočtu jako jeden parametr. Parametry modelu byly odhadnuty metodou logistické regrese a otestovány, jsou-li významné. Pro jednotlivé skupiny taxonů motýlů pak byly sestaveny modely, kvantifikující závislost jejich výskytu na výskytu vyšších cévnatých rostlin. Pomocí těchto modelů pak byly konstruovány areály výskytu vybraných druhů motýlů (více kapitola 7 – Simulace).

#### 3.2.1 $\chi^2$ -test

Zpracováno podle [9].

Základem pro provedení  $\chi^2$ -testu je konstrukce tabulky četností dvou diskrétních proměnných  $A$ ,  $B$  na základě experimentálně zjištěných výskytů jednotlivých kombinací úrovní těchto proměnných. Počet úrovní proměnné  $A$  je označen  $r$ , počet úrovní proměnné  $B$  je označen  $c$ .

	$B$			
$A$	$B_1$	...	$B_c$	Celkem
$A_1$	$n_{11}$	...	$n_{1c}$	$n_{1.}$
...	...	...	...	...
$A_r$	$n_{r1}$		$n_{rc}$	$n_{r.}$
Celkem	$n_{.1}$	...	$n_{.c}$	$n$

Tab. č.1: kontingenční tabulka pro výpočet  $\chi^2$  statistiky

Z této tabulky se vypočítá testovací statistika  $C$ , která má  $\chi^2$  rozdělení:

$$C = n \sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}^2}{(n_i n_j)} - n \quad (\text{vz. } 1-1)$$

Testovací statistika  $C$  se porovnává s kritickou hodnotou  $\chi^2$  rozdělení  $\chi^2_{(r-1)(c-1)}(1-\alpha)$ . Je-li  $C > \chi^2$ , hypotéza  $H_0$  o nezávislosti  $A$  a  $B$  se zamítá. Závislost  $A$  a  $B$  znamená, že pravděpodobnosti, a tím i počty výskytů ve sloupcích pro alespoň jednu úroveň (hodnotu) proměnné  $A$  jsou ovlivněny úrovněmi proměnné  $B$ . Jsou-li  $A$  a  $B$  nezávislé, pak je pravděpodobnost jevu  $p_{ij}$  jevu  $A_j \wedge B_i$  rovna součinu marginálních<sup>2</sup> pravděpodobností  $p_i p_j$ . Tyto pravděpodobnosti se dají vypočítat ze součtů četností v řádcích a sloupcích,  $p_i = n_i/n$ ,  $p_j = n_j/n$ . Platí, že  $\sum p_i = \sum p_j = 1$ .

### 3.2.2 ANOVA

Analýza rozptylu (ANOVA) představuje účinný statistický nástroj pro zkoumání vztahu mezi vysvětlovanými a vysvětlujícími proměnnými. Vysvětlující proměnné se nazývají faktory. Zkoumáme-li vliv jediného faktoru na jednu či více vysvětlovaných proměnných, jde o jednofaktorovou analýzu rozptylu (jednoduché třídění). Při více faktorech mluvíme o vícefaktorové analýze (dvojném, trojném, atd. třídění). Jednorozměrná analýza rozptylu (ANOVA) předpokládá jedinou vysvětlovanou proměnnou, pomocí vícerozměrné analýzy rozptylu (MANOVA) zkoumáme vliv jednoho či více faktorů na několik vysvětlovaných proměnných současně<sup>[2]</sup>.

V práci byla použita metoda jednofaktorové analýzy rozptylu s jednou vysvětlovanou proměnnou.

Cílem analýzy rozptylu je porovnat úroveň měřitelné proměnné v různých skupinách. Abychom zjistili, zda pozorovaná variabilita proměnné podléhá pouze náhodným kolísáním, nebo zda se v ní projevuje také odlišná úroveň hodnot v různých skupinách, rozkládáme celkovou pozorovanou variabilitu na složky, odpovídající různým zdrojům variability. Odtud název analýza rozptylu<sup>[2]</sup>.

#### **Jednorozměrná analýza s jedním faktorem**

Zpracováno podle [9].

Hodnotu sledované veličiny můžeme vyjádřit jako součet neznámé konstanty  $\mu$ ,

<sup>2</sup> Mezních, okrajových. V tabulce 1 jsou marginální pravděpodobnosti hodnoty  $n_{.j}$  až  $n_{.c}$  a  $n_{i.}$  až  $n_{r.}$



odpovídající  $k$ -té úrovni faktoru, a náhodné složky  $\varepsilon_{ki}$ , zahrnující ostatní vlivy. Model má potom tvar

$$y_{ki} = \mu + \alpha_k + \varepsilon_{ki}, \quad (\text{vz. } 2-1)$$

$$i = 1, 2, \dots, n_k, \quad k = 1, 2, \dots, K,$$

kde  $\alpha_k$  vyjadřuje tzv. efekt  $k$ -té úrovně faktoru,  $\mu$  a  $\alpha_k$  jsou neznámé parametry,  $n_k$  značí počet pozorování při  $k$ -té úrovni faktoru. Hypotézu o neexistenci faktoru pak formulujeme jako složenou hypotézu  $H: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$  o nulových efektech všech úrovní faktoru. Test je založen na rozkladu celového součtu čtverců podle vztahu

$$Q_T = Q_B + Q_E, \quad (\text{vz. } 2-2)$$

kde  $Q_T$  značí celkový součet čtverců vyjadřující celkovou variabilitu,  $Q_B$  je meziskupinový součet čtverců popisující rozdíly v úrovni skupin a  $Q_E$  představuje vnitroskupinový součet čtverců vyjadřující náhodné kolísání způsobené jinými vlivy než změnami úrovně faktoru. Celkovou variabilitu proměnné lze tedy vyjádřit jako součet variability způsobené rozdíly v úrovni různých skupin a zbytkové variability, způsobené kolísáním, které se označuje jako náhodné. Čím větší podíl na celkové variabilitě má součet čtverců  $Q_B$ , tím víc lze předpokládat existenci vlivu zkoumaného faktoru.

Při normalitě rozdělení a shodě rozptylů v různých skupinách (homoskedascita), jsou veličiny  $y_{ki}$  rovněž nezávislé a při platnosti testované hypotézy  $H$  má statistika

$$F = \frac{(n-K)(Q_T - Q_E)}{(K-1)(Q_E)} = \frac{(n-K)(Q_B)}{(K-1)(Q_E)} \quad (\text{vz. } 2-3)$$

rozdělení  $F$  s  $(K-1)$  a  $(N-K)$  stupni volnosti. Překročí-li hodnota testované statistiky  $F$  kvantil  $f_{1-\alpha}(K-1, n-K)$ , prokázali jsme rozdílnou úroveň skupin, a tedy i vliv zkoumaného faktoru na hladině významnosti  $\alpha$ .

Předpoklad normality rozdělení a shody rozptylů v různých skupinách lze ověřovat pomocí testů, v praxi se často užívají grafy, které jsou součástí výstupů počítačových procedur.

### 3.2.3 Logistická regrese

Zpracováno podle [11].

Logistická regrese je metoda, navržená v 60. letech 20. stol. jako alternativní postup k metodě nejmenších čtverců pro případ, že vysvětlovaná (závislá) proměnná je binární. V minulosti byla tato metoda aplikována nejčastěji v oblasti medicíny a epidemiologie, kde binární proměnná 0,1 představuje přítomnost, respektive nepřítomnost choroby. Lze ji však využít i pro zpracování spojitých dat. Pro tuto práci byla vybrána s přihlédnutím k binárnímu formátu zdrojových dat.

Logistická regrese se liší od lineární regrese v tom, že predikuje pravděpodobnost, zda se daná událost stala nebo nestala. Vypočtená pravděpodobnost je tedy rovna 0 nebo 1. Aby se vytvořila vazebná podmínka, užívá logistická regrese tzv. logitovou transformaci, která vede na sigmoidální vztah<sup>3</sup> mezi závisle proměnnou  $y$  a vektorem nezávisle proměnných  $x$ . Rozdíl mezi logistickou a lineární regresí spočívá v tom, že logistická regrese používá kategorizovanou závisle proměnnou, kdežto lineární regrese užívá pouze spojitou závisle proměnnou.

#### Logistický regresní model

V logistické regresi je nutné vědět, zda se událost stala nebo nestala. Je-li predikovaná pravděpodobnost větší než 0,50, pak se událost stala. Je-li menší než 0,50, pak se nestala.

Postup vyčísľující logistické koeficienty porovnává pravděpodobnost události odehrané  $L_{(1)}$  vůči pravděpodobnosti události neodehrané  $L_{(0)} = 1 - L_{(1)}$  využitím pravděpodobnostního poměru  $L_{(1)}/L_{(0)}$ , ve kterém je pravděpodobnost  $L_{(1)}$  vyjádřena logistickou funkcí

$$L_{(1)} = \frac{1}{(1 + e^{c-z})} \quad (\text{vz. 3-1})$$

Pravděpodobnostní poměr lze vyjádřit vztahem

$$\frac{L_{(1)}}{L_{(0)}} = \exp(a_0 + a_1 x_1 + a_2 x_2 + \dots + a_p x_p), \quad (\text{vz.3-2})$$

---

3 Sigmoidální křivka má tvar písmene S. U nízkých hodnot nezávisle proměnné se její pravděpodobnost blíží nule, u vysokých hodnot se blíží jedné

kde odhadované koeficienty  $a_0, a_1, a_2, \dots, a_p$  jsou míry změny poměru pravděpodobností, které jsou lineární diskriminační funkce o  $p$  původních nezávisle proměnných

$$Z = a_0 + a_1 x_{a_1} + a_2 x_{a_2} + \dots + a_p x_{a_p}. \quad (\text{vz. 3-3})$$

Po dosazení, zlogaritmování výrazu a úpravě vyjde tzv. vícenásobný logistický regresní model

$$\ln\left(\frac{L_1}{L_0}\right) = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p, \quad (\text{vz. 3-4})$$

kde  $b_0 = -C + a_0$ ,  $b_i = a_i$ , pro  $i = 1, \dots, p$ .

Logistický model lze snadno rozšířit na případ  $K$  tříd.

### Odhady parametrů

Pro odhad parametrů logistického modelu se standardně používá metoda maximální věrohodnosti<sup>4</sup>. Pro obecný model se předpokládá multinomické rozdělení a pro standardní model se dvěma třídami rozdělení binomické. Výchozí data jsou ve formě vektoru  $y$  rozměru  $n \times 1$  a matice  $X$  rozměru  $n \times m$ .

### Interpretace regresních koeficientů

Základní předpoklad v logistické regresi říká, že logaritmus pravděpodobnostního poměru  $\ln(L_{(1)}/L_{(0)})$  je lineární funkcí nezávisle proměnných. Žádné předpoklady o rozdělení nezávisle proměnných  $x$  zde neexistují. Hlavní výhodou metody je fakt, že  $x$  mohou být jak diskrétní, tak i spojité veličiny. Vícenásobný logistický regresní model lze upravit do tvaru

$$L_0 = \frac{1}{1 + \exp[-(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p)]}. \quad (\text{vz. 3-5})$$

Kladné znaménko koeficientu  $b_i$  zvyšuje pravděpodobnost  $L_{(1)}$  a záporné ji snižuje. Je-li  $b_i$  kladné, funkce  $\exp$  je vyšší než 1 a pravděpodobnostní poměr  $L_{(1)}/L_{(0)}$  se zvyšuje.

4 K odhadu používá tzv. věrohodnostní funkci  $L(x_1, x_2, \dots, x_n, \theta)$ , kde  $x_1, x_2, \dots, x_n$  jsou pozorování náhodné veličiny a  $\theta$  je parametr rozdělení. Nejlepší odhad získáme pro takovou hodnotu  $\theta'$ , pro kterou dosahuje věrohodnostní funkce svého maxima<sup>[1]</sup>

$L_{(1)}$  se mění od 0 do 1 a pravděpodobnostní poměr  $L_{(1)}/L_{(0)}$  se mění od 0 do  $\infty$ .

### 3.2.4 Geografický střed

Geografický střed je jednoduchá metoda, která dokáže vypočítat aritmetický střed skupiny bodů. Vhodné použití je u datasetu, kdy nepotřebujeme všechny jednotlivé údaje, ale zajímá nás spíše jejich obecná charakteristika. Například máme-li několik skupin bodů, jejich geografické středy mohou ukázat, jak se liší jejich celkové rozložení, zatímco při zobrazení plného počtu prvků by tato informace byla skryta či zkreslena. Obecná rovnice výpočtu je

$$S_g = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{vz. 4-1})$$

kde  $n$  je počet lokalit a  $x_i$  je souřadnice prvku. Geografický střed se počítá zvlášť pro souřadnici X a Y. Alternativou je vážený geografický střed, který se vypočte následovně

$$S_g = \frac{\sum_{i=1}^n x_i \cdot m_i}{\sum_{i=1}^n m_i} \quad (\text{vz. 4-2})$$

kde  $m_i$  je váha jednotlivých prvků.

### 3.2.5 Shluková analýza

Zpracováno podle [18].

Pojem shluková analýza je souhrnným názvem pro celou řadu výpočetních postupů, jejichž cílem je rozklad daného souboru dat na několik relativně homogenních skupin, shluků. Rozklad množiny dat by měl být proveden takovým způsobem, aby si objekty uvnitř jednotlivých shluků byly co nejvíce podobné. Objekty patřící do různých shluků by si naopak měly být podobné co nejméně.

Jednou ze známých definicí pojmu shluk je definice Van Rijsbergenova:

„ Je dána množina objektů  $X = \{x_1, x_2, \dots, x_n\}$  a libovolný koeficient  $D$  nepodobnosti

objektů. Shlukem nazveme takovou podmnožinu  $A$  množiny objektů  $X$ , pro niž platí:

$$\max D(x_i, x_j) < \min D(x_k, x_l),$$

kde objekty  $x_i, x_j, x_l \in A$  a objekt  $x_k \notin A$ . “

### **Klasifikace metod shlukové analýzy**

Tradičně lze rozdělit metody shlukové analýzy na hierarchické a nehierarchické.

Pro hierarchické je společné to, že shlukovací proces má charakter posloupnosti rozkladů množiny objektů. Dle postupu algoritmu lze tyto metody dělit na aglomerativní a divizní.

V případě aglomerativních algoritmů je každý shluk tvořen právě jedním objektem. V následujícím kroku aglomerativního algoritmu jsou vybrány dva jednoprvkové shluky, které jsou si z jistého hlediska nejpodobnější. Tyto objekty se sloučí do nového shluku. Tento postup se opakuje do té doby, než shlukované objekty tvoří jediný shluk.

Naopak, vychází-li se z jednoho shluku, který se postupně rozkládá na vzájemně disjunktivní podmnožiny, hovoříme o metodách divizních.

Skupina nehierarchických shlukovacích metod nevytváří hierarchickou strukturu, ale rozděluje objekty do shluků tak, aby bylo splněno určité kritérium. Prvotní rozklad původní množiny objektů do několika podmnožin se dále mění. Cílem je dosažení optimální hodnoty jistého, pro danou metodu specifického kritéria.

### **Podobnost objektů**

V souvislosti s metodami shlukové analýzy je důležitý pojem podobnosti mezi jednotlivými statistickými jednotkami. U většiny shlukovacích postupů je však výhodnější využívat duálního pojmu míry podobnosti objektů, tj. míry nepodobnosti objektů. Lze se setkat s třemi základními typy popisu podobnosti - nepodobnosti - objektů. Prvním typem jsou koeficienty asociace. Druhým typem koeficienty korelace. Třetím typem jsou metriky. První dva typy řadíme k mírám podobnosti. Třetí typ řadíme k mírám nepodobnosti objektů.

Pro seskupení dat byla v bakalářské práci vybrána metoda Ward – Wishartova, která se řadí mezi metody hierarchické, jako míra nepodobnosti objektů byl použit čtverec euklidovské vzdálenosti.

### Ward – Wishartova metoda

V anglické literatuře také „Ward’s error sum of squares method“. Ward navrhl metodu, která je založena na ztrátě informací, plynoucí ze seskupování objektů do shluků na základě celkového součtu čtverců odchylek každého pozorování, od vektorů průměrů, tj. centroidu shluku, do něhož náleží. Wardova metoda tudíž používá kritérium kvality rozkladu. Kritériem je zde přírůstek celkového součtu čtverců odchylek objektů od vektoru průměrů příslušného shluku. Přírůstek je vyjádřen jako součet čtverců odchylek objektů od vektoru průměrů v nově vznikajícím shluku  $G (G = A \cup B)$ , od něhož jsou odečteny součty čtverců odchylek v obou zanikajících shlucích  $A, B$ . Přesněji zapsáno:

$$\Delta C = \sum_{i=1}^{n_G} \sum_{j=1}^p (x_{gij} - \bar{x}_{gj})^2 - \sum_{i=1}^{n_A} \sum_{j=1}^p (x_{aij} - \bar{x}_{aj})^2 - \sum_{i=1}^{n_B} \sum_{j=1}^p (x_{bij} - \bar{x}_{bj})^2 \quad (\text{vz. 5-1})$$

kde  $x_{gij}$  je hodnota  $j$ -té proměnné  $i$ -tého objektu v nově vzniklém shluku  $G$ ,  $n_G$  je počet objektů v tomto shluku. Symboly  $\bar{x}_{aj}$ , resp.  $\bar{x}_{bj}$ , resp.  $\bar{x}_{gj}$ , představují průměrné hodnoty  $j$ -té proměnné ve shlucích  $A$ , resp.  $B$ , resp.  $G$ .

Při tomto postupu jsou spojovány shluky, jejichž sloučením minimalizujeme kritérium  $\Delta C$ . Tento postup vede k odstraňování malých shluků a vytváření shluků přibližně stejné velikosti.

Wardova metoda je využívána při práci se stejnorozměrnými proměnnými.

## 4. CHARAKTERISTIKA SLEDOVANÝCH JEVŮ

### 4.1 Charakteristika území

Zájmovým územím pro bakalářskou práci je oblast CHKO Bílé Karpaty, která pokrývá orografický celek Bílé Karpaty a části přilehlých celků Vizovická vrchovina a Dolnomoravský úval. Pohoří Bílé Karpaty tvoří nejzápadnější část karpatského horského systému a je nejvyšším pohořím jeho jihozápadního okraje. Nejvyšším vrcholem je Velká Javořina (970 m n.m.)<sup>[8]</sup>. Orientace oblasti je severovýchod-jihozápad a délka české části CHKO je 70 km<sup>[10]</sup>. Celková rozloha CHKO je 715 km<sup>2</sup> (747 km<sup>2</sup> podle zjištění geografickým informačním systémem<sup>[3]</sup>).

Území CHKO Bílé Karpaty je součástí flyšového<sup>5</sup> pásma Západních Karpat, které vzniklo v důsledku horotvorných pohybů alpského vrásnění. Obsah vápencových zrn v usazeninách se projevuje zejména přítomností pěnovců na prameništích a bohatším druhovým složením rostlinstva<sup>[12]</sup>.

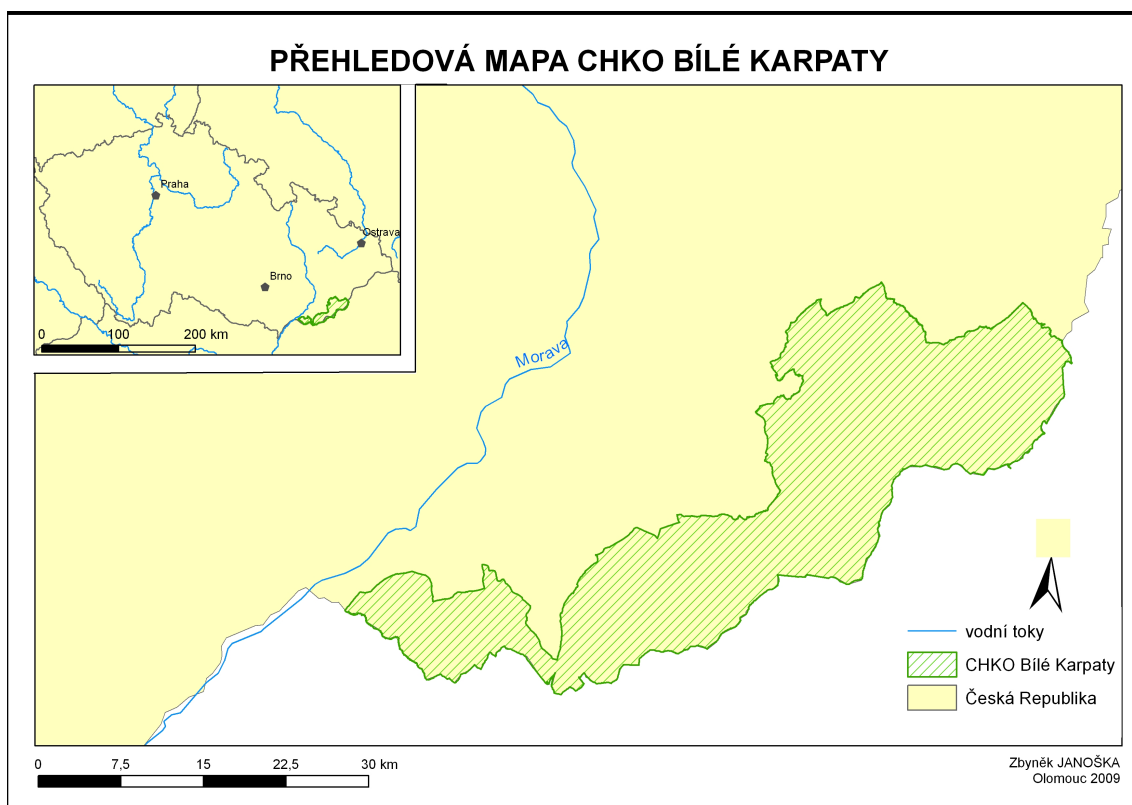
Charakteristickým znakem pro celé pohoří flyšového pásma jsou časté sesuvy v místech měkkých hornin<sup>[12]</sup>.

Celá oblast, ale zejména její jižní část, byla po staletí ovlivňována člověkem, což dalo vzniknout unikátním přírodním společenstvím, jejichž existence je vázána na tradiční způsoby zemědělství a obhospodařování krajiny. Historie osídlování Bílých Karpat sahá až do starší doby kamenné (od počátku lidských dějin až přibližně do 9 000 let př. n. l.). Hlavní vlna kolonizace, kdy bylo založeno velké množství současných sídel (především v jižní části CHKO) začala v 13. stol. V 18. stol. pak proběhlo osídlení i jádrových oblastí pohoří<sup>[12]</sup>. Původně téměř zcela zalesněná oblast se otevřela vznikem pastvin, luk, polí a sídel. Zejména ukázkový vztah mezi člověkem a přírodou byl důvodem pro zařazení CHKO Bílé Karpaty mezi biosférické rezervace UNESCO v roce 1996<sup>[3]</sup>.

Mezi nejhodnotnější oblasti patří z hlediska biodiverzity rozsáhlé, druhově bohaté orchideové louky v jihozápadní části a také staré bučiny v oblasti Vlárského průsmyku.

---

5 flyš = mnohonásobné střídání vrstev jílovců, prachovců, pískovců a slepenců



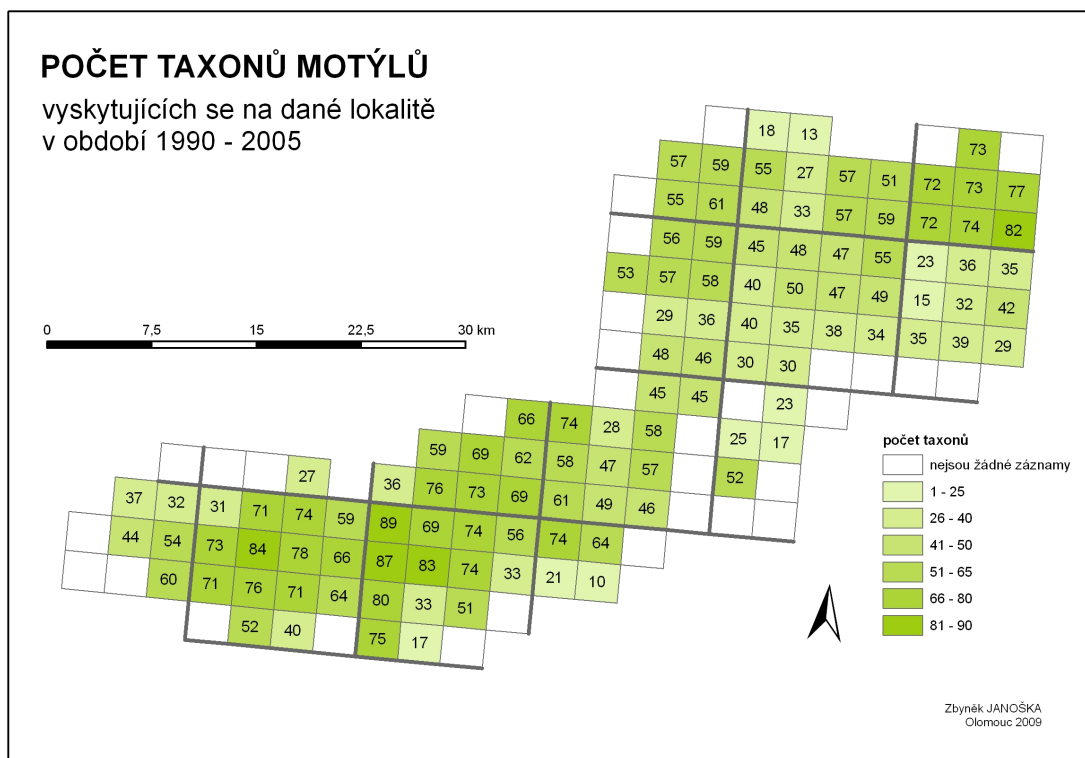
Obr. č.13 – přehledová mapa CHKO Bílé Karpaty

## 4.2 Motýli

Motýli (*Lepidoptera*) jako hmyzí řád jsou na území České republiky zastoupeni počtem 3411 druhů<sup>[19]</sup>, z oblasti CHKO Bílé Karpaty byl publikován výskyt 1964<sup>[3]</sup>. Na území Bílých Karpat se jedinečně vyskytují na stejných lokalitách teplomilné druhy charakteristické pro panonskou bioprovincii a druhy podhorského až horského charakteru, určující pro karpatskou horskou soustavu. První snahy o zmapování motýlů na území spadají na české straně pohoří do začátku 20. stol., na slovenské straně dokonce do století 19., nicméně první systematické mapování bylo uskutečněno až v letech 2003 – 2006 a jeho výsledky byly publikovány v práci *Atlas rozšíření vybraných druhů živočichů CHKO Bílé Karpaty*<sup>[3]</sup>. Pro projekt *Analýza biodiverzity v CHKO Bílé Karpaty jako podklad pro stanovení nové zonace a vhodného managementu cenných území* (VaV/620/12/03) bylo vybráno ke sledování celkem 132 druhů denních motýlů (*Rhopacelora*) a dále dva druhy recentně pozorované v rámci mapování Natura 2000. Pro mapování byla použita KFME síť druhého dělení (obr. 11). V letech 2003 – 2006 byl každý čtverec terénním průzkumem sledován minimálně 6x ročně v období od dubna do září včetně. Z vybraných druhů bylo zaznamenáno 114, zbylých 18 je považováno



pro oblast za vyhynulé. Kromě sledovaných druhů bylo zjištěno 7 druhů pro oblast nových. Z 11 recentně sledovaných druhů chráněných v rámci soustavy Natura 2000 bylo zjištěno 8.



Obr. č.14 – počet taxonů motýlů zjištěných na daných subkvadrátech

### 4.3 Cévnaté rostliny

Na území moravské strany Bílých Karpat zasahují 4 fytogeografické okresy a to Dolnomoravský úval, Bílé Karpaty stepní, Bílé Karpaty lesní a Javorníky. Dosavadní floristický výzkum CHKO Bílé Karpaty prokázal výskyt téměř 1500 druhů cévnatých rostlin, z nichž 30 je dnes považováno za vyhynulé, 50 druhů je nezvěstných. Bílé Karpaty jsou z hlediska floristiky nesmírně rozmanitou oblastí, což je dáno především rozdíly v historickém využívání krajiny v různých oblastech.

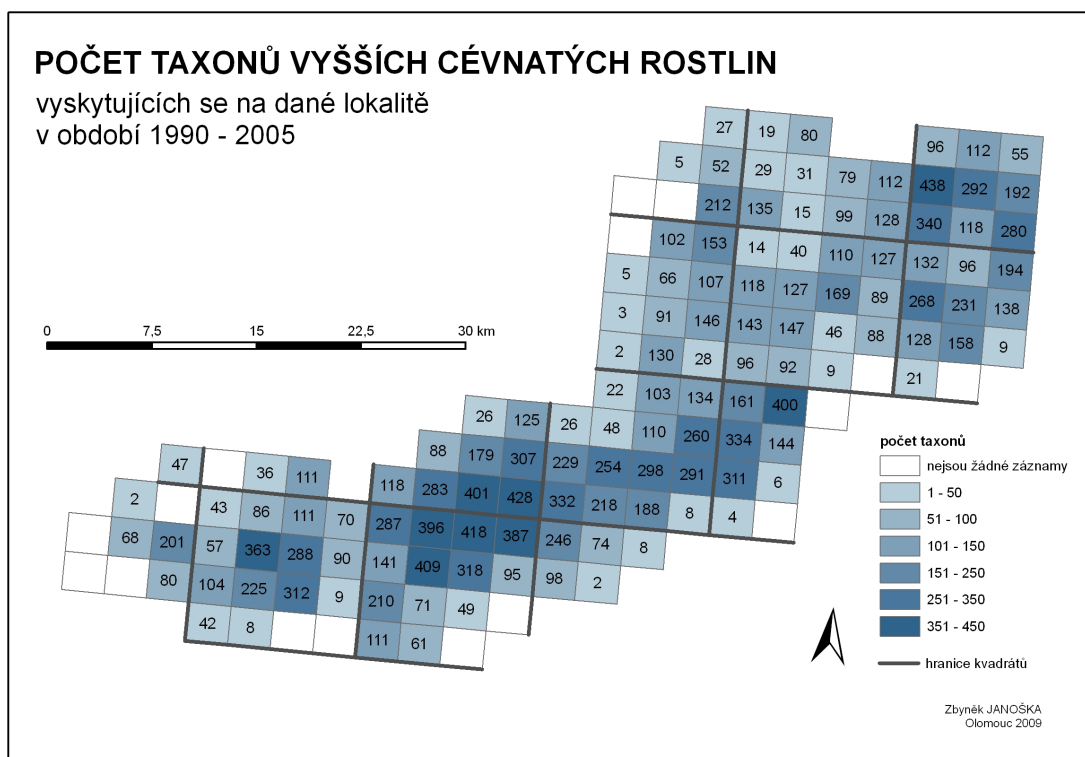
Začátky botanického mapování v Bílých Karpatech spadají do poloviny 19. stol., ale až ve 20. letech 20. století zejména zásluhou S. Staňka byla vegetace podrobně a systematicky zkoumána. Po vyhlášení CHKO Bílé Karpaty v roce 1980<sup>[6]</sup> byla věnována stále větší pozornost věnována výzkumu zachování rozmanitosti flóry. V rámci NATURA 2000 bylo také zahájeno rozsáhlé mapování, jehož výsledkem byl návrh 19 evropsky významných lokalit (u 12 z nich je předmětem ochrany luční biotop)<sup>[6]</sup>.

Nejcharakterističtějším nelesním typem vegetace Bílých Karpat jsou květnaté orchidejové louky, které patří k druhově nejbohatším typům lučních porostů ve střední Evropě. V CHKO Bílé Karpaty se vyskytuje (resp. vyskytovalo) 41 z celkového počtu 56 druhů vstavačovitých, doložených z území České republiky. Zvláštností jsou vzácná plevelová společenstva v oblastech, kde přetrval původní způsob obdělávání políček a záhumenek bez přílišného hnojení a mechanizace.

Lesy pokrývají téměř 45 % plochy CHKO, z toho listnaté lesy jsou zastoupeny necelými 55 %.

V CHKO Bílé Karpaty se vyskytuje celkem 103 chráněných druhů cévnatých rostlin, z toho 27 v kategorii kriticky ohrožených, 37 silně ohrožených a 39 v kategorii ohrožených druhů<sup>[15]</sup>.

Data, použitá v této práci, pocházejí z databáze floristických záznamů, která vznikla v rámci projektu *Analýza biodiverzity v CHKO Bílé Karpaty jako podklad pro stanovení nové zonace a vhodného managementu cenných území* (VaV/620/12/03). Jejím základem jsou pozorování uskutečněná v rámci projektu v letech 2003 – 2006, doplněná o informace z dalších recentních i historických zdrojů<sup>[5]</sup>.



Obr. č.15 – počet taxonů vyšších cévnatých rostlin zjištěných na daných subkvadrátech

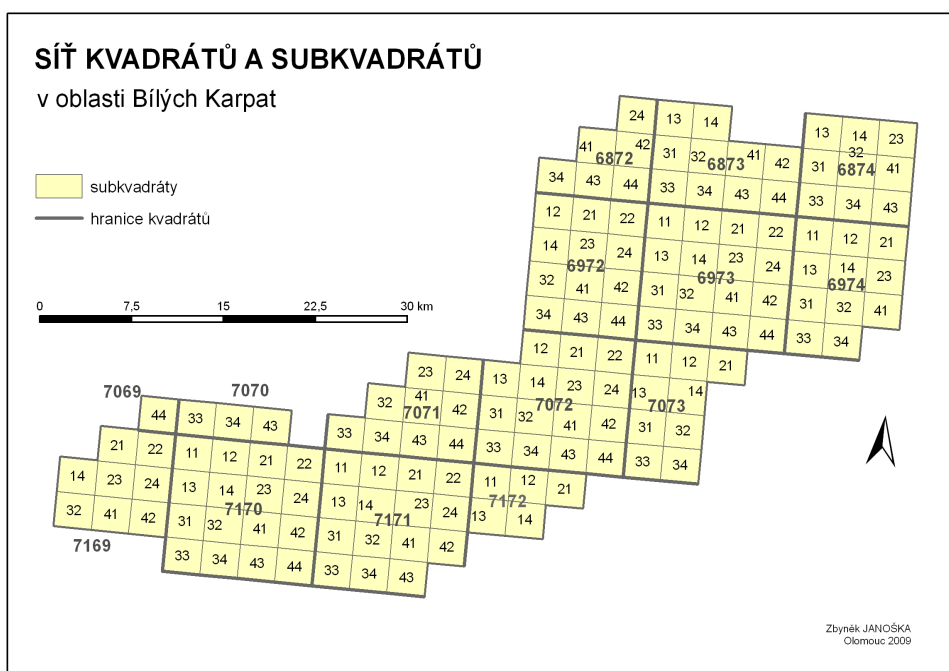
## 5. STATISTICKÉ VYHODNOCENÍ DAT

### 5.1 Zdrojová data

Veškerá data, která byla při práci použita, pochází z datového skladu CHKO Bílé Karpaty.

Data prostorově pokrývají území CHKO a přilehlé oblasti, které jsou z hlediska výskytu sledovaných jevů významné. Jako zdrojová data posloužila:

- polygonová vrstva kvadrátů síťového mapování dle standardu KFME – celé území je rozděleno do kvadrátů, ty pak dále do subkvadrátů. Subkvadráty byly v práci použity jako základní územní jednotka.



Obr. č.16 – síť kvadrátů a subkvadrátů pokrývajících zájmové území

- tabulky botanických pozorování – jedná se o tabulky s výskyty rostlin na zájmovém území. Jako zdrojová data posloužilo 13 tabulek se záznamy o výskytech taxonů vyšších cévnatých rostlin. Záznamy jsou z let 1916 až 2004. Pro zpracování byly vybrány jen záznamy z let 1990 – 2004 včetně. Údaje se většinou vztahují ke konkrétním zeměpisným souřadnicím, nikoli ke kvadrátům.

b_cevnt.dbf						
<i>Id_tax</i>	<i>Pocet</i>	<i>Lokalita</i>	<i>X'</i>	<i>Y'</i>	<i>Prasnost</i>	<i>Pokryvnost</i>
AGREUP_Zzul	0	7071-34	-532068	-1199104	2065	
ARRELA_iLM8	0	7071-34	-532068	-1199104	2065	
ASTDAN_inlpu	0	7071-34	-532068	-1199104	2065	
ASTMAJ_M80	0	7071-34	-532068	-1199104	2065	
BETOFF_yPpft	0	7071-34	-532068	-1199104	2065	
BRAPIN_m9ZD	0	7071-34	-532068	-1199104	2065	
CAMGLO_wMi	0	7071-34	-532068	-1199104	2065	
CAMPER_YBH	0	7071-34	-532068	-1199104	2065	
CENPHR_OUm	0	7071-34	-532068	-1199104	2065	
CENSCA_MyKf	0	7071-34	-532068	-1199104	2065	
CENERY_hLOf	0	7071-34	-532068	-1199104	2065	
TANCOR_AiZti	0	7071-34	-532068	-1199104	2065	

Obr. č.17 – ukázka zdrojových dat – tabulka botanických pozorování

Mezi atributy patří označení taxonu, lokalita, počet výskytů, datum pozorování a další.

Celkový počet záznamů použitých v práci přesahuje 60 000, pro vlastní analýzu ale byly použity jen takové rostliny, které podléhají režimu ochrany. Těchto rostlin je 172.

- tabulky lepidopterologických pozorování – databáze pozorování výskytu motýlů v zájmovém území. Jedná se o dvě tabulky s pozorováními z let 2002 až 2005. V práci byla použita všechna dostupná data. Záznamy se vztahují většinou ke středu subkvadrátů, nikoli ke konkrétním zeměpisným souřadnicím.

a_lepid1.dbf						
<i>Id_tax</i>	<i>Pocet</i>	<i>Lokalita</i>	<i>X'</i>	<i>Y'</i>	<i>Prasnost</i>	<i>File_lok</i>
MARQUE_JWf	1	7170	-543382	-1204995	6000	q01_71
MARQUE_JWf	0	7171	-531207	-1206166	6000	q01_71
SYNSTO_BkfM	1	Radějov u Strážnice	-547283	-1205743	5043	i_spravd
SYNSTO_BkfM	1	Velká nad Veličkou	-534926	-1201114	3992	i_spravd
LEPREA_nGp?	0	Radějov u Strážnice	-547283	-1205743	5043	i_spravd
COLERA_B3gá	0	Petrov	-603112	-1122241	1910	i_spravd
COLERA_B3gá	0	Strážnice na Moravě	-550640	-1199368	5420	i_spravd
COLERA_B3gá	0	Strážnice na Moravě	-550640	-1199368	5420	i_spravd
COLERA_B3gá	0	Tvarožná Lhota	-545898	-1203347	5383	i_spravd
COLERA_B3gá	40	Tvarožná Lhota	-545898	-1203347	5383	i_spravd
IDADEG_u7NL	0	Lopeník	-515463	-1197881	2855	i_spravd
HORRAD_Hny	0	Lopeník	-515463	-1197881	2855	i_spravd

Obr. č.18 – ukázka zdrojových dat – tabulka výskytu motýlů

Mezi atributy patří označení taxonu, lokalita pozorování, počet výskytů, datum pozorování a další.

Dohromady se jedná o více než 20 000 záznamů o výskytech motýlů, které byly v práci použity.

## 5.2 Zpracování dat

Nejprve bylo třeba z nálezových dat odstranit ta, kterým chyběl některý z následujících atributů – lokalita, datum pozorování. Data, u kterých byl počet pozorování nulový, byla brána jako data binární, kde 0 znamená přítomnost druhu. U lepidopterologických pozorování se taková data nevyskytovala, nicméně 12 z 13 tabulek pro botanická pozorování obsahovala právě tento typ záznamů. Proto bylo třeba i veškerá lepidopterologická pozorování upravit do tohoto formátu.

Po standardizaci dat byl každému záznamu přiřazen odpovídající subkvadrát. V rámci subkvadrátů byla data seskupena do tří kontingenčních tabulek (jedna pro botanická pozorování a dvě pro lepidopterologická pozorování). Jeden rozměr v tabulce byla lokalita – subkvadrát, druhý rozměr byla identifikace taxonu. Obsahem jsou pro jednotlivé tabulky:

- tabulka výskytů rostlin – pro každý subkvadrát hodnota 1, pokud rostlina na lokalitě byla pozorována, nebo 0, pokud pozorována nebyla
- tabulka výskytů motýlů – pro každý subkvadrát hodnota 1, pokud motýl na lokalitě byl pozorován, nebo 0, pokud pozorován nebyl
- tabulka počtů motýlů – pro každý subkvadrát celkový počet jedinců pozorovaný na jeho území

motyli_vyskyty_vsechny_q.dbf					
Lokalita	N68872_24	N68872_41	N68872_42	N68872_43	N68872_44
AGLURT_aKxi	0	3.00	4.00	2.00	1.00
ANTCAR_IUKTz	1	9.00	38.00	16.00	46.00
APAILI_4guDZ	0	0.00	0.00	0.00	2.00
APAIRI_w80Bh	0	0.00	1.00	0.00	2.00
APHHYP_tixzZ	0	7.00	8.00	1.00	15.00
ARALEV_s6ck6	0	19.00	6.00	9.00	14.00
ARGADI_ZwbDP	0	0.00	0.00	0.00	0.00
ARGAGL_Dh6p3	0	1.00	1.00	0.00	3.00
ARGNIO_qHXw8	0	0.00	0.00	0.00	0.00
ARGPAP_aKK85	0	12.00	38.00	12.00	63.00
ARIAGE_UqWIL	0	4.00	3.00	2.00	3.00
ARIART_68crf	0	0.00	0.00	0.00	0.00

Obr. č.17 – ukázka kontingenčních tabulek – spojená data

botanika_vyskyty_vsechny_q.dbf					
id_tax_c_1	N6872_24	N6872_41	N6872_42	N6872_43	N6872_44
ABIALB_lfuQJ	0	1.00	1.00	0	1.00
ABISP_Rbegi	0	0.00	0.00	0	0.00
ACECAM_yy4Jj	0	0.00	0.00	0	1.00
ACENEG_vONLY	0	0.00	0.00	0	0.00
ACEPLA_VqRFq	0	0.00	0.00	0	1.00
ACEPSE_NtoKj	0	0.00	0.00	0	1.00
ACESP_hhQHP	0	0.00	0.00	0	0.00
ACOCAL_SKUmW	0	0.00	0.00	0	0.00
ACOLYC_apfey	0	0.00	0.00	0	0.00
ACOLYC_G5HHI	0	0.00	0.00	0	0.00
ACTSPI_5zmhU	0	0.00	0.00	0	0.00
ADDAES_lgt2	0	0.00	0.00	0	0.00

Obr. č.18 – ukázka kontingenčních tabulek – binární data

### 5.3 Sestavení modelu

Výše zmíněné tabulky byly statisticky vyšetřeny s účelem prokázat/zamítnout závislost výskytu motýlů. Dále bylo přikročeno ke konstrukci modelu, který popisuje závislost prostorové distribuce motýlů na výskytu rostlin.

#### 5.3.1 $\chi^2$ -test

Základním předpokladem pro jakékoli další vyšetření dat je úvaha, že výskyt sledovaných jevů na území se neděje náhodně, ale že podléhá nějakému řádu. Pokud by byl výskyt jevů zcela náhodný, nebylo by možné jej modelovat.

K vyšetření nezávislosti proměnných se používá  $\chi^2$ -test, který podává informaci o tom, je-li výskyt náhodný či ne. Hlavním cílem práce je zjistit závislosti výskytu motýlů – proto byly nálezy motýlů vyšetřeny pomocí  $\chi^2$ -testu v prostředí R.

Použito bylo příkazu *chisq.test()* s defaultním nastavením parametrů.

P-hodnota	< 2,2.10 <sup>-16</sup>
Hladina významnosti	0,01
Závěr	Nezávislost proměnných se zamítá

Výsledkem je zjištění, že závislost výskytů motýlů na lokalitách je prokázána. Jejich distribuce je tedy nenáhodná a má smysl analyzovat její závislost na výskytu rostlin.

Není nutné provádět  $\chi^2$ -test pro tabulku výskytu motýlů (binární data), protože se jedná jen o zjednodušený zápis stejných dat, jaká již byla analyzována.

Zároveň je zbytečné testovat výskyty rostlin. Předmětem analýzy je zjistit závislost

výskytu motýlů na výskytu rostlin.  $\chi^2$ -test tuto závislost potvrdit, a ať již je výskyt rostlin na sledovaném území náhodný či ne, na výsledky práce to nebude mít vliv. Mělo by smysl testovat výskyty rostlin v případě, že zjišťujeme závislost jejich výskytu např. na přírodních faktorech.

V této úloze ale není podstatné, zda je distribuce taxonů rostlin v prostoru náhodná či ne.

### 5.3.2 ANOVA

ANOVA (*Analysis Of Variance*, analýza rozptylu) je metoda, která umožňuje ověřit, zda na hodnotu náhodné veličiny má statisticky významný vliv hodnota některého znaku, který se dá u veličiny zpozorovat<sup>[2]</sup>. V tomto konkrétním případě byl pomocí ANOVY vyšetřen vliv jednotlivých druhů motýlů na průměrný výskyt na lokalitě, a dále pak vliv jednotlivých lokalit na průměrný výskyt daného taxonu motýla. Bylo použito jednoduché třídění s jedním faktorem. Spouštěcím příkazem v prostředí R je `aov()` při defaultním nastavení.

Jednoduše řečeno, bylo zjišťováno, zda každá jednotlivá lokalita má vliv na průměrný výskyt motýlů. U všech lokalit byla překročena p-value<sup>6</sup>, což znamená, že každá jedna lokalita má na průměrný výskyt vliv.

Dále byl pomocí ANOVY vyšetřen vliv výskytu jednotlivých taxonů na průměrný výskyt na každé jedné lokalitě. U naprosté většiny motýlů byla p-value překročena, několik z nich ale tuto podmínku nesplnilo. Tito motýli se buďto vyskytují na území rovnoměrně a ve vysokých počtech, nebo se vyskytují pouze v malém množství na velmi malém počtu lokalit.

### 5.3.3 Logistická regrese

Jakmile byla prokázána vzájemná závislost mezi výskyty sledovaných jevů, mohlo být překročeno k sestavení modelu. V potaz bylo bráno více statistických metod. Nakonec byla vybrána metoda logistické regrese, která byla navržena jako alternativa k metodě nejmenších čtverců pro případ, že vysvětlovaná (závisle) proměnná je binární<sup>[11]</sup>. Právě tato vlastnost byla rozhodující, protože data výskytů rostlin jsou v binárním formátu.

V R je spouštěcím příkazem `glm()`. Vstupními daty jsou vysvětlovaná (závislá)

---

6 p-value je míra, která popisuje, kolik evidence máme pro zamítnutí nulové hypotézy. Tradičně se volí zamítací hodnota 0,05 (tedy pokud je p-value nižší, zamítáme nulovou hypotézu ve prospěch alternativy). Čím nižší je p-value, tím větší je pravděpodobnost zamítnutí nulové hypotézy.

proměnná a libovolný počet vysvětlujících (nezávislých) proměnných.

Výstupem je pak vektor obsahující jednu konstantu a seznam koeficientů pro každou ze vstupních nezávislých proměnných.

Vzhledem k velkému počtu vstupních dat se ukázalo jako neúnosné modelovat závislost výskytu na každém jednotlivém botanickém taxonu. Model by pak obsahovat více než 1000 parametrů, jejich hodnoty by byly nízké a zvýšila by se možnost vnesení chyby.

Způsob, jak se s velkým počtem dat vyrovnat, je rozdělit vstupní data do skupin a modelovat pak závislost výskytu motýlů ne na jednotlivých taxonech, ale na skupinách příbuzných taxonů.

#### 5.3.4 Seskupení dat

Redukce dat je účinný nástroj vícerozměrné analýzy dat. V této práci mohlo být zvoleno mezi dvěma postupy – seskupení dat na základě botanické podobnosti a seskupení dat na základě statistického vyšetření. Prvním krokem byl výběr pouze těch taxonů rostlin, které podléhají určitému režimu ochrany. Těchto druhů se na zájmovém území vyskytuje 172 a jsou indikátory kvality ekosystému.

Při tvorbě shluků bylo rozhodováno mezi několika metodami, jak data seskupit. Byla uvažována metoda, kdy by se data rostlin seskupila podle diagnostických druhů vegetačních tříd a svazů<sup>[4]</sup>. Tato metoda byla zamítnuta, protože počet diagnostických druhů je velmi malý a většina informace obsažené v datech výskytů rostlin by se tak ztratila.

Další uvažované metody byly již čistě statistické a vycházejí z teorie shlukové analýzy, které se věnuje velké množství publikací, například <sup>[11]</sup>.

#### **Seskupení na základě geografického středu**

Poněvadž každý taxon (botanický i zoologický) se vyskytuje na mnoho lokalitách, je jakékoli seskupování velmi obtížné a v rámci práce bylo přistoupeno ke snížení počtu rozměrů dat. Pro každý taxon (jak botanický tak zoologický) byl spočítán geografický střed jeho výskytů pomocí skriptu napsaného v *avenue*. Takto byl každému taxonu přiřazen jeden bod a tyto body byly dále seskupeny.

Jako metoda seskupování byla použita metoda Ward – Wishartova. Tato metoda minimalizuje ztrátu informací, které se při seskupování nelze vyhnout a je vhodná k seskupování dat se stejnorozměrnými proměnnými. Její nevýhodou je eliminace malých



shluků a jejich připojení ke shlukům početnějším.

Vlastní shluková analýza byla provedena v trial verzi softwaru QC.Expert 3.0, použitou metrikou byl čtverec euklidovské vzdálenosti, který je pro Ward – Wishartovu metodu doporučován jako nejvhodnější.

Výsledkem byl protokol, který obsahoval vždy původní shluk, nový shluk a míru nepodobnosti mezi nimi. Jako mezní hodnota byla zvolena hodnota 0,6. Pokud míra nepodobnosti mezi dvěma shluky dosáhla této hodnoty, nebyly už tyto shluky seskupeny. Míra nepodobnosti 0,6 byla zvolena proto, že při této hodnotě se začínaly spojovat i větší shluky a tím pádem se rychle snižovala obsažená informace.

The screenshot shows the QC.Expert 3.0 software window with a menu bar (Soubor, Úpravy, Formát, QC.Expert, Uživatelské analýzy, Okno, Nápověda) and a toolbar. The main area displays a table with columns A through H and rows 1 through 24. The table contains data for a dendrogram analysis, including object IDs, distances, and similarity values.

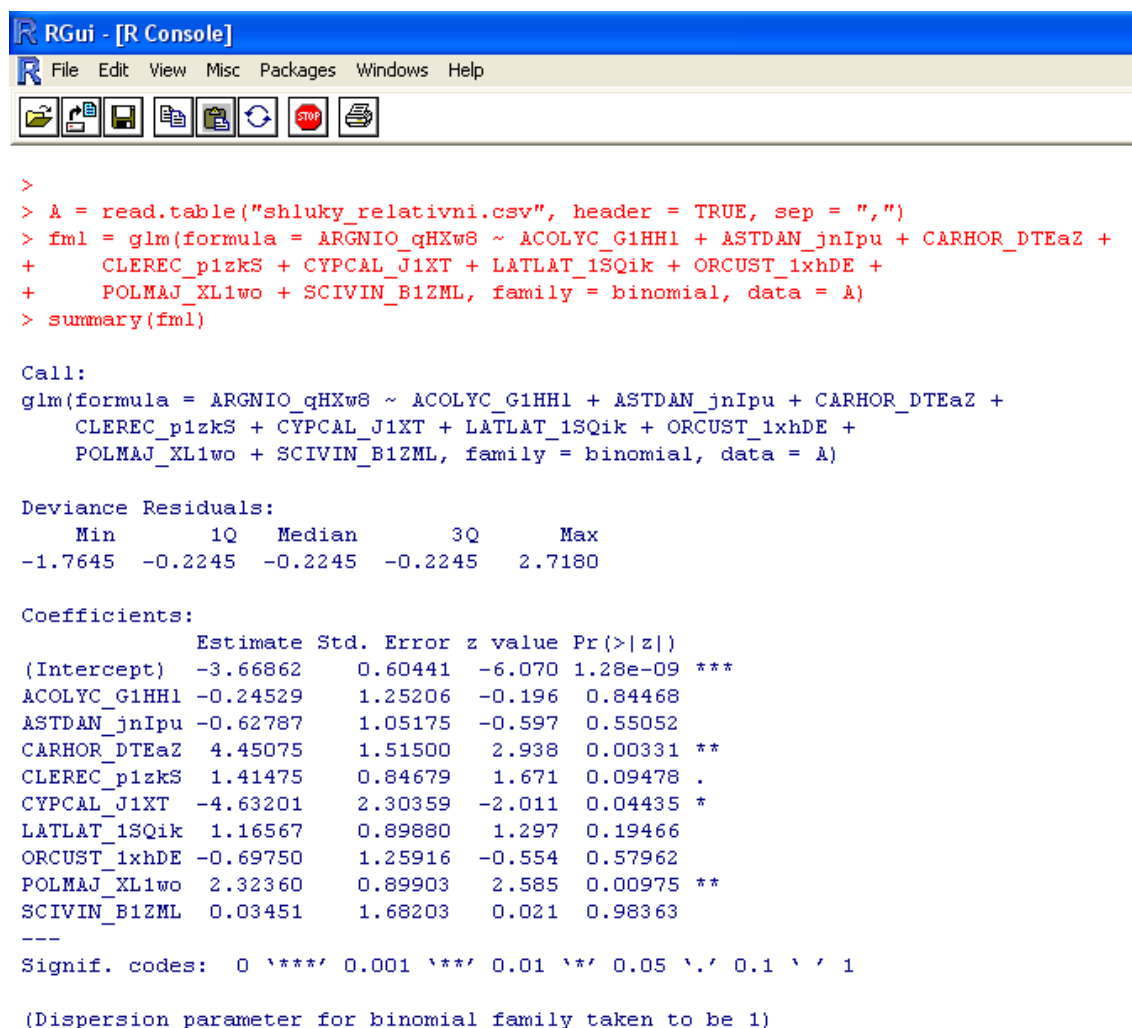
	A	B	C	D	E	F	G	H
1								
2	Dendrogram							
3								
4	Typ vzdálenosti	Čtverec Eukleid. vzd.						
5	Metoda	Ward's Algorithm						
6								
7	Obj.1	Obj.2	Nový shluk	Vzdálenost				
8	49	101	296	0				
9	55	291	297	0				
10	64	110	298	0				
11	298	115	299	0				
12	299	200	300	0				
13	220	254	301	0				
14	252	264	302	0				
15	96	152	303	0				
16	143	175	304	0				
17	20	41	305	0				
18	148	189	306	0				
19	28	181	307	0,0001				
20	221	262	308	0,0001				
21	59	286	309	0,0001				
22	2	173	310	0,0001				
23	92	269	311	0,0002				
24	40	194	312	0,0002				

Obr. č.19 – ukázka protokolu shlukové analýzy

Výsledkem bylo 40 shluků, mezi nimiž ten nejpočetnější obsahoval 18 taxonů, ten nejméně početný pouze jediný. Z těchto shluků byly odstraněny takové, které obsahovaly buďto jen botanické, nebo jen zoologické taxony. Konečný počet shluků byl tedy 28.

### 5.3.5 Modelování

Vlastní modelování bylo prováděno v rámci každého shluku v prostředí R za pomoci metody logistické regrese. Spouštěcím příkazem je *glm()* a parametr *family* byl nastaven jako *binomial*.



```
>
> A = read.table("shluky_relativni.csv", header = TRUE, sep = ",")
> fml = glm(formula = ARGNIO_qHXw8 ~ ACOLYC_G1HH1 + ASTDAN_jnIpu + CARHOR_DTEaZ +
+ CLEREC_p1zkS + CYPICAL_J1XT + LATLAT_1SQik + ORCUST_1xhDE +
+ POLMAJ_XL1wo + SCIVIN_B1ZML, family = binomial, data = A)
> summary(fml)

Call:
glm(formula = ARGNIO_qHXw8 ~ ACOLYC_G1HH1 + ASTDAN_jnIpu + CARHOR_DTEaZ +
     CLEREC_p1zkS + CYPICAL_J1XT + LATLAT_1SQik + ORCUST_1xhDE +
     POLMAJ_XL1wo + SCIVIN_B1ZML, family = binomial, data = A)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7645  -0.2245  -0.2245  -0.2245   2.7180

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.66862    0.60441  -6.070 1.28e-09 ***
ACOLYC_G1HH1 -0.24529    1.25206  -0.196  0.84468
ASTDAN_jnIpu -0.62787    1.05175  -0.597  0.55052
CARHOR_DTEaZ  4.45075    1.51500   2.938  0.00331 **
CLEREC_p1zkS  1.41475    0.84679   1.671  0.09478 .
CYPICAL_J1XT -4.63201    2.30359  -2.011  0.04435 *
LATLAT_1SQik  1.16567    0.89880   1.297  0.19466
ORCUST_1xhDE -0.69750    1.25916  -0.554  0.57962
POLMAJ_XL1wo  2.32360    0.89903   2.585  0.00975 **
SCIVIN_B1ZML  0.03451    1.68203   0.021  0.98363
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

Obr. č.20 – logistická regrese v prostředí R

Do logistické regrese vstupují data v binárním formátu (výskyty rostlin) nebo ve formátu nominálním s tím, že nabývají hodnot 0-1 (takto byla upravena tabulka výskytů motýlů, aby byla co nejvíce zachována původní informace<sup>7</sup>). Prvním vstupním parametrem

7 Pro každý taxon byla nalezena nejvyšší hodnota mezi nálezy a touto hodnotou vydělena každá jednotlivá hodnota na každé lokalitě. Vzniklá matice nabývala hodnot od 0 do 1. Protože logistická regrese počítá hodnoty 0 – 0,5 jako nevýskyt (pravděpodobnost výskytu menší než 0,5) a hodnoty vyšší než 0,5 jako výskyt (pravděpodobnost výskytu vyšší než 0,5), byla ke každé nenulové hodnotě v nové vypočtené matici přičtena hodnota 0,5. Takto vznikla matice, kde nevýskyt byl zaznamenán jako 0 a výskyt nabýval hodnot 0,5 až 1 podle relativního počtu nálezů na každé lokalitě.

je závislá proměnná, v tomto případě taxon motýla, dále za znaménkem  $\sim$  následují jednotlivé taxony rostlin. Všechny vstupní parametry byly zadávány ve formě vektorů.

Vlastní rovnici si pak lze představit následovně:

$$\text{motýl}_1 \sim \text{rostlina}_1 + \text{rostlina}_2 + \dots + \text{rostlina}_n$$

Výsledkem analýzy jsou hodnoty koeficientů pro každý vstupní parametr. Parametr *Intercept* je konstantou. Dále je výsledkem míra spolehlivosti pro každý parametr. Najdeme ji ve sloupci *Pr* ( $> |z|$ ). Jedná se o p-value, čili čím nižší hodnota, tím spolehlivější předpověď.

Celkem byly sestaveny modely pro 109 taxonů motýlů. Z těchto taxonů pouze 22 splňovalo podmínku, že alespoň tři parametry modelu musí být stanoveny s přesností alespoň 90% a pouze u devíti modelů bylo dosaženo této přesnosti u čtyř anebo více parametrů. Areály těchto devíti taxonů byly dále modelovány a zobrazeny v mapách.

## 6. SIMULACE

S pomocí logistického modelu bylo možné pro každý jednotlivý taxon motýlů vypočítat pravděpodobnost jeho výskytu na každé lokalitě. Vizualizovány byly předpokládané areály výskytu devíti taxonů, které splňovaly podmínku, aby alespoň čtyři parametry modelu byly určeny se spolehlivostí vyšší než 90%.

*Modely závislosti vybraných taxonů motýlů:*

$$\text{POLSEM} = -1,1786 + 0,2451*\text{CREPRA} + 1,3697*\text{ERILAT} + 1,0796*\text{ORCMAS}$$

$$\text{BOLSEL} = -0,4577 + 2,0960*\text{ANTRAM} + 1,1774*\text{CAROTR} - 0,9884*\text{CORMAS}$$

$$\text{ARIEUM} = -2,32079 + 1,27456*\text{ORCMOR} + 1,23203*\text{PLABIF} + 1,29577*\text{RANARV}$$

$$\text{COLSP} = -1,5759 + 0,9067*\text{CEPLON} + 1,4440*\text{GENCRU} + 4,4019*\text{MELARV} + 1,4872*\text{TRIALP}$$

$$\text{HETMOR} = -1,0024 + 1,4488*\text{CHAVIR} + 1,0597*\text{LITPUR} + 1,5329*\text{PRUGRA}$$

$$\text{ARGNIO} = -3,66862 + 4,45075*\text{CARHOR} + 1,41475*\text{CLEREC} - 4,63201*\text{CYPCAL} + 2,32360*\text{POLMAJ}$$

$$\text{ARISP} = -3,0993263 + 2,3193028*\text{ASTDAN} + 2,1191592*\text{CLEREC} - 4,5365230*\text{CYPCAL}$$

$$\text{BREINO} = -1,4261 + 1,8157*\text{ASTDAN} + 1,1478*\text{CLEREC} + 1,1849*\text{LATLAT}$$

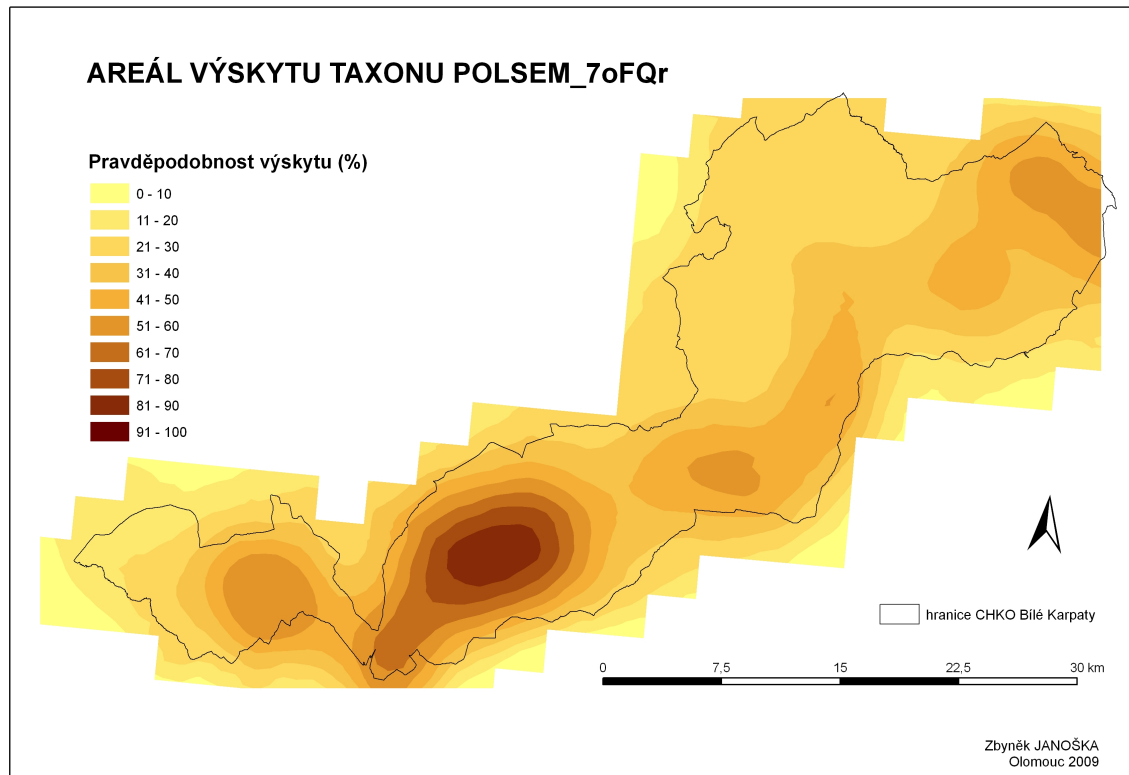
$$\text{MELATH} = -0,6645 + 1,6040*\text{EPIVIR} + 1,2957*\text{GERSAN} + 1,1974*\text{MELMEL}$$

U všech modelů uvedených výše je konstanta (první parametr) záporná. To značí, že základní pravděpodobnost výskytu je menší než 50%. Většina dalších parametrů má kladné znaménko, čili s výskytem taxonu rostliny se zvyšuje i pravděpodobnost výskytu motýla. U dvou taxonů rostlin (ORMAS a CYPCAL) jsou znaménka záporná. S výskytem těchto rostlin klesá pravděpodobnost výskytu motýla. To by nemělo být interpretováno jako záporný vliv rostliny na výskyt motýla, pravděpodobnější interpretace je taková, že výskyt rostliny je závislý na dalších faktorech (například fyzicko-geografických) a jsou to právě tyto faktory, které mají na výskyt motýla záporný vliv. Tyto skryté závislosti bohužel nemohou být z daných dat potvrzeny ani vyvráceny. Provedená analýza ukazuje na souvislosti mezi výskyty vybraných taxonů, ale nevysvětluje je. Takto by měly být výsledky také prezentovány.

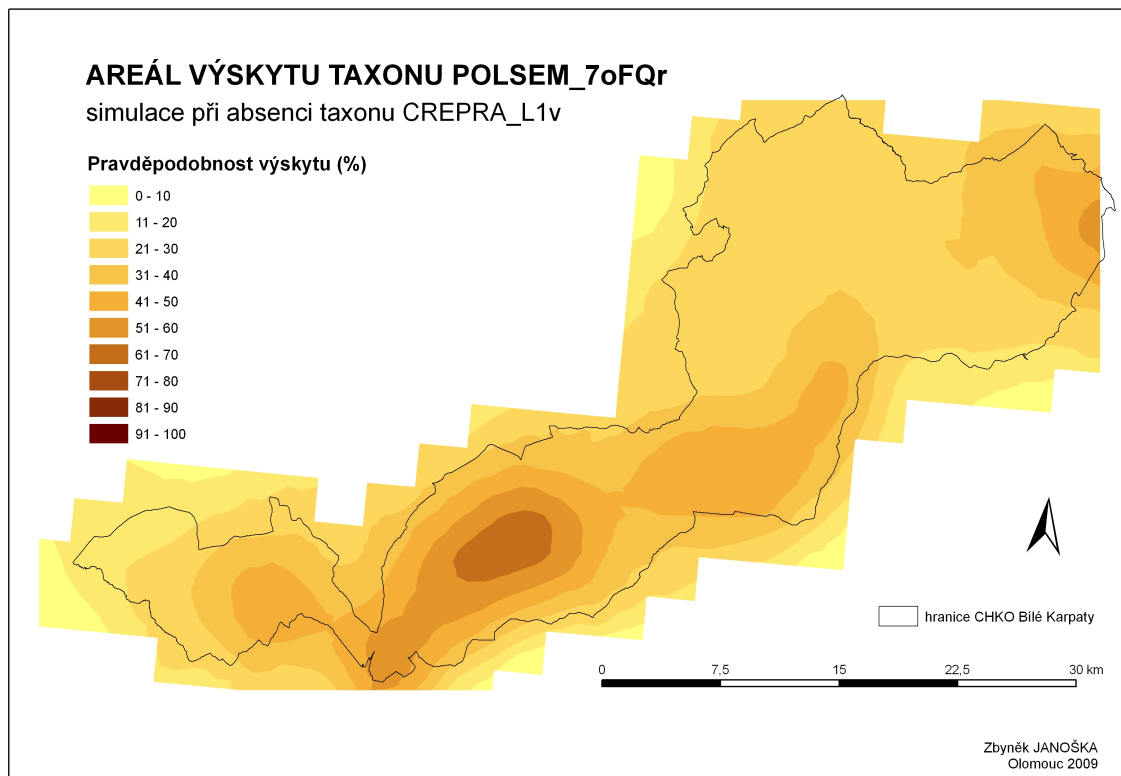
Protože závislost výskytu byla zjišťována na rostlinách podléhajících určitému režimu ochrany, nabízí se otázka, jak by se změnil areál výskytu v případě vyhynutí těchto druhů. Pro simulace byly vybrány taxony rostlin, které podléhají režimu ochrany C1 (kriticky ohrožený druh) nebo C2 (silně ohrožený druh). V modelech uvedených výše jsou tyto taxony podtrženy. Je však možné obecně modelovat změny areálů v závislosti např. na výstavbě silnice (tím pádem změny místních podmínek a úbytku rostlin), nebo

na jakékoli události, která naruší stabilitu ekosystému. Pak je ale potřeba znát souvislosti mezi výskytem rostlin a dalšími faktory.

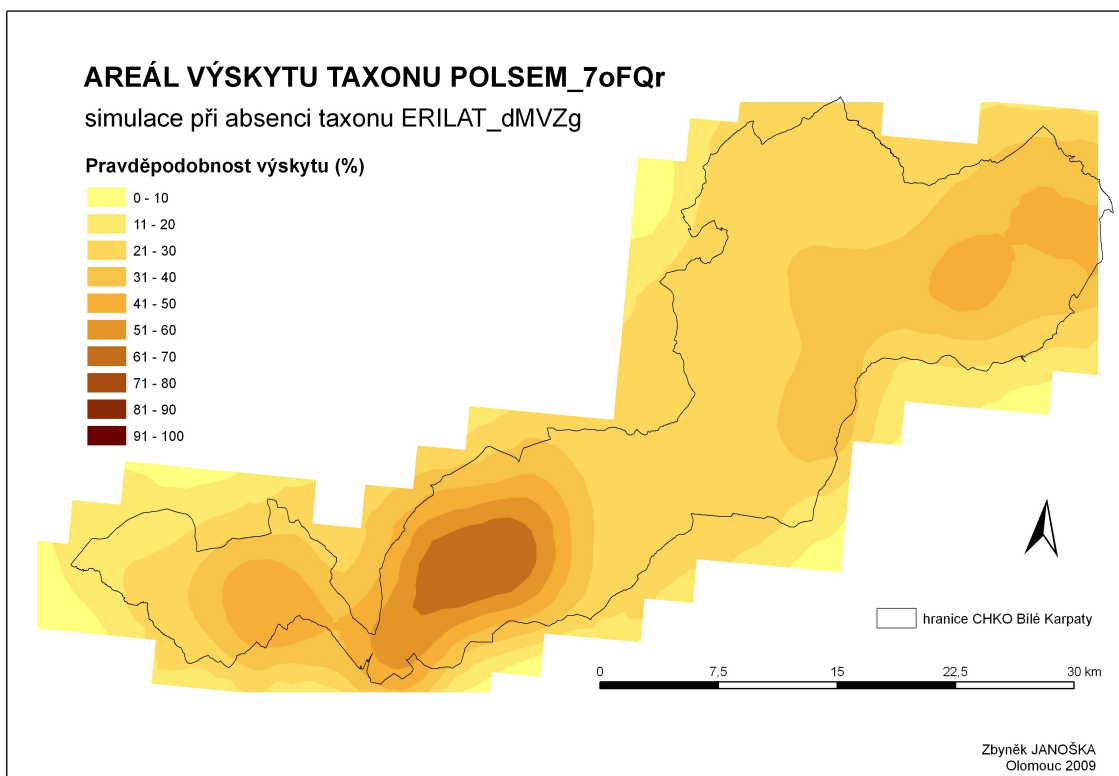
Jako modelový byl vybrán taxon POLSEM, areály předpokládaného výskytu ostatních taxonů jsou umístěny jako přílohy.



Obr. č.21 – areál výskytu taxonu POLSEM\_7oFQr



Obr. č.21 – simulace areálu výskytu při absenci taxonu CREPRA\_L1v



Obr. č.21 – simulace areálu výskytu při absenci taxonu ERILAT\_dMVZg

## 7. DISKUZE

Jako cíl práce byla uvedena analýza výskytu vybraných taxonů na základních přírodních faktorech. Nakonec bylo zadání zjednodušeno na analýzu výskytu pouze mezi taxony navzájem. Důvodem byla jednak výpočetní a také časová náročnost, především hledání metod vhodných pro danou úlohu. Zároveň bylo v zadání uvedeno, že modelové taxony budou vybrány ve spolupráci s Ústavem botaniky a zoologie MU Brno. Tato instituce se nakonec na výběru nepodílela a proto byla použita všechna dostupná data a výběr taxonů byl prováděn až v průběhu prací, viz kapitola 5. I toto bylo důvodem, proč nakonec nebyla provedena analýza výskytu na základních přírodních faktorech.

Základním problémem, který vyvstal při práci s daty, byla samotná jejich kvalita. Data pro výskyt motýlů a vyšších cévnatých rostlin byla dodána v rozdílném formátu (rostliny převážně jako binární, motýli jako nominální). Některé operace nad daty tedy nebylo možné provést, případně je bylo možné provést až po zjednodušení všech dat na data binární, čímž byly některé informace ztraceny. Dále zdrojová data nepokrývala celé zájmové území (viz přiložené mapky) a nebylo možné zjistit, zda-li je to skutečně nepřítomností sledovaných jevů, nebo jen data na tomto území nebyla pořizována. Tyto lokality bylo do výpočtů zahrnuty, pokud se na nich vyskytoval alespoň jeden ze sledovaných jevů (buďto motýl, nebo rostlina).

Toto vše se zcela jistě odrazilo na kvalitě výpočtů, i když metody byly voleny tak, aby odpovídaly vstupním datům. Pro lepší interpretaci výsledků by bylo třeba znát metody sběru dat, toto zejména platí pro botanická pozorování, která byla v průběhu let pořizována řadou odborníků i dobrovolníků a většinou se soustřeďovala na lokality z botanického hlediska významné. Důsledkem toho je různá kvalita i kvantita záznamů v jednotlivých oblastech. Protože většina dat měla nižší kvalitu (nesla méně informací, popřípadě záznamy byly v jednodušším – binárním - formátu), i data kvalitnější musela být na tento formát normalizována.

I přes tyto potíže se povedlo najít vhodné metody zpracování.

S prací se software problém nebyl. Geostatistical Analyst disponuje podrobnou a srozumitelnou nápovědou, k užitku byl i výukový program, který nad ukázkovými daty ukázal základy práce s extenzí. Mohu ho vřele doporučit všem zájemcům o práci s extenzí,

kteří nemají s interpolací v ArcGIS zkušenosti. R se ukázalo jako velmi dobrý prostředek pro zpracování dat. Některé problémy plynuly z nízké počáteční znalosti prostředí, a tudíž horší orientaci v nápovědě, která je pro práci s R klíčová. Samotné zpracování dat v R však bylo velmi rychlé, spolehlivé a s velkou mírou uživatelské volnosti.

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

Rovněž by výsledky neměly být interpretovány tím stylem, že výskyt motýlů je závislý pouze na výskytu rostlin. Jak bylo uvedeno v kapitole 6, lepší je interpretovat výsledky pouze tak, že ukazují na podobnost daných taxonů, ale nevysvětlují závislosti. Je možné, že absence taxonů rostlin tak, jak byla v kapitole 6 modelována, by se na výskytu motýlů neodrazila, ale pro zjištění detailnějších vazeb by bylo třeba použít pokročilejší metody statistické analýzy a zkoumat vliv více faktorů.



## 8. ZÁVĚR

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

Dalším výstupem jsou obecně formulované rovnice, vyjadřující závislost výskytu denních motýlů na vyšších cévnatých rostlinách. Tyto rovnice mohou pomoci zejména při simulacích, kdy budeme zjišťovat, jak by se změnil areál výskytu daného druhu při dosažení určitých podmínek, například kdyby došlo k vyhynutí jistého druhu apod. Tyto rovnice mohou nabýt vyššího významu po dokončení práce Bc. Jana Geletiče, která zkoumá právě výskyt taxonů vyšších cévnatých rostlin na fyzicko-geografických podmínkách. Pak bude možné obě práce propojit a zkoumat například, jak zasáhne výstavba silnice, přehrady apod. do přirozených areálů taxonů rostlin či motýlů. Výsledky práce pak mohou být použity k hodnocení zásahů do krajiny.

Jak již bylo mnohokrát zmiňováno, všechny výsledky práce jsou odvislé od kvality vstupních dat, principy, které byly v práci použity, jsou však obecné, a lze je aplikovat jak při zkoumání vztahu mezi výskytem motýlů a rostlin například na jiném území, ale obecně při výzkumu areálů výskytu jakéhokoli živočišného druhu.

Postupy použité v této práci je možné použít prakticky na jakákoli data, však bez dokonalého pochopení důvodů, proč byly tyto metody zvoleny, může dojít k jejich chybné interpretaci a proto takovéto použití této práce nedoporučuji.

Hlavní využití najde práce u pracovníků Správy CHKO Bílé Karpaty, jako inspirace však může posloužit i při řešení naprosto odlišných úloh.

## 9. POUŽITÉ ZDROJE

- [1] ANDĚL. J., (1978): *Matematická statistika*. – SNTL - Nakladatelství technické literatury, Praha, 346 pp.
- [2] HEBÁK. P., HUSTOPECKÝ. J., (2007): *Vícerozměrné statistické metody [1]*. - INFORMATORIUM, Praha, 240 pp.
- [3] HORAL, D., JAGOŠ, B., RESL, K., UŘIČÁŘ, J., JONGEPIER, J.W. & PECHANEC, V. (2006): *Atlas rozšíření vybraných druhů živočichů CHKO Bílé Karpaty*. - ZO ČSOP Bílé Karpaty, Veselí nad Moravou, 85 pp.
- [4] CHYTRÝ M. & TICHÝ L. (2003): *Diagnostic, constant and dominant species of vegetation classes and alliances of the Czech Republic: a statistical revision*. - Folia Facultatis Scientiarum Naturalium Universitatis Masarykianae Brunensis, Brno, 231 pp.
- [5] JONGEPIER, J. W. & JONGEPIEROVÁ, I. (2006): *Komentovaný seznam cévnatých rostlin Bílých Karpat*. - ZO ČSOP Bílé Karpaty, Veselí nad Moravou, 108 pp.
- [6] JONGEPIEROVÁ, I. (eds.), 2008: *Louky Bílých Karpat (Grasslands of the White Carpathian Mountains)*. - ZO ČSOP Bílé Karpaty, Veselí nad Moravou, 461 pp.
- [7] JUŘIKOVSKÁ, L. (2009): *Geostatistika populární a hravá. Seminář při mezinárodním symposiu GIS Ostrava 2009*. - Institut geoinformatiky, VŠB-TU Ostrava, Ostrava, 11 pp.
- [8] KUČA, P., MÁJSKÝ, J., KOPEČEK, F. & JONGEPIEROVÁ, I. (1992): *Biele Bílé Karpaty*. - Ekológia, Bratislava, 380 pp.
- [9] KUPKA, K., (2006): *QC.Expert Uživatelský manuál*. – TriloByte STATISTICAL SOFTWARE, Pardubice, 253 pp.
- [10] MACKOVČIN, P., JATIOVÁ, M., a kol. (2002): *Zlínsko. Chráněná území ČR, svazek II*. - Agentura ochrany přírody a krajiny ČR a EkoCentrum Brno, Praha, 376 pp.
- [11] MELOUN, M., MILITKÝ, J., HILL, M., (2005): *Počítačová analýza vícerozměrných dat v příkladech* – Akademie věd ČR, Praha, 449 pp.
- [12] PIRO, Z. et WOLFOVÁ, J. (eds.), (2008): *Zachování biodiverzity karpatských luk*. - FOA, Nadační fond pro ekologické zemědělství, Praha, 108 pp
- [13] VOŽENÍLEK, V. (2002): *Diplomové práce z geoinformatiky*. - Univerzita Palackého, Olomouc, 60 pp.
- [14] Using ArcGIS Geostatistical Analyst

### Internetové zdroje

- [15] Bílé Karpaty [online]. ©2009. poslední kontrola 7.5.2009. [citace 2009-5-7]  
URL:<http://bilekarpaty.cz/pages/pomery.htm>
- [16] ESRI Products Overview [online]. ©ESRI. poslední kontrola 7.5.2009. [citace 2009-5-7] URL:<http://www.esri.com/products/index.html>
- [17] FLANTUA, S. G. A., a kol. (2007): *Application of GIS and logistic regression to fossil pollen data in modelling present and past spatial distribution: A case study in the Colombian Savanna*, poslední kontrola 7.5.2009. [citace 2009-5-7] URL:<http://ncg.nu-im.ie/gisruk/materials/proceedings/PDF/2B2.pdf>
- [18] KLUFOVÁ, R. & ROST, M. (2009): *Metody shlukové analýzy*. poslední kontrola 7.5.2009. [citace 2009-5-7]  
URL:[http://www2.zf.jcu.cz/public/departments/kmi/MSMT\\_05/metody%20shlukove%20analyzy.pdf](http://www2.zf.jcu.cz/public/departments/kmi/MSMT_05/metody%20shlukove%20analyzy.pdf)
- [19] LAŠTŮVKA, Z., & LIŠKA, J. (2008): *Seznam motýlů České republiky*, poslední kontrola 7.5.2009. [citace 2009-5-7] URL:[www.lepidoptera.wz.cz](http://www.lepidoptera.wz.cz).
- [20] NASA ADS: Assessment of karst rocky desertification using the radial basis function network model and GIS technique: a case study of Guizhou Province, China [online]. poslední kontrola 29.3.2009. [citace 2009-3-29]  
URL:<http://adsabs.harvard.edu/abs/2006EnGeo..49.1173Q>
- [21] ROBERTS, E. A., a kol. (2003): *Using sampling and inverse distance weighted modeling for mapping invasive plants*, poslední kontrola 7.5.2009. [citace 2009-5-7]  
URL:<http://extension.oregonstate.edu/catalog/html/sr/sr1057/017using.pdf>

## **10. SUMMARY**

The basic problem of the landscape mapping is how to obtain the high-quality data. Prediction models can be very helpful for estimating areas of incidence of selected taxa. Also the structure of natural phenomenon is complex and can not be simply quantified without analysing the processes behind.

Main aim of this bachelor thesis was to find dependence between the incidence of selected botanical taxa and zoological taxa from the family of lepidoptera.

Statistical analysis in the R-Project environment was used to determine these linkages and then ArcGIS 9.3 (Geostatistical Analyst extension) to model the distribution of the phenomenon.

## **PŘÍLOHY**

## Seznam příloh

Příloha 1: Slovníček

Příloha 2: CD-ROM

- textová část bakalářské práce
- zdrojová data: tabulky lepidopterologických pozorování
- zdrojová data: tabulky botanických pozorování
- zdrojová data: polygonová vrstva kvadrátů síťového mapování dle standardu KFME
- kontingenční tabulka motýli
- kontingenční tabulka rostliny
- kontingenční tabulka rostliny C
- kontingenční tabulka vše, relativní hodnoty (formát .csv)
- geografické středy
- shluky vše
- shluky simulace
- modely (protokoly z pracovního prostředí R)
- pracovní protokol z programu QC.Expert (formát .dbf)
- skript pro výpočet geografického středu
  - skript ve formátu .txt
  - zdrojová data pro skript
- metadatový soubor XML
- WWW stránky
- soubor map
  - areál výskytu taxonu ARGNIO\_qHXw8
  - areál výskytu taxonu ARIEUM\_sBFE1
  - areál výskytu taxonu ARISP\_1Neyy
  - areál výskytu taxonu BOLSEL\_HziyG
  - areál výskytu taxonu BREINO\_3Ezgo
  - areál výskytu taxonu COLSP\_aJdtb
  - areál výskytu taxonu HETMOR\_ZF2cr
  - areál výskytu taxonu MELATH\_53cKb
  - areál výskytu taxonu POLSEM\_7oFQr
  - simulace výskytu taxonu ARGNIO\_qHXw8 při absenci CARHOR\_DTEaZ
  - simulace výskytu taxonu ARGNIO\_qHXw8 při absenci CYPCAL\_J1XT
  - simulace výskytu taxonu ARGNIO\_qHXw8 při absenci POLMAJ\_XL1wo
  - simulace výskytu taxonu ARIEUM\_sBFE1 při absenci ORCMOR\_wP1hY
  - simulace výskytu taxonu ARISP\_1Neyy při absenci CYPCAL\_J1XT
  - simulace výskytu taxonu COLSP\_aJdtb při absenci GENCRU\_flsBI
  - simulace výskytu taxonu POLSEM\_7oFQr při absenci CREPRA\_L1v
  - simulace výskytu taxonu POLSEM\_7oFQr při absenci ERILAT\_dMVZg

## **Příloha č. 1: Slovníček**

Text byl zpracován volně podle kapitoly Glossary v publikaci Using ArcGIS Geostatistical Analyst.

### **Anisotrophy (anizotropie)**

Vlastnost prostorových procesů nebo dat, kdy se prostorová závislost mění jak se vzdáleností, tak se směrem mezi lokacemi. (porovnat s isotrophy)

### **Autocorrelation (autokorelace)**

Statistická korelace mezi prostorově náhodnými proměnnými stejného typu, vlastností apod., kde korelace závisí na vzdálenosti a/nebo směru mezi lokacemi. (porovnat s crosscorrelation)

### **Binning**

Je proces, kdy dvojice bodů jsou shlukovány do skupin na základě jejich vzdálenosti od sebe. Každému binu je přiřazena hodnota vzdálenosti a semivariance, které odpovídají průměru všech bodů spadajících do daného binu. Běžně se používají dvě metody binningu – formované buďto čtvercovou, nebo sektorovou metodou (Geostatistical Analyst pracuje s čtvercovou metodou).

### **Cokriging**

Statistická interpolační metoda, která používá data různých datových typů (různé atributy) k předpovědění hodnoty primárního datového typu (primárního atributu). Výsledkem jsou rovněž směrodatné chyby předpovědi - míra toho, jak přesná je předpověď. (více také kriging)

### **Correlation (korelace)**

Je to kovariance odstupňovaná tak, že hodnota -1 znamená žádnou závislost, hodnota 1 pak naprostou závislost dvou vstupních datových sad. (více také covariance)

### **Covariance (kovariance)**

Je statistická tendence proměnných stejných typů, vlastností apod., kdy hodnoty těchto proměnných vykazují závislost vůči sobě. Pozitivní covariance nastává, když obě z proměnných jsou nad jejich očekávaným průměrem, negativní covariance tehdy, když

jedna z proměnných je nad očekávanou, druhá pod očekávanou hodnotou. (porovnat s crossvariance)

### **Crosscorrelation**

Statistická korelace mezi prostorově náhodnými proměnnými různého typu, vlastností apod., kde korelace závisí na vzdálenosti a/nebo směru mezi lokacemi. (porovnat s autocorrelation)

### **Crossvariance**

Je statistická tendence proměnných různých typů, vlastností apod., kdy hodnoty těchto proměnných vykazují závislost vůči sobě.. Pozitivní Crossvariance nastává, když obě z proměnných jsou nad jejich očekávaným průměrem, negativní crossvariance tehdy, když jedna z proměnných je nad očekávanou, druhá pod očekávanou hodnotou. (porovnat s covariance)

### **Cross-validation**

Je metoda, kdy část dat je oddělena od datasetu a zbytek dat je použit k předpovědění hodnot v místě oddělených dat. Výsledky jsou pak porovnány se skutečnými hodnotami. Full Cross-validation znamená oddělit každá jednotlivá data v datasetu a použít zbytek dat k jejich předpovědi.

### **Crossvariogram**

Je funkcí vzdálenosti a směru dělící dvě lokace a používá se ke kvantifikaci cross-correlation. Crossvariogram je definován jako rozptyl rozdílu mezi proměnnými různých typů nebo atributů na dvou lokacích. Crossvariogram většinou roste se vzdáleností a je definován parametry: nugget, sill, range.

### **Deterministic (deterministický)**

Je to typ nebo část modelu, kde výsledek je kompletně a exaktně odvozen ze vstupních hodnot; nepočítá se s náhodnými procesy.

### **Detrending**

Je proces odstraňování trendu. Povrch tvořící trend je odstraněn z originálních dat – výsledné hodnoty po odstranění trendu se nazývají rezidua.



### **Empirical (empirický)**

Ve statistickém pojetí znamená, že kvantita je odvislá od dat, pozorování či experimentů, není to model ani jeho část. Například empirický semivariogram je počítán pouze ze vstupních dat, na rozdíl od teoretického variogramu, který reprezentuje ideální stav.

### **Global Polynomial interpolation**

Jedna z deterministických interpolačních metod, kdy výsledná data neprocházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření.

### **Intrinsic stationarity (vnořená stacionarita)**

Předpoklad, že data jsou výsledek náhodných procesů s konstantním aritmetickým průměrem, a že semivariogram je závislý pouze na vzdálenosti a směru mezi dvěma lokacemi.

### **Inverse Distance Weighted interpolation (IDW)**

Jedna z deterministických interpolačních metod, kdy výsledná data procházejí přesně vstupními body a maximální a minimální hodnoty mohou být dosaženy pouze ve vstupních bodech.

### **Isotrophy (izotropie)**

Je vlastnost přírodních procesů či dat, kdy prostorová závislost (autokorelace) se mění jen se vzdáleností, směr je nepodstatný (porovnat s anisotrophy).

### **Kriging**

Je statistická interpolační metoda, která počítá z dat jednoho typu (hodnoty jednoho atributu) k předpovědi dat stejného typu (hodnoty stejného atributu) na nezměřených lokacích. Výsledkem jsou rovněž směrodatné chyby předpovědi - míra toho, jak přesná je předpověď (více také cokriging).

### **Least-squares fit (metoda nejmenších čtverců)**

Model (regresní linie, povrch, funkce) prochází daty tak, že je minimalizován součet druhých mocnin odchylek od měřených dat.

### **Local Polynomial interpolation**

Jedna z deterministických interpolačních metod, kdy výsledná data neprocházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření.

### **Mean Stationarity**

Vlastnost prostorových procesů, kdy všechny prostorově neusprádané proměnné mají stejnou střední hodnotu.

### **Nugget**

Je parametr semiovariogramu, který reprezentuje náhodnou chybu, chybu měření nebo rozptyl měření v mikroměřítkách, který je příliš malý na to, aby byl zachycen. V grafu je to nespojitost, kdy ve vzdálenosti 0 je prostorová autokorelace nenulová.

### **Outlier**

Jedná se o naměřený jev s výrazně odlišnou hodnotou než jakou mají ostatní měřené body. Globální outlier je hodnota, která výrazně vybočuje z intervalu ostatních měřených hodnot celého datasetu. Lokální outlier je taková hodnota, která sice zapadá do intervalu měřených hodnot, ale vzhledem k nejbližším okolním bodům je nezvykle vysoká či nízká.

Outliery mohou být buďto reálnými abnormalitami, a pak jsou často nejdůležitějšími měřenými daty, nebo mohou být výsledkem chybného měření. Pokud jsou výsledkem chybného měření či záznamu, měly by být před jakoukoli další prací s daty odstraněny.

### **Polynomial (polynomní)**

Funkce, která je složena ze součtu měř jednotlivých proměnných. V prostorovém kontextu, kdy x-souřadnice a y-souřadnice jsou proměnné, polynomní jsou výrazy: 1, x, y,  $x^2$ ,  $y^2$ ,  $x^2y$ , apod., kdy každý z nich je ohodnocen koeficientem  $b_0$  až  $b_n$ .

### **Probability Map**

Je povrch, který vyjadřuje pravděpodobnost, že měřená hodnota se vyskytuje nad/pod prahem, zadaným uživatelem.

### **QQPlot**

Druh grafu, kdy kvantily dvou rozdělení jsou vyneseny relativně k sobě – slouží

k porovnání rozdělení dvou datasetů.

### **Quantile (kvantil)**

Pro  $n$ -tý kvantil ( $n$  je mezi 0 a 1) je hodnota, která má poměr  $n$  dat nad svou hodnotou (pro  $n=0,25$  to bude hodnota, pod níž se bude nacházet 75 % dat)

### **Radial Basis Function**

Jedna z deterministických interpolačních metod, kdy výsledná data procházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření.

### **Regression (regrese)**

Je statistická metoda, kdy proměnná je vytvořena z funkcí jedné nebo více proměnných. Často je to funkce splňující podmínku nejmenších čtverců (tedy je to taková funkce, že celková odchylka od vstupních funkcí je minimální).

### **Residuals (rezidua)**

Jsou hodnoty získané oddělením trendu od původních vstupních dat.

### **Range**

Je parametr semivariogramu, který určuje vzdálenost, ve které přestávají být hodnoty autokorelované (maximální vzdálenost, ve které jsou hodnoty na sobě závislé – ve větší vzdálenosti se už nevyplatí modelovat.)

### **Second-order Stationarity (stacionarita druhého řádu)**

Předpoklad, že data jsou výsledek náhodných procesů s konstantním průměrem, a že prostorová kovariance je závislá pouze na vzdálenosti a směru dvou pozorování, nikoli na jejich poloze.

### **Semicrossvariogram**

Semivariogram dělený dvěma.

### **Semivariogram**

Variogram dělený dvěma.

## **Sill**

Parametr semivariogramu, který určuje hodnotu, ke které se blíží model semivariogramu, při velkých vzdálenostech. Ve velkých vzdálenostech, data přestávají být autokorelovaná, takže Sill je vlastně rozptyl náhodné proměnné. Některé teoretické semivariogramy Sill nemají. V Geostatistical Analyst mají všechny modely semivariogramu Sill.

## **Spatial Dependence (prostorová závislost)**

Poznatek, že věci, které jsou si prostorově bližší, jsou si více podobné než věci vzdálené.

## **Spline interpolation**

Jedna z deterministických interpolačních metod, speciální případ radial basis functions, kdy výsledná data procházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření.

## **Stacionarita**

Předpoklad, že všechny statistické vlastnosti atributů jevu jsou závislé pouze na relativní poloze pozorování.

## **Trend**

Jakýkoli povrch je tvořen dvěma hlavními komponenty – globálním trendem a náhodným rozptylem krátkého dosahu. Trend je povrch tvořený je z fixních parametrů, často polynomů souřadnic  $x$  a  $y$ . Trend zpravidla modeluje jen změny velkých měřítek, malé odchylky jsou pak výsledkem náhodných chyb. Při geostatistickém modelování bývá často trend odstraněn a modelován jen náhodný rozptyl. Důvodem je zachování předpokladu stacionarity.

## **Variogram**

Je funkce vzdálenosti a směru mezi dvěma lokacemi a používá se ke kvantifikování autokorelace. Variogram je definován jako rozptyl rozdílů hodnot proměnné v párech měřených prvků na jejich prostorové vzdálenosti. Většinou narůstá se vzdáleností a je definován pomocí parametrů: nugget, sill a range.

### **Variography (variografie)**

Je proces odhadování teoretického semivariogramu. Skládá se z: průzkumu vstupních dat, výpočtu empirického semivariogramu, binningu, přiřazení modelu k semivariogramu, a ohodnocení výsledného modelu.