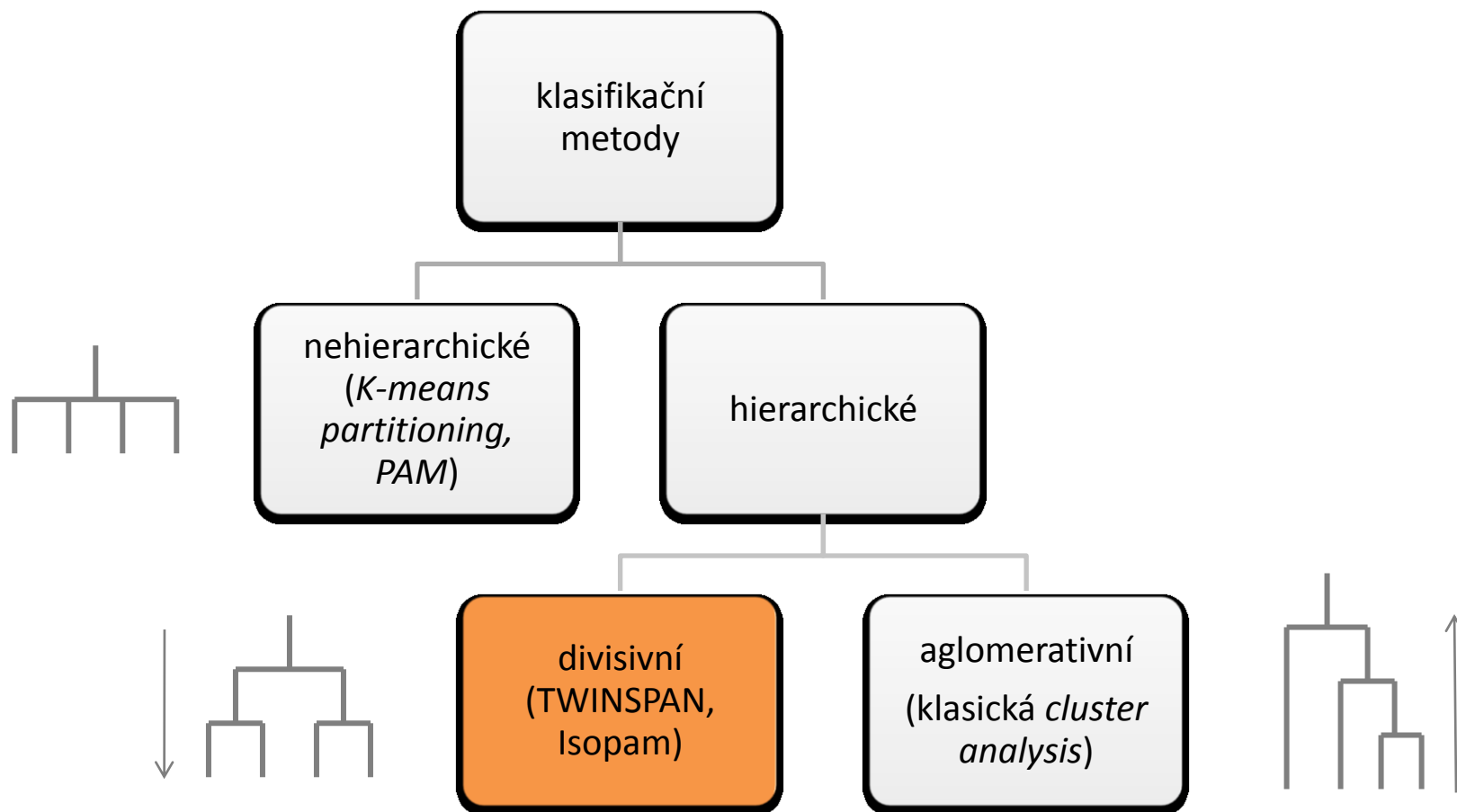


KLASIFIKACE



TWINSPAN

TWO WAY INDICATOR SPECIES ANALYSIS

- divisivní metoda
 - začíná dělením celého souboru vzorků a postupuje směrem dolů
 - skupina se dále nedělí, pokud je příliš malá, nebo bylo dosaženo dělení do požadované úrovně
- polytetická metoda
 - každé dělení závisí na **několika (indikačních) druzích** (x monotetická metoda – dělení ovlivňuje jediný druh)
- metoda velmi oblíbená mezi vegetačními ekology
 - ale – algoritmus je poměrně složitý, ne zcela popsáný a s řadou arbitrárních kroků. Proto má také řadu zarytých odpůrců: "TWINSPAN too unstable and tricky: Better avoided." (Jari Oksanen)
- vzorky jsou uspořádány podle první osy korespondenční analýzy (CA) a podle ní jsou rozděleny do dvou shluků (vzorky s pozitivním skóre a negativním skóre)
- metoda ošetří vzorky, které leží blízko středu osy, a které tak mají velkou pravděpodobnost, že budou špatně klasifikovány

TWINSPAN

TWO WAY INDICATOR SPECIES ANALYSIS

- *pseudospecies*
 - metoda primárně funguje pro **kvalitativní** data
 - kvantitativní informace se dodává rozdělením druhů na *pseudospecies* podle relativní abundance (*cut levels*), např. 1, 5, 10, 20 %.

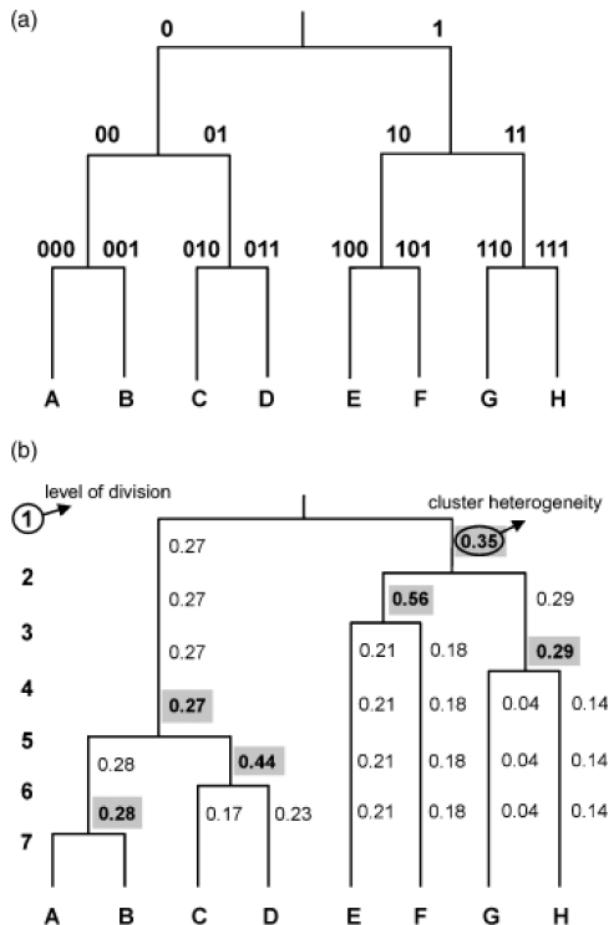
	Species	Sample 1	Sample 2
Original table	<i>Cirsium oleraceum</i>	0	1
	<i>Glechoma hederacea</i>	6	0
	<i>Juncus tenuis</i>	15	25
Table with pseudo-species used in TWINSPAN	Cirsoler1	0	1
	Glechede1	1	0
	Glechede2	1	0
	Junctenu1	1	1
	Junctenu2	1	1
	Junctenu3	1	1
	Junctenu4	0	1

TWINSPAN

TWO WAY INDICATOR SPECIES ANALYSIS

- *pseudospecies*
 - metoda primárně funguje pro **kvalitativní** data
 - kvantitativní informace se dodává rozdělením druhů na *pseudospecies* podle relativní abundance (*cut levels*), např. 1, 5, 10, 20 %.
- výsledkem je (mimo jiné) tabulka podobná fytoecenologické
 - snímky z určitých klastrů a druhy s vysokou fidelitou k dané skupině jsou seskupeny dohromady
- metoda vhodná v případě, že jsou data strukturovaná podle jednoho výrazného gradientu
- vhodné na hledání (několika málo) ekologicky interpretovatelných skupin v datech
- Možnost klasifikovat snímky i druhy
 - Two-way metoda
 - Využívá symetričnosti chi-square distance

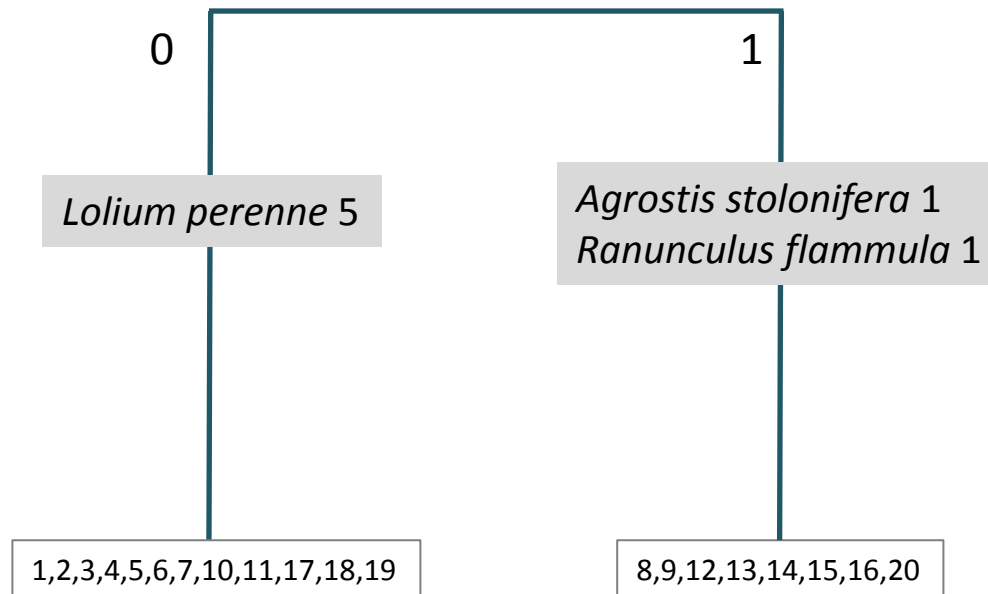
MODIFIKOVANÝ TWINSPAN (ROLEČEK ET AL. 2009)



- na rozdíl od původního algoritmu (a) umožňuje modifikovaný TWINSPAN (b) dopředu stanovit cílový počet skupin
- algoritmus se po každém dělení na dvě skupiny rozhoduje, kterou ze skupin bude dále dělit – vybere tu, která je více „heterogenní“ na základě její betadiverzity
- míru betadiverzity je nutné zvolit (např. Jaccardův index podobnosti)
- V R je možné využít funkci `cuth` v package `twinspan` (ovšem pouze s využitím `total-chisq distance` jako míry nepodobnosti).

TWINSPAN

- kde je to možné, dělení je doplněno indikátorovými druhy



TWINSPAN SOFTWARE

- Originální program v jazyku FORTRAN (Mark Hill 1979)
 - Zcela zastaralý programovací jazyk
 - Algoritmu už přesně nikdo nerozumí (ani původní autor)
 - Algoritmus nelze převést do jiných jazyků (přes snahu nejlepších mozků ve vegetační ekologii)
 - Implementace jinde (R, Juice): pouze wrapper okolo původního Fortran algoritmu
- <https://github.com/jarioksa/twinspan>
 - Autor Jari Oksanen
 - Klasický Twinspan, modifikovaný pouze s omezenými možnostmi
 - Celkem dobře implementováno včetně výstupů
- <https://www.sci.muni.cz/botany/juice/>
 - Program Juice – Luboš Tichý
 - Klikací okna, nutno importovat data do Juice formátu

ISOPAM

- *Isometric feature mapping and partitioning around medoids*
- Kombinuje ordinační metodu Isomap a PAM
 - Isomap – PCoA, která se zaměřuje pouze na nepodobnosti mezi podobnými vzorky. Následně analyzuje nepodobnosti mezi takto vzniklými skupinami
- Hierarchické uspořádání
 - Uživatel volí c.max (maximální počet clusterů v každém dělení) a l.max (maximální počet úrovní)
 - Isomap zkouší různé kombinace a zdaleka ne vždy dojde až k c.max/l.max
 - Sieve – má-li se klasifikace zaměřit na nepodobnosti pouze mezi indikačně významnými druhy; lze nastavit T/F
 - Posiluje význam presence/absence
 - Velmi snižuje význam variability dominance druhů s vysokou frekvencí

ISOPAM

- Vytváří pěkné výsledky – zvlášť jsou-li promítány do PCoA
- Dobře implementovaný v R
- Pěkné výstupy včetně popisu shluků
- Náročný na výkon počítače
 - Maximálně jednotky tisíc vzorků

Journal of Vegetation Science 21: 1162–1171, 2010
DOI: 10.1111/j.1654-1103.2010.01221.x
© 2010 International Association for Vegetation Science

A brute-force approach to vegetation classification

Sebastian Schmidlein, Lubomír Tichý, Hannes Feilhauer & Ulrike Faude

Abstract

Aim: Introduction of a novel approach to the classification of vegetation data (species by plot matrices). This approach copes with a large amount of noise, groups irregularly shaped in attribute space and species turnover within groups.

field ecologists. The derived classes often do not match expert guesses regarding ecologically meaningful partitions. This mismatch may be an explanation for the fact that handcrafted vegetation tables frequently continue to be published, even in times of teraflop computing.