

## NOVÉ A PŘEHLÍŽENÉ MATEMATICKO-STATISTICKÉ METODY PRO GEOBIOCENOLOGII A EKOLOGII

NEW OR IGNORED MATHEMATICAL-STATISTICAL METHODS FOR GEOBIOCENOLOGY AND ECOLOGY

*Karel Matějka*

*IDS, Na Komořsku 2175/2a, 143 00 Praha 4, matejka@infodatasys.cz*

### ABSTRACT

This paper deals primarily with the use of Monte Carlo permutation tests, expanding the possibility of cluster analysis and calculating correlation distances in multidimensional data. All of these procedures can find applications in environmental, geobiocenological and similar research. Mostly these are not entirely new methods but their modifications. The neglect of the Monte Carlo permutation tests is mainly due to their common absence in the statistical software used or to their use requiring more detailed knowledge of the method. The methods are illustrated by several examples: (1) the classification of sequentially ordered phytocenological relevés by the HSCA method for the study of the regeneration of the plant community following the collapse of the tree layer, (2) the comparison of the results of the classification of soil mite communities with external data on soil properties was carried out by a permutation test for variance analysis, (3) the hierarchical agglomerative classification was used to study the spatial structure of the tree layer, (4) differences in the properties of surface soil layers of oak woods were evaluated by correlation distances and subsequent cluster analysis, (5) correlation distances were also used for evaluation of Norway spruce subpopulations according to their relationship of dendrometric variables and age.

**Key words:** classification, correlation distances, Monte Carlo methods, sequentially ordered samples, test of difference

### ÚVOD A PŘEHLED METOD

Cílem tohoto článku je podat přehled o tom, jaké méně časté matematicko-statistické metody mohou být užitečné při zpracování dat v oblasti fytocenologie, geobiocenologie a ekologie. Tyto metody jsou rozvíjeny na pracovišti IDS a byly úspěšně použity v rámci řady studií. Uvedeny jsou některé dílčí výsledky, které ilustrují popisované metody.

### Monte Carlo simulace a statistické testování

Metoda Monte Carlo se používá v rámci statistické analýzy stále častěji, přesto nebývá součástí základního statistického software, jakým je například oblíbená Statistica (dříve Statsoft Inc., nyní firma Dell; [www.statsoft.com](http://www.statsoft.com)). Příslušný permutační test je možno aplikovat v rámci řady analýz (například ROBERT et CASELLA, 2010), již dříve se stal běžnou součástí vyhodnocení výsledků ordinací v programu Canoco (LEPŠ et ŠMILAUER, 2003).

Obecný základ metody Monte Carlo v rámci statistického testování spočívá v následujícím postupu: Necht' máme množinu objektů  $\Theta = \{o_1, o_2, \dots, o_n\}$ . Celá množina je úplně rozdělena do dvou diskrétních podmnožin  $\Theta_1 = \{o_1, \dots, o_{n_1}\}$  a  $\Theta_2 = \{o_{n_1+1}, \dots, o_{n_1+n_2}\}$ ,  $\Theta = \Theta_1 \cup \Theta_2$ , přičemž  $n_1+n_2 = n$ . Předpokládejme, že můžeme vypočítat statistiky  $Z_1$  a  $Z_2$ , které jsou funkcemi obou podmnožin,  $Z_1 = Z(\Theta_1)$ ,  $Z_2 = Z(\Theta_2)$ . Nyní může být cílem testovat hypotézu  $H_0: Z_1 < Z_2$ . Princip metody spočívá ve výpočtu řady náhodných permutací  $P = [o_{P(1)}, o_{P(2)}, \dots, o_{P(n)}]$  množiny  $\Theta$ . Tyto permutace nám definují náhodné rozdělení množiny  $\Theta$  vždy na dvě podmnožiny  $\Theta_1$  (prvních  $n_1$  prvků permutace  $P$ ) a  $\Theta_2$  (následujících  $n_2$  prvků permutace  $P$ ),  $\Theta = \Theta_1 \cup \Theta_2$ . Necht' vypočteme celkem  $N$  takových permutací, přičemž  $Z(\Theta_1)$

$< 2(\Theta_2)$  platí pro  $N_0$  permutací, potom pravděpodobnost chyby prvního druhu  $\alpha = 1 - N_0/N$ . Obdobně lze postupovat i v případě, že počítáme jednu statistiku  $\Phi(\Theta_1, \Theta_2)$  na základě dvou podmnožin: poté počítáme případy permutací, kdy  $\Phi(\Theta_1, \Theta_2) < \Phi(\Theta_1, \Theta_2)$ .

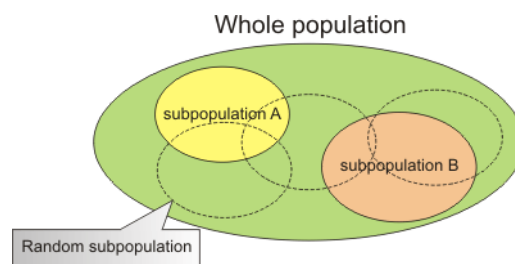
Monte Carlo permutační test je sofistikovaná procedura, kterou lze využít také pro testování rozdílů hodnot mezi dvěma diskrétními podmnožinami pokrývajícími výchozí datovou množinu. Jedná se o obdobu t-testu běžně používaného, jehož předpokladem je však normalita rozdělení dat. Taková normalita bývá často porušena nebo ji vzhledem k charakteru dat nelze očekávat. Příkladem mohou být indexy druhové diversity nebo druhové bohatosti (počet druhů). V takovém případě je vhodný právě Monte Carlo test, který je nezávislý na rozdělení dat. Tento test je v programu DBreleve (MATĚJKA, 2017c) realizován ve dvou formách:

- parametrický test: počítány jsou přímé rozdíly mezi hodnotami -  $D(x_1, x_2) = x_1 - x_2$ . V tomto případě bude průměrná difference rovna rozdílu aritmetických průměrů pro obě podmnožiny. Výsledná pravděpodobnost chyby počítané pomocí Monte Carlo simulace odpovídá pravděpodobnosti chyby párového t-testu za podmínky dostatečného rozsahu výběrového souboru a při shodě rozdělení dat s normálním rozdělením.
- neparametrický test: rozdíly jsou dány funkcí  $D(x_1, x_2) = -1$  pro  $x_1 < x_2$ ,  $0$  pro  $x_1 = x_2$ ,  $+1$  pro  $x_1 > x_2$ . Neparametrický test je vhodný zvláště pro data s extrémně odlehými hodnotami.

Příklad použití testu rozdílů párových hodnot je možno nalézt v publikaci HRIB et al. (2017), kde v tabulce 4 byly testovány například rozdíly pro indikační hodnoty dle Ellenberga (ELLENBERG et al., 1992), které byly počítány na týchž plochách v různých letech (2001 a 2014). Monte Carlo permutační test je v tomto případě vhodný, protože nemusí být řešena otázka normality rozdělení indikačních hodnot.

Specificky byl postavený Monte Carlo test rozdílu genetických parametrů subpopulace vůči zbytku populace a/nebo dvou subpopulací navzájem (IVANEK et MATĚJKA, 2016). Testovat je možné rozdíl mezi hodnotou parametru pro subpopulaci a tímž parametrem v celé populaci, respektive je vypočítávána pravděpodobnost dosažení hodnoty parametru nižší nebo rovné dané konstantě pro náhodně vybranou subpopulaci daného rozsahu (počtu jedinců). Princip testování je možno popsat následovně:

Nechť v rámci analyzované populace  $T$  o rozsahu  $N$  je definována nějaká subpopulace  $A$  o rozsahu  $n_A$  (obr. 1), poté není možné přímo srovnávat libovolné parametry diversity  $D_A$  a  $D_T$ , protože  $n_A < N$ . Proto program generuje řadu náhodných subpopulací o rozsahu  $n_A$  (jako náhodné kombinace  $n_A$  prvků z  $N$  prvků) a pro každou z těchto kombinací vypočte požadovaný parametr ( $D_i$ ). Hodnoty  $D_i$  mají nějaké rozdělení, které je možné aproximovat na základě dostatečně velkého počtu náhodně vygenerovaných kombinací (většinou je dostatečných 1000 náhodných kombinací). Program počítá kvantily tohoto rozdělení, respektive pravděpodobnosti  $\alpha = P[D \leq D_\alpha]$ . Přitom hodnoty  $D_\alpha$  představují  $\alpha$ -kvantily tohoto rozdělení. Hodnota  $D_A$  je statisticky průkazně nižší než  $D_T$  na hladině  $\alpha$ , pokud  $D_A \leq D_\alpha$ . Obdobně je možno testovat průkaznost toho, že  $D_A$  je vyšší než  $D_T$ . Obdobné simulace náhodného vzorkování, tj. tvorby subpopulací je pro analýzu výsledků genetického studia známa i z literatury (například WEGMANN et al., 2010).



Obr. 1. Populace, subpopulace a náhodná populace, jak jsou použity při testování rozdílů genetických parametrů.

Na odlišném základě pracuje program SeqAn (MATĚJKA, 2009) - tam se nejedná o permutační test, ale o výpočet kritických hodnot pořádkových statistik na základě generování hodnot náhodných (při zvoleném typu rozdělení). Vyvinutý postup lze užít například pro hodnocení kvalitativní chromatografie na tenké vrstvě, gelové chromatografie a elektroforézy (tedy i při vyhodnocování rozdílů alelických kombinací při isoenzymové analýze).

### **Hodnocení výsledků klasifikace a porovnání s externími daty**

Test odlišnosti hodnot nezávisle proměnné pro objekty klasifikované hierarchickou procedurou vychází z principu jednofaktorové analýzy rozptylu. MCCUNE et GRACE (2002) doporučují takový test pro verifikaci výsledků shlukové analýzy

Při použití tohoto postupu je však potřebné být opatrný na to, aby klasifikace vzorků a pro testování používaná charakteristika vzorků nevycházely ze stejných dat (MCCUNE et GRACE, 2002), respektive nebyly přímo závislé. Nemělo by tedy smysl testovat klasifikaci fytocenologických snímků v závislosti na pokryvnosti nějakého druhu. Naopak může být dobře použitelné testování závislosti klasifikace snímků podle složení bylinné etáže v závislosti na pokryvnosti (zápoji) stromového patra.

Testování závislosti výsledků hierarchické klasifikace (dendrogramů) na externích datech bylo zařazeno jako jedna z procedur navazujících na konstrukci dendrogramů v programu DBreleve (MATĚJKA, 2017c).

### **Klasifikace sekvenčně uspořádaných objektů**

Běžné metody hierarchické aglomerativní klasifikace vytvářejí shluky objektů bez ohledu na jejich reálné vztahy, které nemusí být popsány zpracovávanými daty. Přitom je v mnoha případech patrné, že klasifikované objekty jsou uspořádané. Takové uspořádání může vycházet například z pořadí ploch podél transektu nebo z časového uspořádání vzorků pořizovaných v průběhu sukcese (LEGENDRE et LEGENDRE, 2012: pp. 711-858). Proto byly již v minulosti popsány metody *constrained clustering* (GORDON, 1973). Tento postup však je, bohužel, v ekologických studiích opomíjen.

Obdobná metoda HSCA (MATĚJKA, 1993) slouží též pro klasifikaci vzorků sekvenčně uspořádaných. Takovým uspořádáním může být například seřazení snímkovaných ploch podél transektu v terénu nebo seřazení fytocenologických snímků opakovaně snímkované sukcesní plochy v čase. Používané indexy nepodobnosti mají navíc úzký vztah k  $\beta$ -diversitě (MAGURRAN, 2004), některé lze přímo považovat za míru  $\beta$ -diversity.

Klasifikace metodou HSCA je realizována v programu, který se stal součástí balíku DBreleve, kde je dále možno vykreslit i výsledné dendrogramy (MATĚJKA, 2017c).

### **Porovnání podmnožin objektů popsaných mnohorozměrnými daty**

Jsou známy případy, kdy korelace mezi dvěma proměnnými se liší mezi dvěma nebo více množinami sledovaných objektů, pro které jsou zmíněné dvě proměnné měřeny. Například může být zjištěna rozdílná korelace mezi výškou a výčetní tloušťkou u smrku v horských polohách oproti polohám nížinným. Další zdokumentovaný příklad z pedologie nabízí rozdílná korelace mezi obsahem celkových popelovin a rozpustných popelovin ve svrchní vrstvě půdy podle toho, jestli se jedná o lokalitu horských smrčín (8. LVS) nebo potenciálně smíšených porostů nižšího vegetačního stupně (7. LVS). Rozdíl těchto korelací je spojen s probíhajícími podzolizačními procesy v 8. LVS (MATĚJKA et STARÝ, 2009). V případě, že objekty jsou popsány více proměnnými, je potřeba použít speciální mnohorozměrné analýzy. Jeden takový možný postup je popsán dále.

Článek MATĚJKA (2017a) přinesl popis nové metody klasifikace a ordinace objektů rozdělených do několika podmnožin a to na příkladu populací či subpopulací dřevin, které jsou charakterizovány dendrometrickými veličinami. Obdobný postup lze použít pro podmnožiny nejrozličnějších objektů, kde objekty jsou popsány několika charakteristikami, které mohou být různě korelovány. Při vyhodnocení podobnosti podmnožin není uvažována průměrná hodnota zjišťovaných charakteristik, ale rozdíl mezi korelačními koeficienty. Může tak být zjištěna například vysoká podobnost mezi dvěma populacemi smrku, které jsou různě staré a tedy i s různě velkými jedinci, ale kde je u obou subpopulací obdobně korelovaný vztah mezi výškou a tloušťkou.

Protože metoda byla v článku MATĚJKA (2017a) popsána pouze ve vztahu ke klasifikaci (sub)populací dřevin, je v následujícím proveden její zobecněný popis.

Mějme množinu objektů, které jsou rozděleny do  $m$  podmnožin. Každý objekt je popsán pomocí  $n$  proměnných. Vztah jednotlivých proměnných je hodnocen pomocí Pearsonova korelačního koeficientu ( $r$ ). Statistická významnost rozdílu dvou odpovídajících si korelačních koeficientů  $r_1$  a  $r_2$  ve dvou srovnávaných podmnožinách je vyjádřena pomocí proměnné  $U$

$$U = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

kde

$$z_1 = \frac{1}{2} \ln \frac{1 + r_1}{1 - r_1}$$

Proměnná  $z_2$  je počítána obdobně.

Protože proměnná  $U$  má přibližně normální rozdělení  $N(0;1)$ , lze ji testovat proti kritickým hodnotám normálního rozdělení. Kritické hodnoty pro  $U$  jsou tedy  $\pm 1,960$  na hladině významnosti  $\alpha = 0,05$  a  $\pm 2,576$  na hladině významnosti  $\alpha = 0,01$  (pro oboustranný test). Současně lze absolutní hodnotu  $|U|$  považovat za míru nepodobnosti dvou hodnocených populací.

Vzhledem k výše uvedenému je možno pro  $m$  srovnávaných podmnožin vytvořit matici

$$\mathbf{U} = \begin{bmatrix} 0 & U_{1,2}^2 & \dots & U_{1,m}^2 \\ U_{2,1}^2 & 0 & \dots & U_{2,m}^2 \\ \dots & \dots & \dots & \dots \\ U_{m,1}^2 & U_{m,2}^2 & \dots & 0 \end{bmatrix}$$

a tu užít pro hierarchické aglomerativní shlukování.

Vzhledem k tomu, že korelace je možno počítat mezi řadou  $n$  proměnných (například pro proměnné  $v$  a  $w$ ;  $1 \leq v, w \leq n$ ), pro jednu dvojici podmnožin  $p$  a  $q$  ( $1 \leq p, q \leq m$ ) je příslušných diferencí několik ( $U_{(p,q),(v,w)}$ ). Tyto rozdíly lze pro jednu dvojici podmnožin počítat jako sumu kvadrátů

$$X_{p,q}^2 = \sum_{v=2}^n \sum_{w=1}^{v-1} U_{(p,q),(v,w)}^2$$

Vzhledem k typu rozdělení proměnných  $U$ , má proměnná  $X^2$  rozdělení  $\chi^2$  o  $(n(n-1)/2)-1$  stupňů volnosti. Též  $X_{p,q}^2$  lze považovat za vzdálenost (nepodobnost) mezi srovnávanými podmnožinami  $p$  a  $q$  ( $1 \leq p, q \leq m$ ) a lze ji užít pro klasifikaci celé skupiny podmnožin, případně pro ordinaci metodou jako Non-metric multidimensional scaling - NMS (MCCUNE et GRACE, 2002). Proměnnou  $X^2$  je možno nazvat korelační vzdáleností.

Výpočet korelačních vzdáleností mezi podmnožinami je možno provést v novém programu CorrelDist, který je součástí nejnovější verze software PlotOA (MATĚJKA, 2017b). Tento program umožňuje mimo jiné počítat distance nejen pomocí celé matice *U*, ale též pomocí výběru určitých korelačních koeficientů podle toho, jaké korelace mohou být logicky zdůvodnitelné. Dendrogramy byly konstruovány na základě matic vzdálenosti klasifikovaných objektů v programu Statistica, verze 8 (StatSoft Inc.; [www.statsoft.com](http://www.statsoft.com)). Matici *U* je možno zpracovat též metodami ordinační analýzy, především metodou non-metric multidimensional scaling (NMS; MCCUNE et GRACE, 2002).

#### PŘÍKLADY APLIKACÍ A DISKUSE

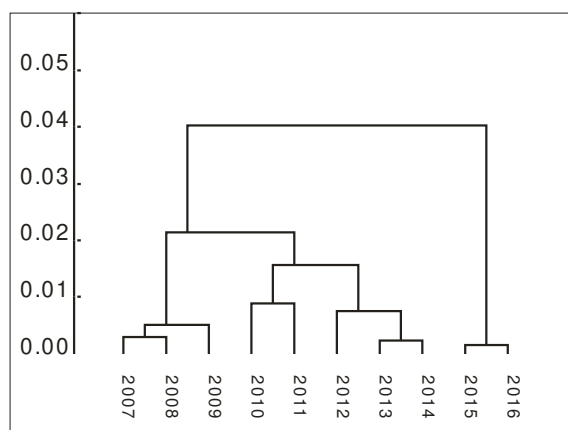
##### **Kdy dochází k nejvýraznější změně rostlinného společenstva po rozpadu stromového patra horské smrčiny?**

Každoročně byly pořizovány fytoecologické snímky trvalých ploch v povodí Plešného jezera na Šumavě (MATĚJKA, 2011a, 2015, 2016). K prvnímu snímkování došlo v roce 2007. Snímky z jednotlivých ploch byly klasifikovány metodou HSCA s užitím míry nepodobnosti *Sq*, která vyjadřuje rozdílnost v kvantitativním zastoupení jednotlivých druhů, odpovídá tedy míře celkové  $\beta$ -diversity.

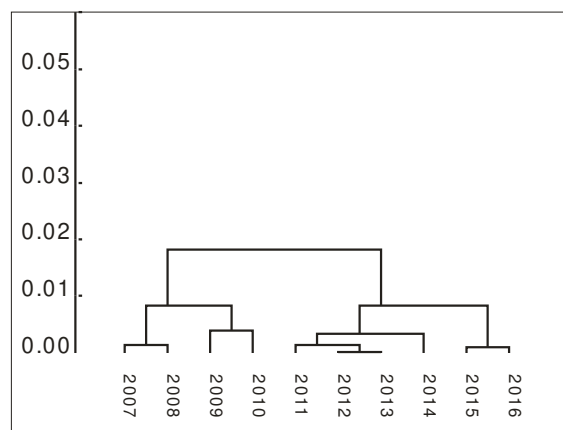
V povodí Plešného jezera došlo vlivem gradace lýkožrouta smrkového od roku 2005 k postupnému rozpadu původního stromového patra, přičemž rozpad proběhl v různých částech povodí v rozdílných letech.

Výsledky se liší podle lesního vegetačního stupně (LVS). Nejmenší změny rostlinného společenstva byly pozorovány ve smrkovém stupni (8. LVS; obr. 4), přičemž k výraznější změně došlo mezi 5. a 6. rokem od kůrovcové disturbance. Plocha PJ4 (obr. 5) představuje sice 7. LVS, ale stanoviště je značně extrémní, takže i zde lze očekávat zhoršené růstové podmínky pro dřeviny, což se odrazilo v celkově nízké změně složení rostlinného společenstva. K první výraznější změně vegetace však zde došlo již 2 a 3 roky po disturbanci. Naproti tomu na ploše PJ2 (7. LVS; obr. 3) byla pozorována výrazná změna společenstva až cca mezi 7. a 8. rokem po disturbanci, což bylo spojeno s úplným zapojením nově vznikajícího dřevinného patra. Tato změna byla mnohem větší, nežli ty, které byly vidět ve vyšších polohách, kde nedošlo a zřejmě ani v budoucnu nedojde ke kompletnímu zapojení nového dřevinného patra. Obdobně v ještě níže položené ploše PJ1 (obr. 2) byla výrazná změna pozorována až cca 9. rok po disturbanci. Méně výraznou změnu je možno pozorovat cca 4 roky po disturbanci.

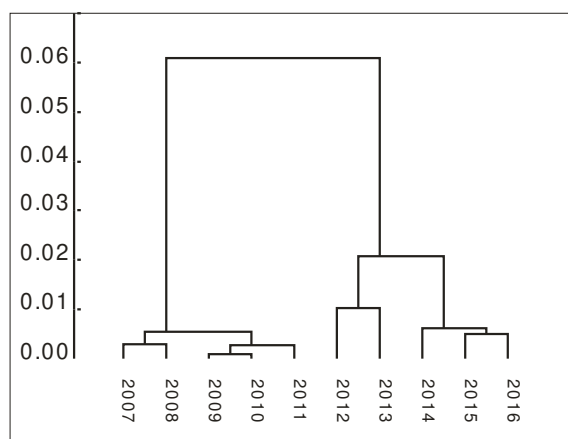
Tyto výsledky jsou v souladu se zjištěními, která byla publikována dříve (např. MATĚJKA, 2015): velký vliv na dynamiku vegetace smrkových lesů po rozpadu stromového patra má poloha plochy, respektive přiřazení k lesnímu vegetačnímu stupni.



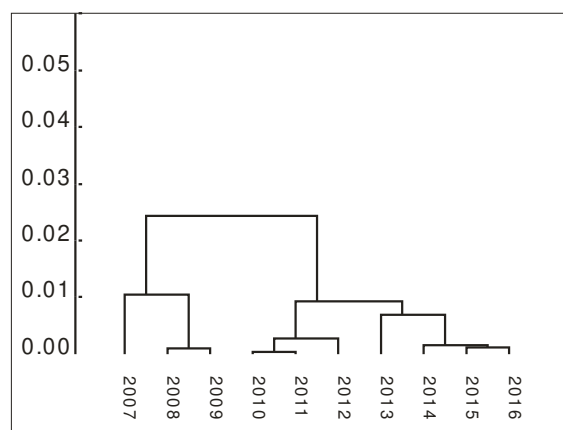
Obr. 2. Sekvenční klasifikace snímků na lokalitě PJ1 metodou HSCA/Sq. LVS 6/7. Rozpad stromového patra 2005-6.



Obr. 4. Sekvenční klasifikace snímků na lokalitě PJ3 metodou HSCA/Sq. LVS 8. Rozpad stromového patra 2005.



Obr. 3. Sekvenční klasifikace snímků na lokalitě PJ2 metodou HSCA/Sq. LVS 7. Rozpad stromového patra 2004.



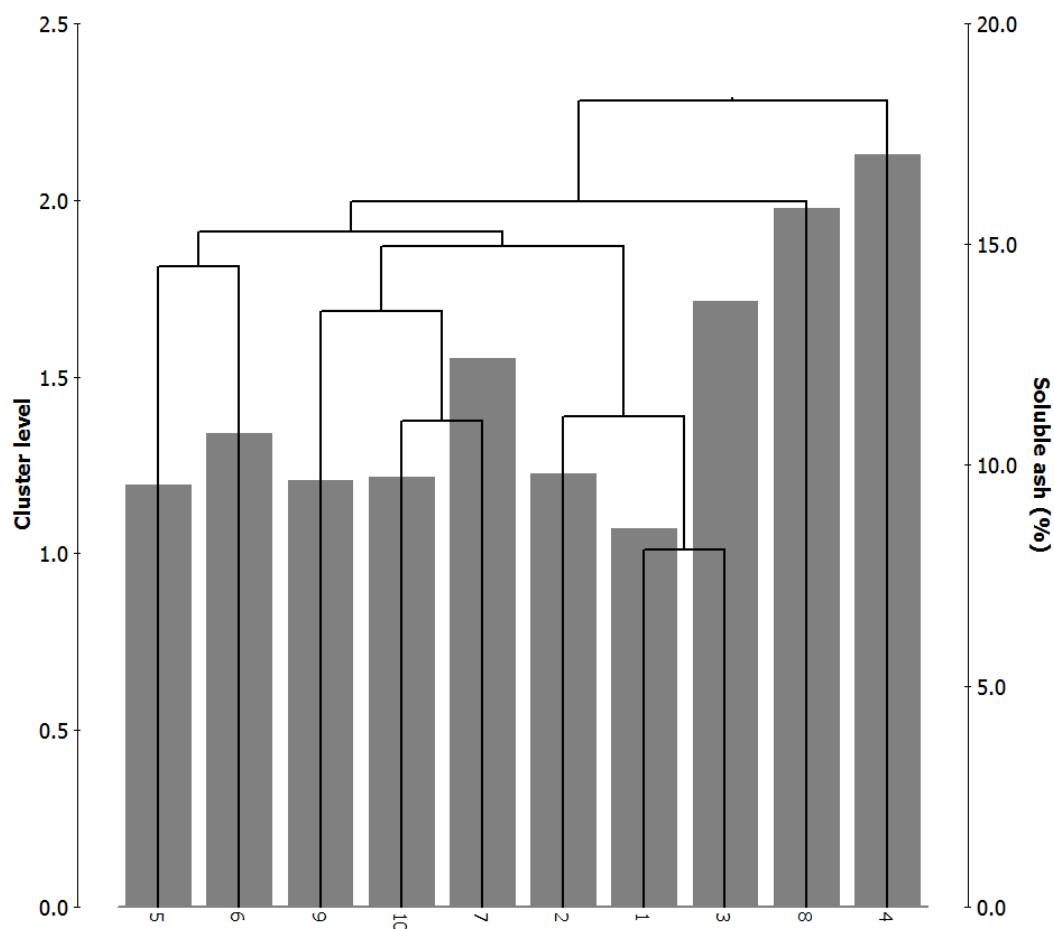
Obr. 5. Sekvenční klasifikace snímků na lokalitě PJ4 metodou HSCA/Sq. LVS 7. Rozpad stromového patra 2007-8.

### Porovnání výsledků klasifikace s externími daty

Při studiu teplomilných doubrav ČR bylo provedeno na lokalitě CK1 v Českém krasu šetření transektu vedoucího mezi bezlesím s rozptýlenými dřevinami a zapojeným lesem (MATĚJKA et al., 2016). Na deseti pravidelně rozmístěných místech byly sondou odebírány půdní vzorky povrchových horizontů do hloubky 10 cm. Ve vzorcích byli determinováni pancířníci (*Oribatida*). Zbýlý půdní vzorek byl použit pro chemickou analýzu. Vzorky byly po vysušení na vzduchu rozemlety a přesáty sítím s průměrem ok 2 mm (skelet byl odstraněn). Následně byl v naváženém vzorku (1 g) stanoven obsah celkových popelovin. Celkové popeloviny byly rozpuštěny v 1 ml koncentrované HCl. Tento výluh byl filtrován a zbytek i s filtrem byl opětovně spálen a tak byly stanoveny nerozpustné popeloviny, dopočítány byly rozpustné popeloviny. Provedeny byly i další analýzy, ale ty zde nejsou prezentovány.

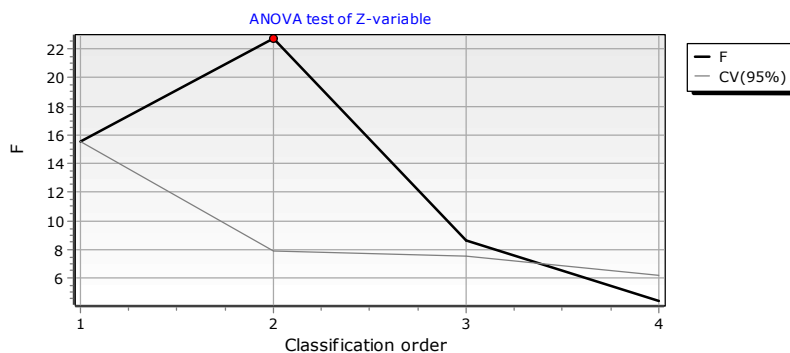
Otázkou je, jestli může existovat vztah mezi charakterem společenstva pancířníků a obsahem rozpustných popelovin v půdě. Společenstva odpovídající půdním vzorkům byla numericky klasifikována metodou průměrné vzdálenosti (obr. 6). Vztah mezi rozdělením vzorků do klasifikačních skupin podle získaného dendrogramu a obsahem rozpustných popelovin byl statisticky testován analýzou variance, přičemž hodnota F-testu byla vyhodnocena pomocí Monte Carlo permutačního testu (obr. 7). Výsledky ukazují, že výrazně (statisticky signifikantně,  $P = 99,6\%$ ) jsou odlišné vzorky při dělení na druhé nejvyšší

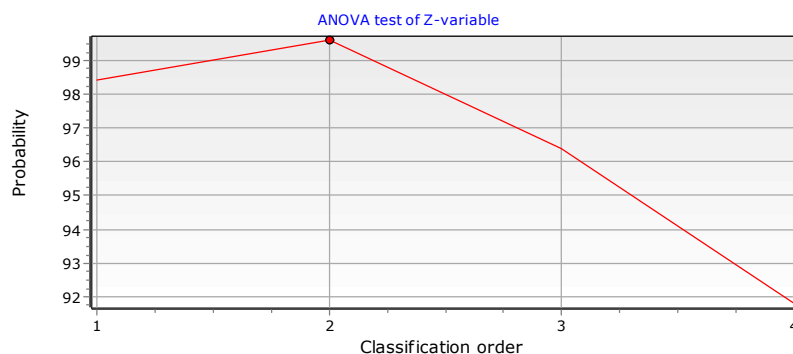
shlukovací hladině. Tato hladina odděluje vzorky 4 (součást vnějšího ekotonu sledovaného gradientu) a 8 (součást vnitřního ekotonu). Je tedy možno tvrdit, že v rámci sledovaných dat mají společenstva půdních pancířníků nejodlišnější strukturu v těch půdních vzorcích, kde je současně nalézán nejvyšší obsah rozpustných popelovin.



Obr. 6. Klasifikace společenstev pancířníků podél transektu na lokalitě CK1 srovnávaná s obsahem rozpustných popelovin v půdě. Metoda průměrné vzdálenosti s Euklidovskou distancí. Data početnosti byla před výpočtem logaritmicky transformována.

A

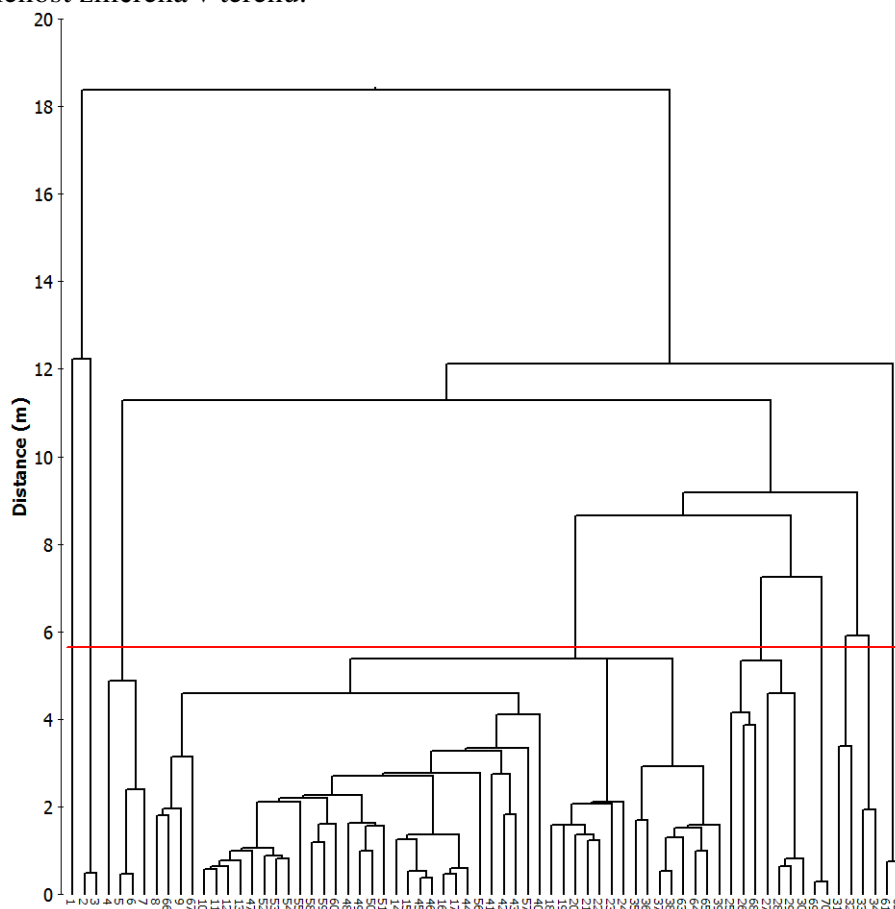


**B**

Obr. 7. Významnost vztahu klasifikace společenstev pancířníků (obr. 6) a obsahu rozpustných popelovin v půdě podél transektu na lokalitě CK1 hodnocená pomocí Monte Carlo testu statistiky F. A – porovnání vypočtené statistiky F a kritické hodnoty na hladině 95 %. B – Pravděpodobnost (%) testu. Vyznačena je nejvyšší dosažená pravděpodobnost  $P = 99,6\%$  pro druhou nejvyšší úroveň klasifikace.

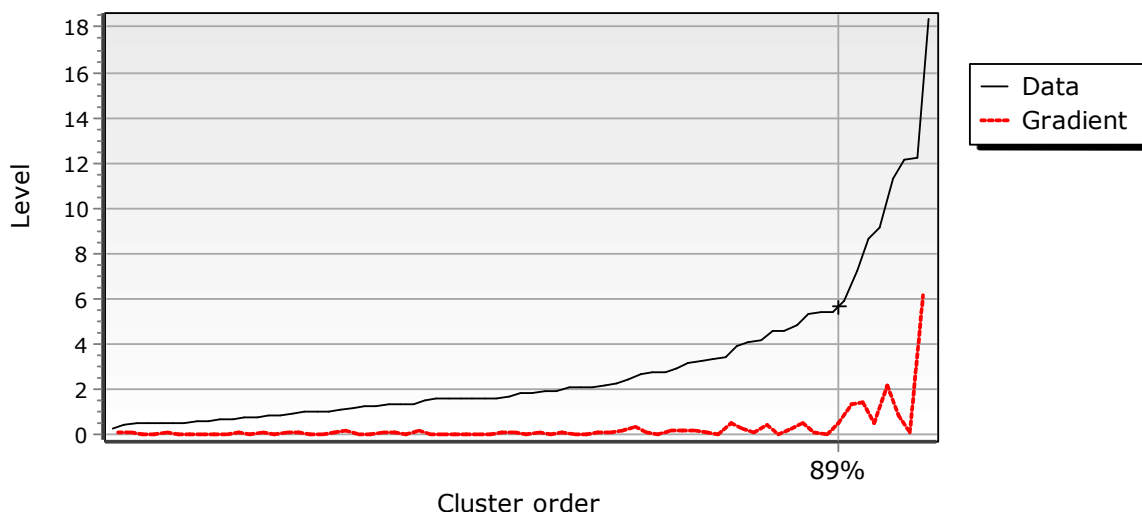
### Jaká je prostorová struktura stromového patra?

Běžně se řeší otázka, jestli je struktura stromového patra (tedy prostorové rozmístění stromů) pravidelná, náhodná nebo shlukovitá (např. CLARK et EVANS, 1954, PIELOU, 1959). V případě, že příslušným testem prokážeme shlukovitou prostorovou strukturu, má smysl pokládat otázku na velikost shluků. Tento úkol je opět řešitelný pomocí hierarchické klasifikace, přičemž nejvhodnější je užití metody nejbližšího souseda (nearest neighbor clustering), protože potenciální shluky mohou vykazovat různý, i silně protáhlý tvar (například může být smrk rostoucí v liniích dříve odumřelých padlých stromů, přičemž délka těchto linií může být větší, nežli je separační vzdálenost mezi shluky). Mírou distance je běžná vzdálenost změřená v terénu.



Obr. 8. Klasifikace stromů na ploše HH1 podle jejich skutečné vzdálenosti. Vyznačena je hladina 5,7 m.





Obr. 9. Závislost hladiny shlukování na pořadí shlukování (obr. 8) ukazuje na první výraznější změnu hladiny při hodnotě 5,7 m.

Následuje příklad šetření výzkumné plochy HH1 na Šumavě (MATĚJKA 2011b), která představuje porost smrku vzniklý spontánní sukcesí v bývalém bezlesí. Všechny stromy na kruhové ploše byly zaměřeny technologií FieldMap. Provedená klasifikace obr. 8) vykazuje poměrně plynule vzrůstající hladiny až do hodnoty 5,7 m, poté následuje výraznější vzestup následujících hladin, který ukazuje na prostorový odstup shluků stromů na výzkumné ploše (obr. 9). Bylo tak prokázáno nejen, že porost má shlukovitou strukturu, ale i to, že shluky jsou od sebe navzájem odděleny mezerami většími než 5,7 m.

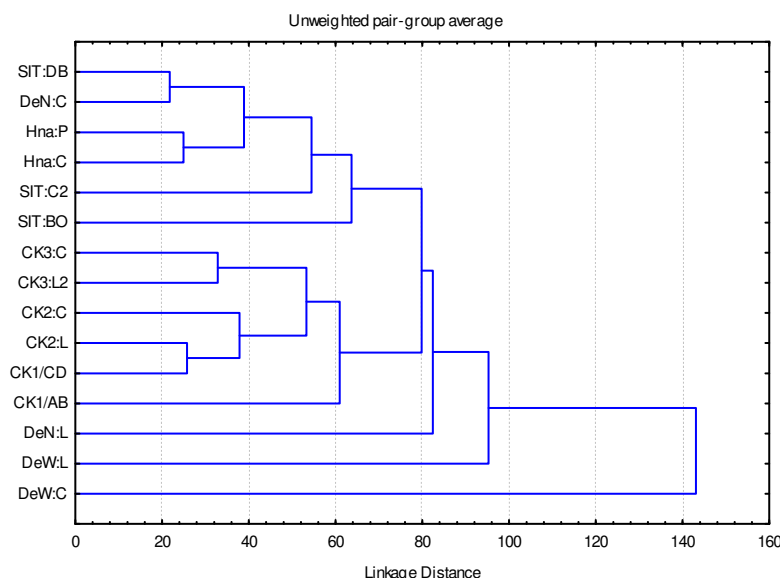
### Rozdíl půd na různých lokalitách

První příklad s výpočtem korelačních vzdáleností popisuje hodnocení různých lokalit nížinných lesů (MATĚJKA et al., 2016), na nichž byly sondou odebírány půdní vzorky povrchových vrstev do hloubky 10 cm. Na každé lokalitě bylo odebráno 5 vzorků, které byly po vysušení na vzduchu rozemlety a přesáty sítím s průměrem ok 2 mm (skelet byl odstraněn). Následně byl v naváženém vzorku stanoven obsah celkových popelovin. Celkové popeloviny byly rozpuštěny v 1 ml koncentrované HCl. Tento výluh byl filtrován a zbytek i s filtrem byl opětovně spálen a tak byly stanoveny nerozpustné popeloviny. Ve výluhu byly stanoveny Al, Fe, Ca, Mg a P. Celkový obsah Ca, Mg, K a P byl stanoven ze vzorků po jejich rozpuštění v lučavce královské.

Pro každou lokalitu byla vypočtena korelační matice mezi zjištěnými charakteristikami půdních vzorků. Rozdíly mezi korelačními maticemi byly počítány jako proměnná  $X^2$ . V tomto případě byla stanovena kritická hodnota  $X^2$  na hladině 95% rovná 73,3. Z toho též vyplývá hodnocení dendrogramu konstruovaného procedurou UPGA (obr. 10): Existují dva výrazné shluky lokalit - (a) převážně lokality s kyselými půdami, Sítovka (SIT), Hnanice (Hna) a kontrolní plocha lokality Děvín-sever (DeN:C) a (b) lokality karbonátových půd Českého krasu, CK1, CK2 a CK3. Mimo tyto dva shluky leží tři plochy z oblasti Děvína, které jsou výrazně odlišné. Je tedy vidět, že zatímco sledované české a moravské acidofilní lokality jsou si z hlediska zjišťovaných půdních procesů podobné, jsou lokality karbonátových půd Českého krasu a Děvína zřejmě signifikantně odlišné, přestože průměrné hodnoty jednotlivých půdních charakteristik mohou být obdobné.

Dobře lze interpretovat i některé další detaily patrné z dendrogramu, respektive z vypočtených distancí:

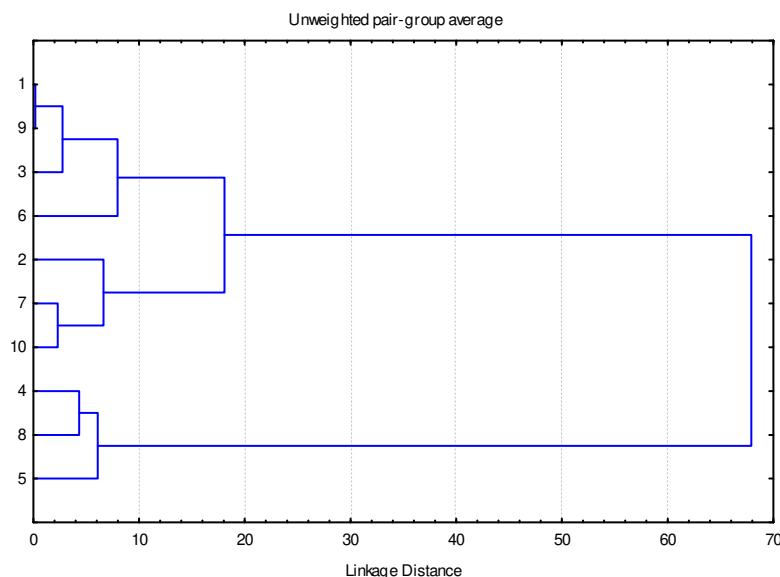
- Na lokalitě Sítovka jsou si výrazně podobnější vzorky z volné půdy (SIT:C2) se zbytky rozkládajícího se dřeva dubu (SIT:DB), nežli je tomu o vzorků se zbytky dřeva borovice (SIT:BO).
- V rámci Českého krasu má nejextrémnější pozici ze sledovaných lokalit bezlesí zachycené v segmentech A a B transektu na lokalitě CK1 (CK1/AB), což je dobře patrné i při porovnání rostlinných společenstev.



Obr. 10. Klasifikace výzkumných ploch v doubravách podle korelační struktury dat o chemickém složení povrchových půdních horizontů. Použito 11 chemických charakteristik. Kritická hodnota pro  $X^2$  na hladině 95% je 73.3161.

### Rozdíly subpopulací horského smrku

Obdobně, jako v článku MATĚJKA (2017a), je možno klasifikovat populace/subpopulace dřevin pouze na základě vybraných korelací. V následujícím byly stejné subpopulace *Picea abies* na deseti plochách v oblasti Medvědí hory na Šumavě, jako ve jmenovaném článku, klasifikovány na základě korelací čtyř dendrometrických charakteristik (logaritmy změřených hodnot DBH, výšky, korunové projekce a výšky nasazení koruny) s věkem stromu v roce 2009 hodnoceným na základě dendrochronologických vývrtů. Příslušná kritická hodnota ukazuje na fakt, že smysluplně lze hledat vysvětlení pro existenci tří typů subpopulací (obr. 11). Takto byly odlišeny především subpopulace na zamokřených půdách (plochy 4, 5 a 8).



Obr. 11. Klasifikace subpopulací *Picea abies* na výzkumných plochách v oblasti Medvědí hory na Šumavě. Použity byly korelace čtyř dendrometrických veličin s věkem v roce 2009. Kritická hodnota pro  $X^2$  na hladině 95% je 9.4927.

Program	Nápověda	Metoda	Příklad použití
DBreleve	MATĚJKA (2017c)	Monte Carlo testování diferencí	HRIB ET AL. (2017)
IsoEnz	MATĚJKA (2010)	Monte Carlo testování diferencí genetických parametrů populací	IVANEK, MATĚJKA (2016)
DBreleve	MATĚJKA (2017c)	Hodnocení výsledků klasifikace a porovnání s externími daty	
HSCA (součást DBreleve)	MATĚJKA (1993)	Klasifikace sekvenčně uspořádaných objektů	MATĚJKA (1993)
CorrelDist (součást DBreleve)	MATĚJKA (2017c)	Porovnání podmnožin objektů popsanych mnohorozměrnými daty	MATĚJKA (2017a)
SeqAn	MATĚJKA (2009)	Testování významnosti diferencí uspořádaných hodnot	IVANEK, MATĚJKA (2016)

Tab.1: Přehled uplatnění zmíněných matematicko-statistických metod.

## ZÁVĚR

Tento text se zabývá především uplatněním Monte Carlo permutačními testy, rozšířením možnosti shlukové analýzy a výpočtem korelačních distancí v mnohorozměrných datech. Všechny tyto postupy mohou najít široké uplatnění v rámci ekologického, geobiocenologického a podobného výzkumu. Většinou se nejedná o zcela nové metody, ale o jejich modifikace. Opomíjení Monte Carlo permutačních testů je dáno především jejich častou absencí ve využívaném statistickém software (například program Statistica) nebo jejich nepřítelným intuitivním použitím vyžadujícím podrobnější znalost metody (R; ROBERT ET CASELLA, 2010).

**Poznámka:** Presentace k tomuto textu obsahující některé další příklady je dostupná na webu [https://www.infodatasys.cz/presentation\\_cz/geobiocenologie2018.htm](https://www.infodatasys.cz/presentation_cz/geobiocenologie2018.htm).

## LITERATURA

- CLARK P.J., EVANS F.C. (1954): Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35: 445-453.
- ELLENBERG, H., WEBER, H.E., DÜLL, R., WIRTH, V., WERNER, W., PAULIBEN, D. (1992): *Zeigewerte von Pflanzen in Mitteleuropa*. 2nd Ed. Scripta Geobotanica XVIII. Göttingen, Verlag Erich Goltze KG, 258 p.
- GORDON, A.D. (1973): Classification in the presence of constraints. - *Biometrics*, 29: 821-827.
- IVANEK, O., MATĚJKA, K. (2016): Genetic structure of spruce population and forest management. URL: [https://www.infodatasys.cz/public/Ivanek\\_Matejka2016.pdf](https://www.infodatasys.cz/public/Ivanek_Matejka2016.pdf)
- LEGENDRE, P., LEGENDRE, L. (2012): Numerical ecology. 3rd Ed. Developments in Environmental Modelling, Vol. 24. Elsevier, Amsterdam etc., 990 p.
- LEPŠ, J., ŠMILAUER, P. (2003): Multivariate analysis of ecological data using CANOCO. Cambridge University Press, Cambridge, 269 p.
- MAGURRAN, A.E. (2004): Measuring biological diversity. Blackwell Publishing, Malden/Oxford/Carleton, 256 p.
- MATĚJKA, K. (1993): Hierarchical semi-cluster analysis (HSCA): a new method of gradient analysis. - *Ekológia Bratislava*, 12: 131-152.
- MATĚJKA, K. (2009): Náповѣда k programu SeqAn. Sequential data distribution analyse. URL: [https://www.infodatasys.cz/software/hlp\\_SeqAn/SeqAn.htm](https://www.infodatasys.cz/software/hlp_SeqAn/SeqAn.htm)
- MATĚJKA, K. (2010): Program IsoEnz. Databáze isoenzymových analýz (Isoenzyme analysis database). URL: [https://www.infodatasys.cz/software/hlp\\_isoenz/isoenz.htm](https://www.infodatasys.cz/software/hlp_isoenz/isoenz.htm)
- MATĚJKA, K. (2011a): Vegetace v povodí Plešného a Čertova jezera v letech 2007–2010. URL: <https://www.infodatasys.cz/proj002/results2010.pdf>
- MATĚJKA, K. (2011b): Management biodiversity v Krkonoších a na Šumavě - zpráva spoluřešitele za rok 2010. URL: <https://www.infodatasys.cz/biodivkrsu/IDSreport2010.pdf>
- MATĚJKA, K. (2015): Disturbance-induced changes in the plant biomass in forests near Plešné and Čertovo Lakes. - *Journal of Forest Science*, 61: 156-168.
- MATĚJKA, K. (2016): Sledování vegetace v povodí Plešného a Čertova jezera do roku 2016. URL: <https://www.infodatasys.cz/proj002/results2016.pdf>
- MATĚJKA, K. (2017a): Multivariate analysis for assessment of the tree populations based on dendrometric data with an example of similarity among Norway spruce subpopulations. - *Journal of Forest Science*, 63: 449-456.
- MATĚJKA, K. (2017b): Náповѣда k programu PlotOA. URL: [https://www.infodatasys.cz/software/hlp\\_PlotOA/PlotOA.htm](https://www.infodatasys.cz/software/hlp_PlotOA/PlotOA.htm)
- MATĚJKA, K. (2017c): Náповѣда k programu DBreleve. Databáze fytocenologických snímků, verze 2.5. URL: [https://www.infodatasys.cz/software/hlp\\_dbreleve/dbreleve.htm](https://www.infodatasys.cz/software/hlp_dbreleve/dbreleve.htm)
- MATĚJKA, K., STARÝ, J. (2009): Differences in top-soil features between beech-mixture and Norway spruce forests of the Šumava Mts. - *Journal of Forest Science*, 55(12): 540-555.
- MATĚJKA, K., STARÝ, K., BOHÁČ, J., LEPŠOVÁ, A., ŠPULÁK, O. (2016): Ukázkové a výzkumné plochy pro sledování vlivu managementu v lesích chráněných území. URL: <https://www.infodatasys.cz/BiodivLes/Demoobjects.pdf>
- MCCUNE, B., GRACE, J.B. (2002): Analysis of ecological communities. MjM Software Design, Gleneden Beach, Oregon, 300 p.
- PIELOU, E.C. (1959): The use of point-to-plant distances in the pattern of plant populations. *Journal of Ecology*, 47: 607-613.
- ROBERT, C.P., CASELLA, G. (2010): Introducing Monte Carlo methods with R. Springer, New York etc., 283 p.

WEGMANN, D., LEUENBERGER, C., NEUENSCHWANDER, S., EXCOFFIER, L. (2010): ABCtoolbox: a versatile toolkit for approximate Bayesian computations. - BMC Bioinformatics, 11.