Poznámky k používání průměrných Ellenbergových indikačních hodnot při analýze vegetačních dat

Notes to the use of mean Ellenberg indicator values in vegetation analyses

David Z e l e n ý

Ústav botaniky a zoologie, Přírodovědecká fakulta, Masarykova univerzita, Kotlářská 2, CZ-611 37 Brno; e-mail: zeleny@sci.muni.cz

Abstract

Ellenberg indicator values (EIV) represent valuable expert information about ecological behaviour of vascular plant species in western and central part of Europe. The mean of EIV for species in vegetation relevé or floristic inventory are considered to be a good proxy of habitat characteristics and relevant surrogate for measured environmental factors, especially when the latter are not available. However, as mean EIV are derived from data about species composition, caution must be applied when using them in analyses with other variables derived from the same data set. Such analyses include, e.g., correlation of mean EIV with ordination axes in unconstrained ordination, analysis of variance of mean EIV between vegetation types or correlation of mean EIV and species richness. As a consequence, such analyses tend to return biased and more optimistic results, especially if their significance was tested using traditional tests, and their interpretation may be therefore inappropriate. In this study, examples using real vegetation data are given and possible remedy is introduced. Due to the biased results, mean EIV are not recommended for use in abovementioned analyses, unless appropriate modified permutation tests are applied.

Keywords: analysis of variance, circularity of reasoning, permutation test, unconstrained ordination

Úvod

Ellenbergovy indikační hodnoty (EIH, Ellenberg et al. 1992) představují soubor hodnot pro druhy cévnatých rostlin, které vypovídají o pozici jejich realizovaného životního optima podél základních ekologických gradientů jako je světlo, teplota, kontinentalita, vlhkost, živiny, půdní reakce a také salinita. Koncept indikačních hodnot vychází z běžné terénní zkušenosti, že řada druhů je svým výskytem vázána na stanoviště, která svými vlastnostmi vyhovují jejich ekologickým nárokům. Pokud tedy známe ekologické nároky jednotlivých druhů, jsme z nich do určité míry schopni odvodit ekologické vlastnosti stanovišť, na kterých tyto druhy rostou. Heinz Ellenberg (1913–1997), německý vegetační ekolog, nebyl první, který se úvahami o indikačních hodnotách rostlinných druhů zabýval. Zmínky o indikačním potenciálu různých druhů rostlin se objevují už na začátku 20. století (např. Cajander 1926), a v padesátých letech o nich uvažoval v souvislosti s vývojem přímé ordinační analýzy i Robert Whittaker (1920–1980), americký vegetační ekolog (Whittaker 1956). Pro českou vegetační vědu je však stěžejní právě práce H. Ellenberga, který propracoval systém indikačních hodnot pro rostlinné druhy, které se vyskytují v Německu a přilehlých státech (Ellenberg 1974, Ellenberg et al. 1992). Indikační hodnoty druhů jsou na ordinální škále od jedné do devíti, s výjimkou vlhkosti, která má stupňů dvanáct (poslední tři stupně jsou pro druhy do různé míry ponořené ve vodě). Nejedná se přitom o přesná měření, ale o odhady, které jsou, ale spíše výjimečně, upřesněny výsledky kultivačních experimentů a měření v terénu. Každá hodnota má svoje slovní vyjádření, které například u EIH pro teplotu odráží rozšíření druhů ve vztahu k nadmořské výšce a zeměpisné šířce a vypadá takto: 1 – rostlina indikující nízké teploty, rostoucí pouze v horách, v alpínském nebo niválním stupni nebo v boreálně-arktických oblastech; 3 – indikátor chladného podnebí, většinou subalpínského stupně; 5 – indikátor poměrně teplého podnebí, vyskytující se od nížin do hor, především ale v submontánním stupni; 7 – indikátor teplého podnebí v nížinách a kolinním stupni; 9 – indikátor extrémně teplého podnebí, z Mediteránu zasahuje jen do nejteplejších oblastí horního Porýní; stupně 2, 4, 6 a 8 pak představují prosté mezistupně mezi výše uvedenými (původní hodnoty jsou odvozeny pro Německo, proto to Porýní u stupně 9). Podobně kontinentalita vyjadřuje rozšíření druhů na gradientu od pobřeží Atlantiku po vnitrozemí Eurasie. Zbývající faktory nevyjadřují geografické rozšíření druhů, ale jejich výskyt na stanovišti určitých vlastností. Nároky jednotlivých druhů na světlo odrážejí stupeň zastínění stromovým patrem, které

druhy tolerují (a který se v případě druhů dřevin vztahuje k chování jejich semenáčků), vlhkost odráží zamokřenost stanoviště, živiny a půdní reakce zase charakter půdy a půdotvorného substrátu. Pokud druh nemá pro některý z gradientů definovánu žádnou hodnotu, je to většinou proto, že se ve vztahu k danému gradientu chová nevyhraněně a není snadné určit jeho optimum (případně chybí dostatek informací).

Používání Ellenbergových indikačních hodnot se u nás (a prakticky v celé Evropě) stalo rutinní záležitostí a neobejde se bez nich téměř žádná popisná vegetační práce. Různé aspekty jejich používání byly v minulosti často kritizovány i obhajovány, například jestli jsou použitelné i mimo střední Evropu, které měřitelné proměnné prostředí vlastně nahrazují nebo jestli počítat průměr EIH vážený abundancí druhů nebo průměr neválený (pro souhrn této problematiky viz Diekmann 2003). V rámci Evropy existují i další soubory indikačních hodnot, které jsou kalibrovány pro použití v jiných územích. Pro Českou republiku jsou zajímavou alternativou například Borhidiho indikační hodnoty publikované pro Maďarsko (Borhidi 1995), protože zahrnují řadu moravských druhů, které se v Německu nevyskytují a v souboru Ellenbergových indikačních hodnot proto chybějí. Výpočtem průměrných EIH pro druhy přítomné ve fytocenologickém zápisu (nebo floristickém soupisu) je možné získat odhad vlastností stanoviště (respektive lokality), na kterém byl zápis (soupis) pořízen. Pokud nejsou současně se sběrem vegetačních dat měřeny nebo odhadovány hodnoty proměnných prostředí, představují průměrné EIH jejich potenciální náhradu (např. Schaffers & Sýkora 2000), a protože ve většině popisných vegetačních studií jde o vztah mezi vegetací a faktory prostředí, jedná se o náhradu více než potřebnou. Stále častěji jsou ale průměrné EIH používány také pro složitější a sofistikovanější analýzy vegetačních dat, ve kterých už nejde o jednoduchý popis ekologických podmínek stanoviště. Jejich součástí bývají statistické testy, které analyzují vztah mezi průměrnými EIH a dalšími proměnnými odvozenými z vegetačních dat. Problém, který je v těchto analýzách ukryt, bude detailně rozebrán v následujících řádcích.

Průměrné Ellenbergovy indikační hodnoty a problém argumentace kruhem

Nejčastější kritika průměrných Ellenbergových indikačních hodnot naráží na fakt, že jejich použití do určité míry připomíná argumentaci kruhem. V širším smyslu jde o to, že vlastní

publikované indikační hodnoty jsou odvozeny z chování rostlin odpozorovaného v přírodě, konkrétně jejich výskytu na stanovištích, které mají určité ekologické parametry. Z vlastností těchto stanovišť tedy byly odvozeny vlastnosti druhů, ze kterých my opět odvozujeme vlastnosti stanoviště, v tomto případě námi studovaného. Na kritiku tohoto postupu se dá ale namítnout, že vlastní indikační hodnoty byly odhadovány v jiném kontextu (geografickém, krajinném, nakonec snad i v jiném měřítku) než jaký má námi studovaná vegetace, a problém kruhové argumentace v tomto smyslu je tak spíše jen teoretickou záležitostí. V užším smyslu se argumentace kruhem týká situace, kdy vlastním datový soubor fytocenologických zápisů z různých stanovišť a pro jednotlivé zápisy vypočtu jejich průměrné EIH. Pokud takto vypočtené hodnoty použiji nějakým způsobem k interpretaci těch samých vegetačních dat, ze kterých jsem je vypočetl, argumentace kruhem přestává být teoretickým problémem. Díky aritmetickému způsobu, jakým jsou průměrné EIH počítány, obsahují totiž tyto hodnoty více informace, než bychom čekali.

Průměrné Ellenbergovy indikační hodnoty a podobnost mezi vegetačními snímky

Pokud dva vegetační zápisy zahrnují úplně stejné druhy, jejich vypočtená průměrná EIH musí být přesně stejná (pokud používáme průměr vážený pokryvnostmi jednotlivých druhů, platí to v okamžiku, kdy mají jednotlivé druhy i stejnou pokryvnost). Pokud se dva zápisy liší jedním nebo několika málo druhy, na rozdílu v jejich průměrných EIH se to příliš neprojeví – průměrná hodnota spočítaná např. z 15 hodnot (druhů) se při změně jedné nebo několika málo hodnot příliš nezmění. Přitom se sice dá předpokládat, že stanoviště, na kterých byly pořízeny dva identické vegetační zápisy, se nebudou výrazně lišit podmínkami prostředí, je ale nepravděpodobné, že by vlastnosti těchto stanovišť byly naprosto stejné a při měření faktorů prostředí bychom naměřili přesně stejné hodnoty. Tímto se vypočtené průměrné EIH liší oproti měřeným hodnotám faktorů prostředí, protože ve vypočtených hodnotách se kromě ekologických vlastností stanoviště, které jsou odvozeny z vlastností jednotlivých druhů, *navíc* odráží i podobnost v druhovém složení mezi oběma stanovišti. Můžeme si to představit tak, že průměrná EIH má dvě informační složky: první, odvozenou z externí informace přiřazené jednotlivým druhům (Ellenbergovy indikační hodnoty), a druhou, odrážející vlastní druhové složení vegetačních zápisů. Pouze první informační složka je přitom

relevantní při použití průměrných EIH pro ekologickou interpretaci vegetace (Zelený & Schaffers 2012).

Dá se namítnout, že druhové složení přeci odráží ekologické vlastnosti stanoviště, a je proto pochopitelné, že se projeví při výpočtu průměrných EIH. To je pravda, ale jen částečná. Druhové složení společenstva na daném stanovišti je výsledkem dvou hlavních typů procesů: deterministických, tedy těch, které můžeme nějakým způsobem předpovědět, a stochastických, které předpovědět nedokážeme (např. Chase 2003). Deterministické procesy souvisí s teorií ekologické niky, která předpokládá, že jednotlivé druhy se od sebe liší svými ekologickými nároky, ve společenstvu obsazují různé niky a tím unikají před konkurencí ostatních druhů. Pokud tedy známe ekologické vlastnosti stanoviště, můžeme z nich odvodit i spektrum druhů, které se na stanovišti mohou vyskytovat (*niche assembly*); na stejných předpokladech ostatně stojí i vlastní princip použití EIH. Pak jsou tu ale ještě procesy stochastické, které souvisí jednak s šířením druhů v prostoru, a jednak s tzv. ekologickým driftem. Stanoviště může být pro daný druh ideální z ekologického hlediska, ale druh na něm neroste, protože se na něj z různých důvodů nedostal, třeba kvůli izolovanosti stanoviště nebo špatné schopnosti druhu šířit se (*dispersal limitation*). Nebo se naopak na stanovišti může vyskytnout druh, který na něm normálně neroste, protože jeho ekologická nika je jinde, ale na stanoviště se dostal a přežívá zde. Stane se to třeba tak, že příhodné stanoviště se zdrojovou populací tohoto druhu leží někde v blízkém okolí, a díky neustálému přísunu diaspor se druh dostane i na stanoviště námi studované. Druh zde možná ani nekvete nebo neplodí a obnovení jeho populace je závislé na pravidelném přísunu diaspor ze sousedství (*sink-source dynamic*, *spatial mass effect*), na stanovišti a potažmo v našem vegetačním zápisu druh ale prostě je. Tak zvaný ekologický drift pak představuje soubor náhodných procesů, které fungují na úrovni jednotlivých populací. Populace nějakého druhu může na daném stanovišti například úplně vymizet, i když příhodné podmínky zde přetrvávají. Stane se to prostě náhodou, přičemž je to podobný typ náhody, jako když člověk opakovaně hází korunou – pokud padne panna, počet jedinců v populaci vzroste o jednoho, pokud padne orel, o jednoho jedince klesne. Pokud máte pocit, že při hodu mincí se víceméně pravidelně střídá panna a orel a populace tak zůstává stabilní, je na čase vzít minci do ruky a přesvědčit se o opaku.

Vztah mezi druhovým složením a ekologickými vlastnostmi stanoviště není tedy tak přímočarý, jak by se mohlo na první pohled zdát. Výpočet průměrných EIH však s ničím takovým nepočítá a vychází pouze z informace o druhovém složení. Ve výsledku tak průměrné EIH odrážejí podobnost ve druhovém složení mezi stanovišti, přičemž jen část této podobnosti je způsobena skutečnou ekologickou podobností těchto stanovišť. Vypočtené hodnoty průměrných EIH pro dvě stanoviště jsou si tím pádem jaksi více podobné než stanoviště samotná. Principiálně to nijak nesnižuje význam průměrných EIH jako odhadů stanovištních vlastností – jde přeci jen o odhady, a toto je další důvod, proč je brát s rezervou a jejich význam ani přesnost nepřeceňovat. Problém ale nastává v okamžiku, kdy jsou průměrné EIH použity v analýze, které se spolu s nimi účastní i další proměnné přímo odvozené z druhového složení.

Průměrné Ellenbergovy indikační hodnoty jako pasivní proměnné v nepřímé ordinaci

Příkladem takové analýzy je nepřímá ordinace vegetačních zápisů, která je součástí řady vegetačních studií. Cílem nepřímé ordinace je snaha o odhalení skrytých gradientů v druhovém složení a objasnění, které faktory prostředí tyto gradienty způsobují. Průměrné EIH se zde používají k interpretaci jednotlivých ordinačních os v případě, že nemáme žádné měřené faktory prostředí (často ale i pokud je máme). Většinou se to provádí pasivním promítnutím průměrných EIH do ordinačního diagramu, což znamená, že průměrné EIH pro jednotlivé vegetační zápisy jsou korelovány se skóre (pozicemi) těchto zápisů podél hlavních ordinačních os. Tento postup navrhl Persson (1981) jako efektivní nástroj pro interpretaci ordinačních diagramů, a od té doby se hojně používá. Pro měřené faktory prostředí bývá zvykem, že průkaznost jejich korelace s ordinačními osami je otestována a pouze průkazné faktory jsou následně interpretovány (a vykresleny jako vektory do ordinačního diagramu). Stejný postup se začal používat i pro průměrné EIH, i když takový test vlastně porušuje základní předpoklad pro testování průkaznosti korelace, totiž že obě korelované proměnné musí být na sobě nezávislé (připomínám, že průměrné EIH jsou korelovány se skóre jednotlivých ploch na hlavních ordinačních osách, které jsou – stejně jako samotné průměrné EIH – vypočteny přímo z druhového složení). Na tento problém ostatně upozorňují Herben & Münzbergová (2003) ve svých skriptech o analýze vegetačních dat. Vegetační ekologové si

ale často s porušováním statistických pravidel hlavu moc nelámou, možná proto, že jim přijdou zbytečně přísná. Podívejme se tedy, jak taková analýza a její výsledky vypadají.

Vegetační data použitá v následujících analýzách byla sbírána na transektech napříč údolím Vltavy severně od Zlaté Koruny v jižních Čechách. Jde o fytocenologické snímky lesní vegetace, zapisované na ploše 10 × 15 m, a společně s nimi byly na stejných plochách měřeny nebo odhadovány některé faktory prostředí. Celkem bylo zaznamenáno 97 fytocenologických snímků na 27 transektech, umístěných v různých úsecích údolí. Podrobný popis sběru dat a metodiky měření faktorů prostředí najde čtenář jinde (Zelený & Chytrý 2007, Zelený 2008), pro účely tohoto článku není důležitý. Pro nepřímou ordinaci dat z údolí Vltavy byla použita detrendovaná korespondenční analýza (DCA, Hill & Gauch 1980). Původně kvantitativní druhová data byla před analýzou převedena na data presenčně-absenční. Všechny statistické analýzy byly provedeny v programu R (R Development Core Team 2011, Vienna, Austria); DCA byla spočtena funkcí *decorana*, která je součástí knihovny *vegan* (Oksanen et al. 2011), s použitím přednastavených parametrů (pro dentrendaci a přeškálování ordinačních os bylo použito jejich rozdělení na 26 segmentů). Vztah mezi průměrnými EIH pro jednotlivé faktory a prvními dvěma ordinačními osami byl spočten pomocí funkce *envfit* taktéž z knihovny *vegan*; nejedná se přitom o výpočet korelace s jednotlivými ordinačními osami, jak bývá zvykem, ale o mnohonásobnou regresi, ve které jsou skóre snímků na první a druhé ose použity jako vysvětlující proměnné a průměrná EIH jako závisle proměnná. Důležitými hodnotami, které tato analýza udává, jsou 1) množství variability v průměrných EIH vysvětlené ordinačními osami (vyjádřené jako koeficient determinace R^2), 2) výsledek permutačního testu, který ukazuje, jestli je vztah mezi ordinačními osami a průměrnou EIH pro daný faktor průkazný nebo ne, a 3) standardizované regresní koeficienty pro jednotlivé ordinační osy (nejedná se ale o korelační koeficienty s těmito osami – ty získáme tak, že standardizovaný regresní koeficient pro danou osu vynásobíme koeficientem determinace). Interpretace výsledků mnohonásobné regrese je ale stejná jako interpretace korelací, a pro jednoduchost proto budu v dalším textu místo správného termínu „mnohonásobná regrese mezi skóre jednotlivých vegetačních zápisů na ordinačních osách a průměrnou EIH“ používat termín „korelace průměrné EIH s ordinačními osami“.

<Tab. 1 a Obr. 1 umístit někam sem>

Průměrné EIH pro všech šest faktorů prostředí jsou průkazně korelovány s prvními dvěma ordinačními osami (tab. 1, P_{orig}), a až na kontinentalitu je hodnota průkaznosti vysoká ($P < 0{,}001$). Všechny proměnné jsou proto zobrazeny v ordinačním diagramu (obr. 1a), a interpretace ordinačních os se přímo nabízí: první osa představuje kombinovaný gradient

živin a vlhkosti, druhá teploty a světla, a obě osy jsou korelovány s půdní reakcí i kontinentalitou. Otevírá se také široký prostor pro úvahy, proč tomu tak vlastně je. Vypadá to, že hlavní ordinační osa představuje gradient způsobený pozicí vegetačních ploch v rámci údolního svahu, tedy kontrastem mezi vlhkými a živinami bohatými lužními a suťovými lesy při bázích svahů a acidofilními bory a doubravami na neúživných a vysychavých půdách horních částí svahů. Druhá osa zřejmě představuje gradient mezi jižněji orientovanými teplými a rozvolněnými porosty teplomilných doubrav a severně orientovanými chladnými a zapojenými porosty jedlin. Půdní reakce je korelovaná s oběma osami – bazično je jak při bázi svahů, kam stékají živiny, tak v teplomilných doubravách, které se zde vyskytují na půdách obohacených vápníkem. Interpretovat vliv kontinentality je na tak malém území docela problém, její signifikantní korelace může být ale například způsobena tím, že kontinentální druhy spíše rostou na suchých a výhřevných stanovištích v horních částech svahů s výraznějšími teplotními výkyvy, zatímco oceánicky laděné druhy budou spíše na vlhčích bázích svahů.

Interpretace výsledků je sama o sobě docela zajímavá vědecká disciplína. Při troše nadsázky se dá říci, že interpretace se najde prakticky pro jakýkoli výsledek, chce to jen trochu úsilí a fantazie. Čas od času se stane, že člověk na poslední chvíli najde chybu v analýze, která dávala pěkné výsledky, a po jejím přepočítání vyjdou výsledky přesně opačné. Někdy je pak jen otázka okamžiku, než se pro nové výsledky najde interpretace, která může znít ještě přesvědčivěji než ta předchozí. Zkusme si nyní udělat krátký myšlenkový pokus. Kolega nám svěřil svoje data (třeba vegetaci makrofyt v sérii různě eutrofních vodních nádrží), abychom mu spočetli a otestovali korelaci průměrných EIH s ordinačními osami. Trochu si z něj vystřelíme a před vlastním výpočtem průměrů zpřeházíme Ellenbergovy indikační hodnoty mezi druhy v tabulce. Výsledkem bude například to, že druhy, které jsou ve skutečnosti náročné na živiny, budou mít nové hodnoty pro živiny jak vysoké, tak nízké, a vlastní ekologický smysl této proměnné se tak vytratí. Nebo můžeme EIH rovnou nahradit náhodnými čísly, kolega to stejně nepozná. Dál postupujeme normálně – z každého snímku spočítáme průměrné EIH pro všech šest faktorů, vypočteme jejich korelaci s ordinačními osami a tuto korelaci otestujeme. K našemu překvapení čtyři z šesti proměnných stejně vyjdou průkazné. Kolega nám poděkuje a na výsledky obratem vymyslí interpretaci. Bude si sice trochu lámat hlavu, proč zrovna kontinentalita je u jeho vegetace makrofyt tak

důležitým faktorem, ale když už vyšla tak pěkně signifikantně, něco na tom asi je. Na následujícím příkladu si ukážeme, že výše zmíněný příklad s kolegou a jeho makrofyty není úplně nereálný.

Jak bylo řečeno výše, průměrné EIH obsahují dvě informační složky, jednu odvozenou z externí informace o ekologickém chování jednotlivých druhů (vlastní Ellenbergovy indikační hodnoty) a druhou odvozenou z druhového složení jednotlivých zápisů, ze kterých jsou hodnoty počítány. Obě tyto složky je přitom možné od sebe oddělit tak, že před vlastním výpočtem průměrných EIH promícháme druhové Ellenbergovy indikační hodnoty (v rámci daného faktoru) mezi druhy v tabulce s vegetačními zápisy (obr. 2b), čímž znehodnotíme ekologickou informaci o jejich chování (případně můžeme, s podobným efektem, vlastní druhové indikační hodnoty nahradit náhodnými čísly). Takto promíchané druhové hodnoty (nebo náhodná čísla přiřazená jednotlivým druhům) použijeme k výpočtu průměrných EIH pro jednotlivé snímky, stejně jako bychom to udělali s původními indikačními hodnotami druhů. Výsledná hodnota v sobě neobsahuje smysluplnou externí ekologickou informaci, zůstává v ní ale obsažena informace o druhovém složení, protože je z druhového složení vypočtena. Pro použití v dalším textu budeme této proměnné říkat *průměr znáhodněných EIH*. Nakonec si zavedeme ještě jednu proměnnou, která půjde ještě více do extrému – nejen že nebude obsahovat žádnou smysluplnou externí ekologickou informaci, nebude ale ani ovlivněna druhovým složením jednotlivých zápisů. Získáme ji tak, že promícháme *průměr znáhodněných EIH* mezi jednotlivými snímky v našem datovém souboru (obr. 2c). Podobného účinku bychom dosáhli, kdybychom jednoduše ke každému vegetačnímu zápisu vygenerovali náhodné číslo, a této proměnné proto budeme říkat *náhodné číslo*. Vedle *průměrné EIH* pro daný faktor (např. půdní reakci) tedy máme další dvě proměnné, *průměr znáhodněných EIH* pro půdní reakci a *náhodné číslo*. Co se stane, když použijeme *průměr znáhodněných EIH* pro jednotlivé faktory, vypočteme jeho korelaci s ordinačními osami a otestujeme průkaznost této korelace? A co se stane, když uděláme to samé s *náhodným číslem*?

<Obr. 2 umístit někam sem>

Aby nedošlo k nedorozumění – nejde o promíchání druhových EIH mezi druhy v rámci *jednoho* vegetačního zápisu, protože výsledná průměrná EIH by byla samozřejmě stejná jako při použití hodnot nepromíchaných (pokud tedy počítáme průměry neválené pokryvností druhů). Indikační hodnoty jsou míchány mezi druhy v rámci celé tabulky vegetačních zápisů použitých v analýze (ta se většinou uvádí ve formátu druhy × vzorky, kdy druhy tvoří řádky tabulky a vzorky (zápisy) tvoří sloupce). Výsledný *průměr znáhodněných EIH* pro jeden vegetační záznam nemá coby absolutní hodnota žádný význam, ten získává až použitím v analýze s vegetačními daty. Při každém promíchání (randomizaci) indikačních hodnot mezi druhy a výpočtu průměru vyjde jiné číslo, a pro získání představy, jak se tato hodnota v analýze chová, je třeba takto znáhodněných hodnot vygenerovat velké množství. To samé platí pro *náhodné číslo*. V následující analýze použijeme *průměr znáhodněných EIH* pro půdní reakci a otestujeme jeho vztah k prvním dvěma osám DCA, a to stejným způsobem, jako jsme to v předchozí analýze provedli s *průměrnými EIH* pro jednotlivé faktory. Tuto analýzu zopakujeme 1000 krát, pokaždé s průměrem vypočteným z nově znáhodněných indikačních hodnot. Podobně zopakujeme 1000 krát analýzu vztahu mezi ordinačními osami a *náhodným číslem*, pokaždé s nově vygenerovanými hodnotami *náhodného čísla*. Pro obě analýzy zaznamenáme počet testů průkazných na hladině průkaznosti $p < 0,05$.

Z tisíce testů průkaznosti korelace mezi *průměrem znáhodněných EIH* pro půdní reakci a prvními dvěma ordinačními osami vyšlo 684 testů průkazných na hladině signifikance $p < 0,05$. Ze stejného počtu testů pro *náhodné číslo* jich vyšlo průkazných 58, což téměř přesně odpovídá použité 5% hladině signifikance. *Průměr znáhodněných EIH* je přitom proměnná, která pro nás z hlediska interpretace nemá žádný význam, protože indikační hodnoty, ze kterých je vypočtena, jsou znáhodněné a tím pádem nesmyslné. Skutečnost, že i přesto v 68 procentech případů vychází taková proměnná průkazně, je varující. Velmi snadno se totiž může stát, že ordinační osy interpretujeme pomocí proměnné, která ve skutečnosti nenese smysluplnou ekologickou informaci.

Klasicky používaný test průkaznosti má tedy tendenci vycházet daleko optimističtěji (rozuměj častěji průkazný) než jak by odpovídalo skutečné informační hodnotě použitých dat. Důvod je přitom jednoduchý – průměrné EIH a ordinační osy na sobě nejsou nezávislé, a jejich vztah tedy nelze tradičním způsobem testovat. Co s tím? Pokud to není třeba, je lepší test průkaznosti v takových případech vůbec neprovádět. Pokud test provést potřebujeme, třeba proto, že to recenzenti vyžadují, nebo abychom rozhodli, které z šesti průměrných EIH použít pro interpretaci našich výsledků a které ne, je možné použít jednu z následujících alternativ.

První alternativou je nepoužívat korelaci průměrných EIH se skóre vegetačních zápisů na ordinačních osách, ale korelovat přímo druhové indikační hodnoty se skóre jednotlivých druhů na ordinačních osách a tuto korelaci testovat. Tímto způsobem, který doporučují Herben & Münzbergová (2003), obejdeme elegantně problém se závislostí mezi průměrnými EIH a ordinačními osami. Vlastně je s podivem, jak málo se tato metoda používá. Souvisí to zřejmě s tím, že v porovnání s průměrnými EIH vychází jaksi „hůř", tedy méně často průkazně (tab. 2, obr. 3), a při honbě za průkaznými výsledky se tak stává méně atraktivní. Jak bylo ale řečeno výše, problém ve skutečnosti není v tom, že by korelace druhových indikačních hodnot se skóre druhů vycházela „hůř", ale v tom, že korelace průměrných EIH se skóre zápisů vychází „příliš dobře". Určitou nevýhodou této metody je ale fakt, že korelace se vztahuje ke skóre druhů a ne zápisů, takže interpretace s ohledem na gradienty v druhovém složení dat není tak přímočará. Rozdíl je také v tom, že řada druhů nemá pro některou z EIH přiřazenu žádnou hodnotu, a tyto druhy z vlastní korelace vypadnou – počty druhů s chybějícími EIH se přitom liší mezi jednotlivými faktory, a korelace (potažmo její průkaznost) je tak pro jednotlivé faktory počítána z různého počtu hodnot.

<Tab. 2 a Obr. 3 umístit někam sem>

Druhou alternativou při testování průkaznosti korelace průměrných EIH se skóre vzorků je použití modifikovaného permutačního testu (Zelený & Schaffers 2012), který je založen na výše vysvětleném promíchávání druhových indikačních hodnot mezi druhy v tabulce dat. Korelace *průměrných EIH* je přitom porovnávána s opakovaně generovanými korelacemi *průměrů znáhodněných EIH* s cílem zjistit, jestli jsou *průměrné EIH* průkazně (tedy dostatečně často) lepší. Další podrobnosti k tomuto testu včetně způsobu, jak a v čem jej spočítat, najdete v příloze 1. Při porovnání s elegantní jednoduchostí první alternativy se může modifikovaný permutační test zdát zbytečně složitý, jeho výhodou ale je, že na stejném principu lze postavit testy i v dalších typech analýz, nejen v nepřímé ordinaci. Výsledky modifikovaného permutačního testu (tab. 1., P_{modif}) vycházejí podobně jako výsledky korelace druhových indikačních hodnot se skóre druhů na ordinačních osách (tab. 2). Pouze živiny a vlhkost zůstaly vysoce průkazné ($p < 0,001$), průkaznost teploty a světla se snížila,

kontinentalita se stala neprůkaznou a půdní reakce slabě průkaznou. V případě korelace druhových EIH se skóre druhů na ordinačních osách (tab. 2) se půdní reakce stala také neprůkaznou.

Testování rozdílů mezi vegetačními typy vyjádřených průměrnými Ellenbergovými indikačními hodnotami

Další často používanou analýzou, ve které jsou průměrné Ellenbergovy indikační hodnoty přímo použity při interpretaci vegetačních dat, je srovnání ekologických nároků jednotlivých vegetačních typů. Vegetační typ je tvořen vegetačními zápisy, které jsou si druhovým složením podobné, a různé vegetační typy se od sebe druhovým složením více či méně odlišují. Častým způsobem, jak bývají vegetační typy s platností pro studované území definovány, je použití některé z metod numerické klasifikace jako TWINSPAN (Hill 1979, Roleček et al. 2009) nebo shluková analýza, které rozdělí použitý soubor vegetačních zápisů do předem definovaného počtu skupin. Následně mohou být porovnány průměrné EIH zápisů v jednotlivých skupinách a tyto rozdíly bývají často i statisticky testovány, buď analýzou variance (ANOVA) nebo jejími neparametrickými analogiemi (např. Kruskal-Wallisův test). Výsledkem je zjištění, které ekologické faktory nejlépe popisují rozdíly mezi vegetačními typy. Jak se v hodnotách vybraného faktoru jednotlivé vegetační typy liší, může být následně zjištěno *post-hoc* srovnávacím testem (např. Tukey HSD).

Problém těchto testů je obdobný jako v případě korelace mezi průměrnými EIH a ordinačními osami. Příslušnost vegetačního záznamu do jednoho z vegetačních typů je kategoriální proměnná odvozená z vegetačních dat, buď přímo pomocí numerické klasifikace, nebo i nepřímo pomocí klasifikace manuální, používané v tradiční fytocenologii (při které se snažíme ručně roztřídit zápisy do skupin, které jsou z hlediska druhového složení co nejvíce homogenní). Pokud tedy vztah mezi průměrnými EIH a příslušností zápisu do vegetačního typu statisticky otestuji, mohu opět očekávat, že výsledky budou zkreslené a příliš optimistické.

Ukážeme si to stejně jako v předchozím případě na vegetačních datech z údolí Vltavy, která jsou pomocí shlukové analýzy rozdělena na 5 vegetačních typů. Pro shlukovou analýzu byla použita kombinace Euklidovské distance pro výpočet nepodobnosti mezi zápisy a Wardova shlukovacího algoritmu, který tvoří dobře vymezené shluky. Shluková analýza byla provedena v programu R pomocí příkazu *hclust* a výsledek byl nakreslen ve formě dendrogramu s vylišením pěti vegetačních typů (obr. 4). Pro jednotlivé zápisy byly vypočteny průměrné EIH pro všech šest faktorů a rozdíly mezi skupinami byly zobrazeny pro každý faktor zvlášť pomocí krabicových diagramů. Pro testování průkaznosti rozdílů mezi skupinami byla použita jednocestná ANOVA, průkaznost byla testována F-testem. Podobně jako v případě DCA byl proveden také modifikovaný permutační test (viz Příloha 1), který zohledňuje skutečnost, že testované proměnné na sobě nejsou nezávislé (Zelený & Schaffers 2012).

<Obr. 4 a Tab. 3 vložit někam sem>

Očekávané optimistické výsledky nás rozhodně nezklamaly (tab. 3, P_{param}), protože pro všech šest průměrných EIH vyšly vysoce ($p < 0,001$) průkazné rozdíly mezi jednotlivými vegetačními typy (obr. 5). Mohl bych se tedy pustit do interpretací, které by byly do značné míry podobné interpretacím použitým u nepřímé ordinace. Pokud ale provedu i modifikovaný permutační test, situace se výrazně změní (tab. 3, P_{modif}). Jak kontinentalita, tak i teplota se staly neprůkaznými, a průkaznost půdní reakce se také výrazně snížila. Může se to zdát jako škoda, vždyť při pohledu na krabicový diagram pro kontinentalitu (obr. 5) je vidět, že nejvíce je kontinentální vegetace teplomilných doubrav (typ 3) a acidofilních borů a doubrav (typ 1), což pěkně souhlasí s teorií, kterou jsem rozvíjel u výsledků nepřímé ordinační analýzy. Pokud se ale rozhodnu, že se při interpretacích budu řídit výsledky testů průkaznosti, měl bych použít výsledky modifikovaného permutačního testu.

<Obr. 5 vložit někam sem>

Další typy analýz

Problém s používáním průměrných EIH je ve skutečnosti daleko širší, než by se ze dvou výše uvedených příkladů mohlo zdát. Při jejich použití jako pasivních proměnných v nepřímé

ordinaci se musíme vedle zkreslených výsledků testů průkaznosti vypořádat ještě s faktem, že v porovnání s měřenými faktory prostředí mohou analogické průměrné EIH korelovat s ordinačními osami lépe a jejich vektory v ordinačních diagramech tak budou delší, ačkoliv to nemusí odpovídat jejich skutečné informační hodnotě. Pokud se tedy pozastavujete třeba nad tím, že průměrná EIH pro půdní reakci je ve vašich datech lépe korelovaná s ordinačními osami než měřené pH, může to být způsobeno právě vnitřní provázaností mezi průměrnými EIH a ordinačními osami, o které byla řeč výše. Ze stejných důvodů mají navíc průměrné EIH tendenci být více korelované s ordinačními osami vyššího řádu, tedy více s první než s druhou osou. První ordinační osa totiž obsahuje více informace extrahované z druhového složení než osa druhá (proto je také první, protože vysvětlí více variability v druhových datech než všechny ostatní, podobně druhá vysvětlí více než třetí atd.), a její korelace s průměrnými EIH bude proto těsnější.

Mimo výše uvedené typy analýz se může problém s optimistickými výsledky testů průkaznosti projevit také při testování korelace mezi průměrnými EIH navzájem, případně při testování korelace (nebo regrese) průměrných EIH a druhové bohatosti, spočítané ze stejných dat (i když zde není zkreslení tak výrazné a záleží na heterogenitě dat; Zelený & Schaffers 2012). Dále se může projevit při použití průměrných EIH jako vysvětlujících proměnných v klasifikačních a regresních stromech (CART; De'ath & Fabricius 2000), pokud je závisle proměnná odvozena přímo z druhových dat (např. počty druhů v zápisech v případě regresních stromů nebo jejich roztřídění do vegetačních typů v případě stromů klasifikačních). V tom případě může být variabilita vysvětlená průměrnými EIH nadhodnocená, což se pak projeví při použití společně s měřenými faktory prostředí, které se jeví jako horší.

Může se samozřejmě stát, že průměrné EIH jsou skutečně lepšími proměnnými prostředí než měřené faktory. Měřené faktory nemusí vždy měřit to, co je pro rostliny skutečně důležité. Při stanovení obsahu živin v půdě se například může stát, že chemický rozbor půdy sice odhalí koncentraci nějakého prvku v půdě, my ale nevíme, jestli je tento prvek v dané formě rostlinou skutečně využíván a v jaké míře (např. Axmanová et al. 2011). Stanovištní podmínky navíc mohou výrazně kolísat v průběhu roku, a jednorázově měřená hodnota (například vlhkosti) nemůže toto kolísání zachytit. O Ellenbergových indikačních hodnotách

se naopak předpokládá, že zachycují jakýsi průměrný pohled na stanovištní podmínky z hlediska zastoupených druhů.

Shrnutí

Na Ellenbergovy indikační hodnoty rozhodně není třeba zanevřít, protože jejich použití nám může výrazně usnadnit ekologickou interpretaci jevů, se kterými se při studiu vegetace setkáváme. Při použití průměrných EIH je ale nutné mít na paměti, že tyto proměnné jsou vypočteny z druhových dat, a jejich použití v analýze spolu s jinými proměnnými vypočtenými z těchto dat může vést k závěrům, které jsou optimističtější, než by ve skutečnosti měly být.

Nejjednodušším doporučením je průměrné EIH nepoužívat vůbec. Například pokud jsou k dispozici relevantní měřené faktory prostředí, není třeba používat zároveň i průměrné EIH jen proto, že je tak snadné je vypočíst. Pokud průměrné EIH přeci jen použít chceme, neměli bychom testovat průkaznost jejich vztahu s jinými proměnnými, které jsou odvozeny ze stejných druhových dat. A pokud se přeci jen rozhodneme test průkaznosti použít, měli bychom použít jeho modifikovanou podobu, která bere v potaz skutečnost, že testované proměnné na sobě nejsou nezávislé. Zároveň platí, že průměrné EIH by v takových analýzách neměly být bez dalšího statistického ošetření srovnávány s analogickými měřenými faktory prostředí, protože se oproti nim mohou neoprávněně jevit lepšími, než ve skutečnosti jsou (například tím, že jsou lépe korelované nebo častěji a více průkazné). Týká se to například testování vztahu mezi průměrnými EIH a ordinačními osami v nepřímé ordinační analýze (korelace nebo regrese s ordinačními osami PCA, CA, DCA, NMDS aj.), testování rozdílů v průměrných EIH mezi vegetačními typy (ANOVA nebo Kruskal-Wallisův test), testování vztahu průměrných EIH a druhové bohatosti (korelace, regrese) nebo testování vztahů mezi průměrnými EIH navzájem (korelace).

Poděkování

Rád bych poděkoval mým kolegům za jejich názory a připomínky, které přispěly k hlubšímu pochopení celé problematiky používání Ellenbergových indikačních hodnot ve vegetačních

analýzách; jmenovitě jsou to Martin Diekmann, Jörg Ewald, Milan Chytrý, Ching-Feng Li, Jan Roleček, André P. Schaffers a Lubomír Tichý. Dík patří také anonymnímu recenzentovi, Tomášovi Černému a ostatním členům redakční rady Zpráv ČBS za jejich připomínky k rukopisu tohoto článku. Vznik textu byl podpořen prostředky Grantové agentury České republiky (GAP505/11/0732 a GAP505/12/1022).

Summary

The use of Ellenberg indicator values (EIVs) in vegetation analysis has long lasting tradition, at least in European countries. Although criticized for multiple issues, mean EIVs are generally considered to represent a reliable estimate of habitat ecological characteristics, mainly if these are not directly measured. In this paper, I pointed out several issues related to the use of mean EIVs in analyses with other data derived from the same dataset, which is often seen in vegetation studies. Such analyses include correlation of mean EIVs and ordination axes in an unconstrained ordination analysis, ANOVA of mean EIVs between vegetation types, correlation of mean EIVs and species richness or correlation of mean EIVs between each other. Due to the way how mean EIVs are calculated, these values inherit the information about the species composition of the plot from which it is calculated. If related to other variables derived from the same species composition data, attention must be paid to the obvious fact that these two sets of variables are not independent. As a result, the statistical analysis of this relationship, especially if tested for significance, often yields biased results toward more optimistic values (e.g. higher correlation coefficients, higher coefficients of determination and highly significant statistical tests). As such results are “too good to be true”, the interpretation based on them may be misleading. I propose several recommendations for using mean EIVs in vegetation analysis together with other variables derived from species composition. First, if there are relevant environmental factors which are directly measured, don’t use simultaneously mean EIVs just because it is so simple to calculate them. If you decide to use them, don’t test the significance of their relationship with other variables derived from the same species composition data. And if you decide to test these relationships anyway, use modified permutation tests, which takes into consideration the dependency of both sets of variables. There is no need to abandon the

using of mean EIVs in vegetation analysis, as they often bring useful insights into the structure and ecological interpretation of the data, especially in descriptive vegetation studies. However, their precision and importance should not be overrated.

Literatura

Axmanová I., Zelený D., Li C.F. & Chytrý M. (2011): Environmental factors influencing herb layer productivity in Central European oak forests: insights from soil and biomass analyses and a phytometer experiment. – Plant Soil 342: 183–194.

Borhidi A. (1995): Social behaviour types, the naturalness and relative ecological indicator values of the higher plants in the Hungarian Flora. – Acta Bot. Hung. 39: 97–181.

Cajander A.K. (1926): The theory of forest types. – Acta For. Fenn. 2: 11–108.

Chase J.M. (2003): Community assembly: when should history matter? – Oecologia 136: 489–498.

De'ath G. & Fabricius K.E. (2000): Classification and regression trees: powerful yet simple technique for ecological data analysis. – Ecology 81: 3178–3192.

Diekmann M. (2003): Species indicator values as an important tool in applied plant ecology – a review. – Basic Appl. Ecol. 4: 493–506.

Ellenberg H. (1974): Zeigerwerte der Gefäßpflanzen Mitteleuropas. – Scr. Geobot. 9: 1–97.

Ellenberg H., Weber H.E., Düll R., Wirth V., Werner W. & Paulißen D. (1992): Zeigerwerte von Pflanzen in Mitteleuropa. – Scr. Geobot. 18: 1–258.

Herben T. & Münzbergová Z. (2003): Zpracování geobotanických dat v příkladech. Část 1. Data o druhovém složení. – Ms., skripta, URL: ftp://botany.natur.cuni.cz/skripta/zpracovani_geobot_dat/multivar.pdf.

Hill M.O. (1979): TWINSPAN – A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. – Ecology and Systematics, Cornel University, Ithaca, NY, US.

Hill M.O. & Gauch J.G. (1980): Detrended correspondence analysis: an improved ordination technique. – Vegetatio 42: 47–58.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H. & Wagner H. (2011): Vegan: Community Ecology Package. R package version 1.17-6. URL: http://cran.r-project.org/web/packages/vegan/index.html.

Persson S. (1981): Indicator values as an aid in the interpretations of ordination diagrams. – J. Ecol. 69: 71–84.

Roleček J., Tichý L., Zelený D. & Chytrý M. (2009): Modified TWINSPAN classification in which the hierarchy respects cluster heterogeneity. – J. Veg. Sci. 20:596–602

Schaffers A.P. & Sýkora K.V. (2000): Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: comparison with field measurements. – J. Veg. Sci. 11: 225–244.

ter Braak C.J.F. & Šmilauer P. (2002): CANOCO reference manual and CanoDraw for Windows user's guide. Software for Canonical Community Ordination (version 4.5.). – Biometris, Wageningen & České Budějovice.

Tichý L. (2002): JUICE, software for vegetation classification. – J. Veg. Sci. 13: 451–453.

Whittaker R.H. (1956): Vegetation of the Great Smoky Mountains. – Ecol. Monogr. 26: 2–80.

Zelený D. (2008): Lesní vegetace v údolí Vltavy severně od Zlaté Koruny (okres Český Krumlov) – Zprávy Čes. Bot. Společ. 43: 111–169.

Zelený D. & Chytrý M. (2007): Environmental control of the vegetation pattern in deep river valleys of the Bohemian Massif. – Preslia 205–222.

Zelený D. & Schaffers A.P. (2012): Too good to be true: pitfalls of using mean Ellenberg indicator values in vegetation analyses. – J. Veg. Sci., Doi: 10.1111/j.1654-1103.2011.01366.x.

Příloha 1 – Princip modifikovaného permutačního testu a jeho praktické použití.

Appendix 1 – Theory on the background of modified permutation test and it's practical use.

Teoretický princip modifikovaného permutačního testu

Principem permutačních testů je, že randomizují (promíchávají) hodnoty jednotlivých pozorování a na základě analýzy takto randomizovaných dat dělají závěry o nulové hypotéze. Tomuto typu analýz se říká Monte Carlo a jejím příkladem je permutační test pro testování průkaznosti vysvětlující proměnné v přímé ordinační analýze (např. v programu CANOCO for Windows; ter Braak & Šmilauer 2002). Zatímco parametrické testy vyžadují, aby testované proměnné vykazovaly určité (např. normální) rozdělení, permutační testy tuto podmínku nemají a jejich použití je proto daleko širší. Permutační testy simulují stav, kdy platí nulová hypotéza, totiž že vztah mezi proměnnými se neliší od vztahu, který by vznikl náhodným přiřazením hodnot jednotlivých pozorování. Následně pak zjišťují, jestli se reálná (nepermutovaná) data od tohoto vztahu dostatečně liší. V prvním kroku je třeba vybrat testovou statistiku, jejíž průkaznost se bude testovat (pro korelaci to může být například korelační koeficient *r*, pro regresi koeficient determinace R^2, pro analýzu variance hodnota *F*). Ve druhém kroku je vytvořena distribuce hodnot testové statistiky, které mohou být očekávány, pokud nulová hypotéza platí. Každá jedna hodnota této testové statistiky je získána právě permutací pozorování jedné proměnné a její analýzou s druhou proměnnou (v případě korelace tedy vypočteme korelační koeficient *r* mezi dvěma proměnnými, z nichž jedna má mezi sebou promíchaná pozorování), a distribuci těchto hodnot získáme mnohonásobným opakováním tohoto kroku. Takto vygenerovanou distribuci hodnot testové statistiky nakonec srovnáme s hodnotou reálné testové statistiky, kterou vypočteme z původních (nepermutovaných) dat. Úkolem je rozhodnout, jaká je pravděpodobnost, že by hodnota reálné testové statistiky vznikla z náhodně přiřazených dat. Pokud je tato pravděpodobnost dostatečně malá (např. tradičně používaných 5 %, 1 % nebo 0,1 %), můžeme nulovou hypotézu zamítnout a tvrdit, že vztah mezi studovanými proměnnými se průkazně liší od náhodného. Hladina průkaznosti (např. 5 %) je pravděpodobnost, se kterou jsme ochotni riskovat, že zamítneme nulovou hypotézu, která ve skutečnosti platí (tzv. chyba prvního druhu).

Modifikaci tohoto permutačního testu pro průměrné EIH si ukážeme na příkladu korelace průměrné EIH pro půdní reakci s první osou DCA (přesněji řečeno se skóre jednotlivých pozorování na této ose). Máme tedy dvě proměnné, průměrnou EIH pro půdní reakci a skóre pozorování na první ose, a vypočteme jejich korelační koeficient *r.* Pokud bychom použili klasické schéma permutačního testu na otestování průkaznosti tohoto korelačního koeficientu, pak bychom skutečnou hodnotu *r* porovnávali s nulovou distribucí hodnot *r*, které vzniknou randomizací jedné proměnné (např. průměrných EIH) a jejich korelací s druhou proměnnou. Modifikace tohoto testu je založená na úpravě nulové hypotézy, kterou chceme zamítnout. Původní nulová hypotéza zněla takto: „*Mezi proměnnou X (skóre zápisů na ordinačních osách) a průměrnou EIH není žádný vztah*“. Jak je ale vysvětleno v hlavním textu, z principu výpočtu obou proměnných vyplývá, že na sobě obě proměnné nejsou nezávislé, a testovat takto znějící nulovou hypotézu proto postrádá smysl. Nulovou hypotézu proto upravíme do následující podoby: „*Proměnná X nemá vztah k té části informace v průměrné EIH, která je odvozená z ekologických vlastností jednotlivých druhů*“. Takto definovaná nulová hypotéza tedy bere v potaz fakt, že mezi proměnnou X a tou částí informace v průměrných EIH, která je díky způsobu výpočtu odvozena

z druhových dat, vztah z principu je a jeho testování pro nás tedy není zajímavé. Permutační schéma původního testu je upraveno tak, že místo permutování hodnot průměrné EIH mezi jednotlivými pozorováními jsou permutovány přímo druhové EIH mezi druhy ve vegetační tabulce, ze kterých je následně spočten *průměr znáhodněných EIH* pro každý vegetační zápis a z těchto hodnot je následně spočtena hodnota statistického kritéria (v tomto případě korelační koeficient). Proces permutace se tedy „vnořil hlouběji“ do algoritmu výpočtu průměrných EIH, a místo původní permutace už vypočtených průměrných EIH jsou permutovány EIH pro jednotlivé druhy, které jsou poté použity k výpočtu průměrů pro jednotlivé zápisy. V případě analýzy variance (ANOVA) je test založen na stejném principu, pouze místo korelačního koeficientu *r* je jako testová statistika použita hodnota *F*.

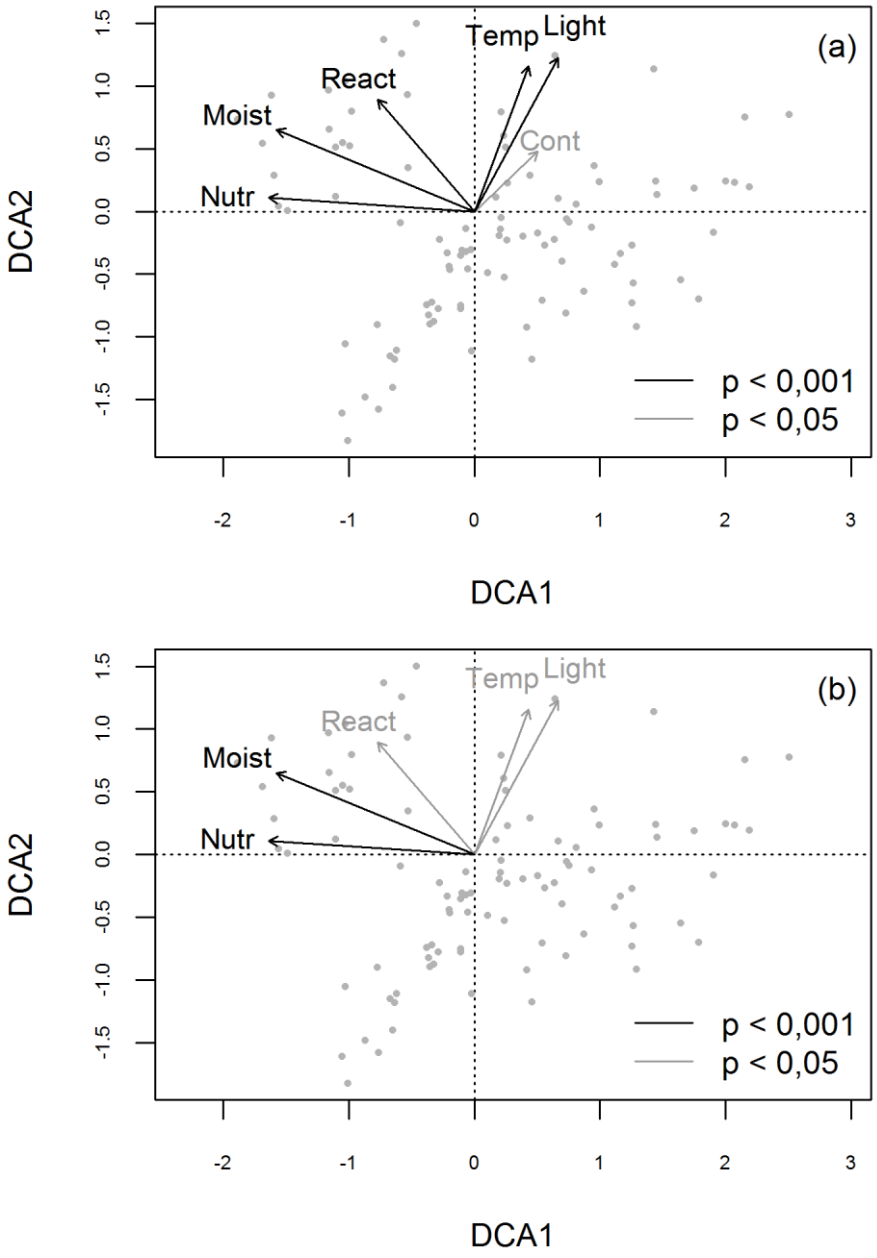
Praktické použití modifikovaného permutačního testu

Modifikovaný permutační test pro oba typy analýz zmiňované v tomto článku je možné provést v programu R pomocí funkcí *envfit.iv* a *summary.aov.iv*, které jsou k dispozici v přílohách článku Zelený & Schaffers (2012). Funkce *envfit.iv* slouží k otestování vztahu mezi ordinačními osami a průměrnými EIH (vztah je vyjádřen vícenásobnou lineární regresí, kde průměrná EIH vystupuje jako závisle proměnná a skóre zápisů na ordinačních osách jako vysvětlující proměnné) a v kombinaci s funkcí *plot.envfit* umožňuje jejich promítnutí do ordinačního diagramu. Funkce *summary.aov.iv* počítá modifikovaný permutační test pro analýzu variance mezi průměrnými EIH pro jednotlivé vegetační typy.

Oba testy je také možné provést pomocí výše uvedených funkcí v programu JUICE 7.0 (Tichý 2002) v kombinaci s JUICE-R skripty, které jsou pro tyto metody dostupné na internetové adrese http://bit.ly/juice-r (v sekci *Available R scripts* pod názvem *Modified permutation test for Ellenberg values*). JUICE-R skript je určen pro program R, který je třeba mít také nainstalován na počítači, uživatel jej ale nemusí umět ovládat. Příslušný JUICE-R skript je nutné stáhnout z výše uvedených internetových stránek a uložit v počítači. Vegetační data je třeba naimportovat do programu JUICE, ve kterém je možné k jednotlivým druhům přiřadit jejich Ellenbergovy indikační hodnoty. Z programu JUICE jsou následně tato data spolu s JUICE-R skriptem zaslána do programu R, kde dojde k vlastnímu výpočtu. Podrobnosti k použití této metody v programu JUICE najdete na webových stránkách programu JUICE (http://juice.trenck.cz/?idm=7) nebo přímo na výše uvedených JUICE-R stránkách. Skript je možné spustit také samostatně v programu R a data do něj nahrát jako tabulky uložené v textových souborech ve formátu *.txt; program JUICE v tom případě není třeba (více podrobností naleznete na http://bit.ly/eiv-modif-perm-test).

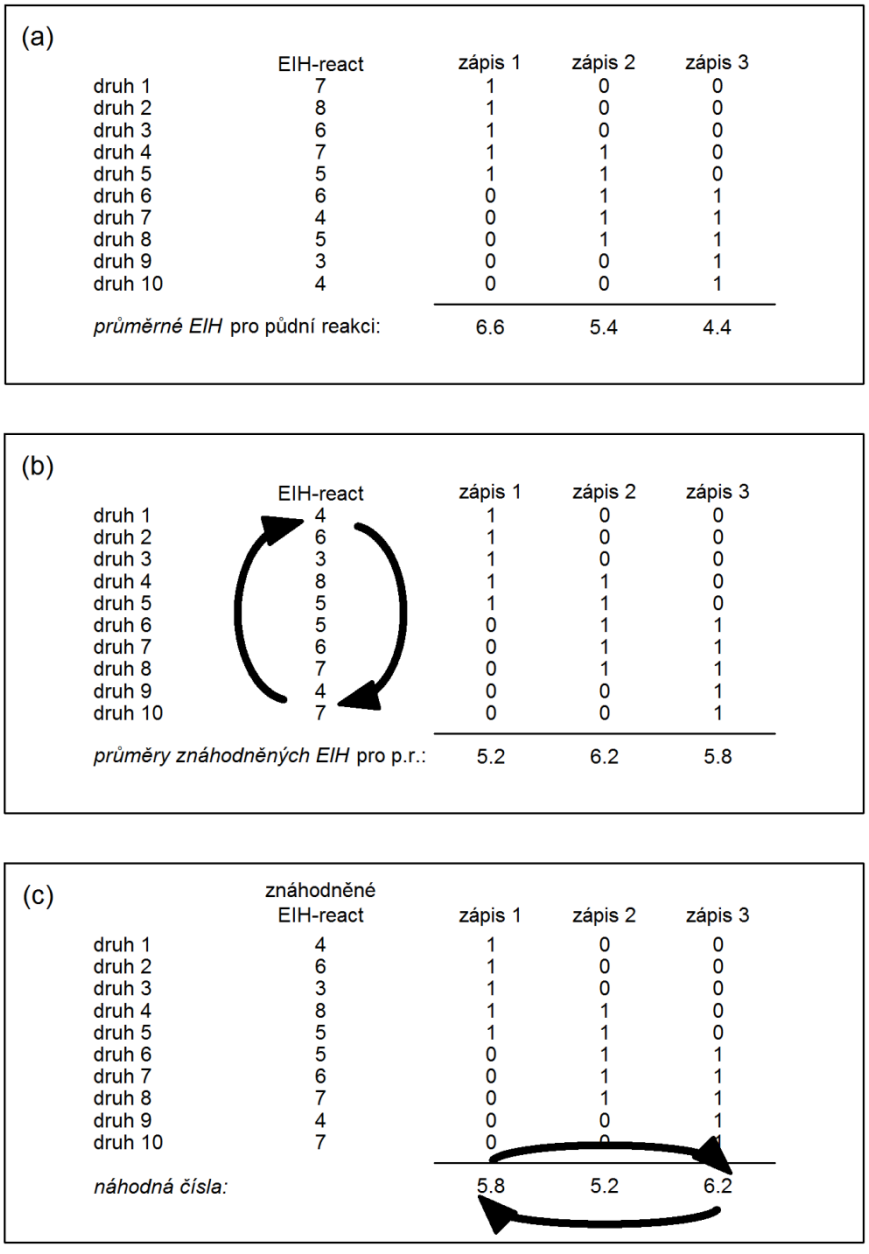
Obr. 1. – Ordinační diagram detrendované korespondenční analýzy (DCA) s pasivně proloženými průměrnými Ellenbergovými indikačními hodnotami pro světlo (Light), teplotu (Temp), kontinentalitu (Cont), vlhkost (Moist), živiny (Nutr) a půdní reakci (React). Míra průkaznosti vztahu průměrných EIH k ordinačním osám je vyjádřena následovně: černé vektory zobrazují proměnné průkazné na hladině $p < 0,001$, šedé na $p < 0,05$, neprůkazné proměnné nejsou zobrazeny. Pro hodnoty průkaznosti viz tab. 1. (a) Průkaznost počítána původním permutačním testem, (b) průkaznost počítána modifikovaným permutačním testem. Body v diagramu zobrazují pozice jednotlivých vegetačních zápisů.

Fig. 1. – Ordination diagram for detrended correspondence analysis (DCA) with passively projected mean Ellenberg indicator values for light (Light), temperature (Temp), continentality (Cont), moisture (Moist), nutrients (Nutr) and soil reaction (React). Significance of the relationship between the mean EIV and ordination axes is expressed as following: black vectors are for variables significant at the level $p < 0,001$, grey vectors at $p < 0,05$, not significant variables are not displayed. See Table 1 for significance values. (a) Significance calculated using the original permutation test, (b) significance calculated by modified permutation test. Points in the diagram represent position of individual vegetation samples.


(a)
Temp
Light
React
Moist
Cont
Nutr
p < 0,001
p < 0,05
DCA2
DCA1
(b)
Temp
Light
React
Moist
Nutr
p < 0,001
p < 0,05
DCA2
DCA1

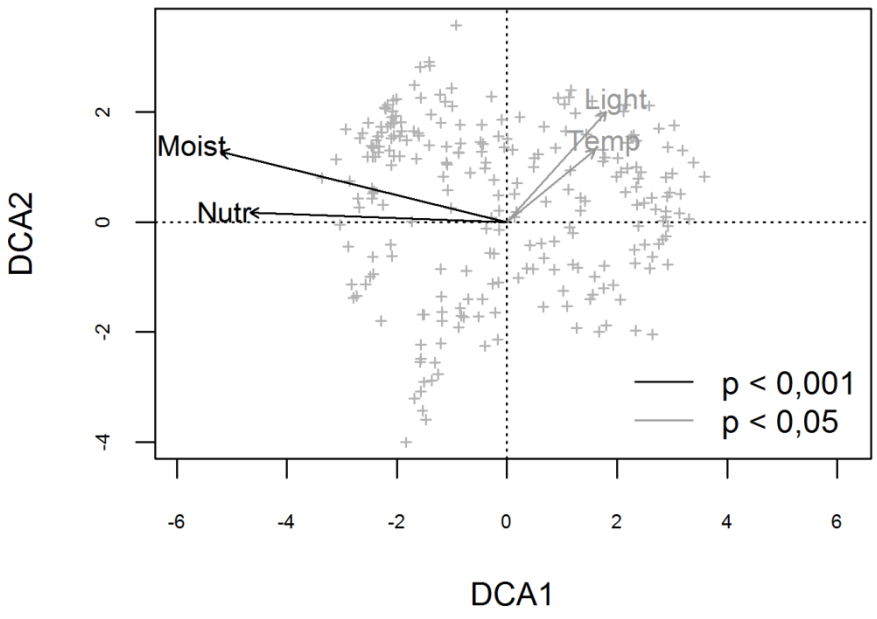
Obr. 2. – Schematické znázornění výpočtu (a) *průměrných Ellenbergových indikačních hodnot* pro půdní reakci, (b) *průměrů znáhodněných EIH* pro půdní reakci a (c) *náhodných čísel*. Pro podrobnosti výpočtu viz hlavní text.

Fig. 2. – Schematic representation of the calculation of (a) *mean real Ellenberg indicator values* for soil reaction, (b) *mean randomized EIV* for soil reaction and (c) *random numbers*. See the main text for details.



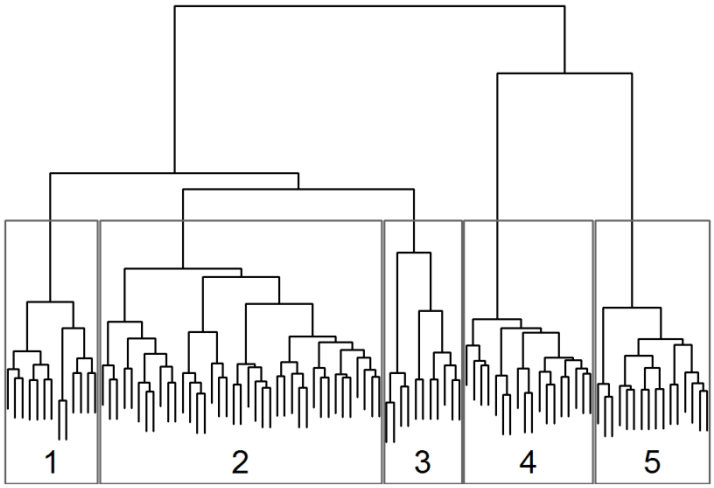
Obr. 3. – Ordinační diagram detrendované korespondenční analýzy (DCA) s pasivně promítnutými druhovými Ellenbergovými indikačními hodnotami (zkratky jako v obr. 1). Na rozdíl od obr. 1 nejde o vztah mezi průměrnými EIH a skóre jednotlivých vegetačních zápisů na ordinačních osách, ale o vztah mezi EIH pro druhy a skóre těchto druhů. Míra průkaznosti tohoto vztahu je vyjádřena následovně: černé vektory zobrazují proměnné průkazné na hladině $p < 0,001$, šedé na hladině $p < 0,05$, neprůkazné proměnné nejsou zobrazeny. Pro hodnoty průkaznosti viz tab. 2. Křížky v ordinačním diagramu zobrazují pozice jednotlivých druhů.

Fig. 3. – Ordination diagram for detrended correspondence analysis (DCA) with passively projected species Ellenberg indicator values (abbreviations the same as in Fig. 1). In contrast to Fig. 1 the vectors do not represent relationships between mean EIV and sample scores on ordination axes, but relationship between EIV for individual species and species scores. Significance of this relationship is expressed as following: black vectors are for variables significant at the level $p < 0,001$, grey at $p < 0,05$, not significant variables are not displayed. See Tab. 2 for significance values. Grey crosses in diagram represent position of individual species.



Obr. 4. – Dendrogram shlukové analýzy, která rozdělila vegetační zápisy z údolí Vltavy do pěti vegetačních typů, ohraničených šedými obdélníky s číslem vegetačního typu. Čísla vegetačních typů jsou vysvětlena v popisku k obr. 5.

Fig. 4. – Dendrogram of cluster analysis, dividing vegetation samples from Vltava river valley into five vegetation types, which are separated by grey rectangles with number of the vegetation type inside. The numbering of vegetation types is explained in captions of Fig. 5.



Obr. 5. – Krabicové diagramy, které zobrazují rozdíly průměrných Ellenbergových indikačních hodnot mezi jednotlivými vegetačními typy. Průkaznost rozdílů, testovaná parametrickým i modifikovaným permutačním testem, je uvedena v tab. 3. Vegetační typy: 1 – acidofilní bory a doubravy, 2 – suťové lesy a dubohabřiny, 3 – teplomilné doubravy, 4 – lužní lesy, 5 – jedliny.

Fig. 5. – Box-and-whisker diagrams, representing differences of mean Ellenberg indicator values between particular vegetation types. For the significance of the differences, tested by both parametric and modified permutation test, see Tab. 3. Vegetation types: 1 – acidophilous pine- and oak forests, 2 – ravine forests and oak-hornbeam forests, 3 – thermophilous oak forests, 4 – alluvial forests, 5 – fir forests.

Světlo / Light
Teplota / Temperature
Kontinentalita / Continentality
Vlhkost / Moisture
Živiny / Nutrients
Půdní reakce / Soil reaction

Tab. 1. – Srovnání průkaznosti korelace mezi průměrnými Ellenbergovými indikačními hodnotami a osami DCA, počítané původním permutačním testem (P_{orig}) a modifikovaným permutačním testem (P_{modif}).

Tab. 1. – Comparison of the significance of the relationship between mean Ellenberg indicator values and DCA ordination axes, calculated by both original permutation test (P_{orig}) and modified permutation test (P_{modif}).

	DCA1	DCA2	R^2	P_{orig}	P_{modif}
Světlo / Light	0,477	0,879	0,600	**< 0,001**	**0,004**
Teplota / Temperature	0,350	0,937	0,471	**< 0,001**	**0,011**
Kontinentalita / Continentality	0,726	0,688	0,148	**0,004**	0,452
Vlhkost / Moisture	-0,925	0,381	0,897	**< 0,001**	**< 0,001**
Živiny / Nutrients	-0,998	0,066	0,831	**< 0,001**	**< 0,001**
Půdní reakce / Soil reaction	-0,653	0,757	0,429	**< 0,001**	**0,032**

Tučně jsou vyznačeny hodnoty průkazné na hladině signifikance $P < 0,05$. DCA1 a DCA2– standardizované regresní koeficienty s první, resp. druhou ordinační osou. R^2 – koeficient determinace. P_{orig} – hladina průkaznosti původního permutačního testu. P_{modif} – hladina průkaznosti modifikovaného permutačního testu. / Values significant at the level $P < 0,05$ are printed bold. DCA1 and DCA2, respectively – standardized regression coefficients with the first and the second ordination axis, respectively. R^2 – coefficient of determination. P_{orig} – significance of the original permutation test. P_{modif} – significance of the modified permutation test.

Tab. 2. – Průkaznost korelace mezi Ellenbergovými indikačními hodnotami pro jednotlivé druhy a skóre druhů na prvních dvou ordinačních osách, testované (původním) permutačním testem.

Tab. 2. – Significance of the relationship between Ellenberg indicator species values and individual species scores on the first two ordination axes, tested by (original) permutation test.

	DCA1	DCA2	R^2	P
Světlo / Light	0,668	0,744	0,140	**0,002**
Teplota / Temperature	0,772	0,635	0,083	**0,017**
Kontinentalita / Continentality	0,646	0,763	0,027	0,295
Vlhkost / Moisture	-0,971	0,238	0,555	**< 0,001**
Živiny / Nutrients	-0,999	0,037	0,421	**< 0,001**
Půdní reakce / Soil reaction	-0,700	0,714	0,026	0,299

Tučně jsou vyznačeny hodnoty průkazné na hladině signifikance $P < 0,05$. DCA1 a DCA2– standardizované regresní koeficienty s první, resp. druhou ordinační osou. R^2 – koeficient determinace. P – hladina průkaznosti permutačního testu. / Values significant at the level $P < 0,05$ are printed bold. DCA1 and DCA2, respectively – standardized regression coefficients with the first and the second ordination axis, respectively. R^2 – coefficient of determination. P – significance of the permutation test.

Tab. 3. – Výsledky jednocestné analýzy variance (ANOVA) pro rozdíly průměrných Ellenbergových indikačních hodnot mezi pěti vegetačními typy.

Tab. 3. – Results of one-way ANOVA for differences of mean Ellenberg indicator values between five vegetation types.

	F	P_{param}	P_{modif}
Světlo / Light	36,242	**< 0,001**	**< 0,001**
Teplota / Temperature	7,149	**< 0,001**	0,148
Kontinentalita / Continentality	5,858	**< 0,001**	0,511
Vlhkost / Moisture	91,823	**< 0,001**	**< 0,001**
Živiny / Nutrients	42,982	**< 0,001**	**< 0,001**
Půdní reakce / Soil reaction	31,741	**< 0,001**	**0,014**

F – hodnota kritéria pro parametrický F-test. P_{param} – průkaznost vypočtená parametrickým F-testem. P_{modif} – průkaznost vypočtená modifikovaným permutačním testem. Hodnoty průkazné na hladině $P < 0,05$ jsou tištěny silně. / F – the value returned by parametric F-test. P_{param} – significance of the analysis, calculated by parametric F-test. P_{modif} – significance calculated by modified permutation test. Significant values ($P < 0,05$) are in bold.