

JAROSLAVA HORVÁTHOVÁ, JURAJ HRAŠKO

**TEORETICKO-METODOLOGICKÉ PRINCÍPY NUMERICKEJ TAXONÓMIE  
PŮD****I. VSTUPY A MODEL NUMERICKEJ TAXONÓMIE**

Jaroslava Horváthová, Juraj Hraško: Theoretical-methodological Principles of Numerical Soil Taxonomy. I. Inputs and Model of Numerical Taxonomy. Geogr. Čas., 36, 1984, 3; 1 figs, 3 tabs, 26 refs.

Numerical taxonomy of soils is based on the modified Adansonian principles. An unavoidable supposal of numerical classification methods are the selection of objects and characteristics, mathematic-statistic data elaboration, quantification and coding of characteristics. The problem solution alone originates from the cluster analysis, the model of which is based on three supposals: 1. forming of data matrix, 2. data matrix reduction to the matrix of similarity, 3. selection of suitable agglomerative strategy. The specific problem of soil profiles comparison is being considered.

## 1. ÚVOD

Numerické metódy klasifikácie sú v súčasnosti modernými smermi v rôznych vedných disciplínach. Spomeňme napr. systematickú botaniku a zoológiu, bakteriológiu, geológiu, fytoekológiu, medicínu, psychológiu a pedológiu. Problematika numerickej taxonómie s možnosťou aplikácie na pôdy predstavuje jeden z najnovších smerov spracovania terénnych výsledkov v pôdoznanectve. Prvé práce, ktoré sa pokúsili o nový spôsob klasifikácie, vznikli v 60. rokoch v USA, neskôr vo Veľkej Británii, Austrálii a Holandsku. Hole, Hironaka [14] ako prví aplikovali princípy novej taxonómie na klasifikáciu pôd, keď ordináčnou metódou Goodalla [10] a Braya, Curtisa [5] klasifikovali pôdy tzv. Miamskej série a 25 reprezentatívnych pôdných typov sveta, čím dokázali, že metódy numerickej taxonómie sú aplikovateľné na pôdy. Neskôr Bidwell, Hole [4] tou istou metódou klasifikovali 29 pôdných typov štátu Kansas, pričom pre ich usporiadanie vybrali 30 pôdných vlastností.

Súčasný vývoj budovania numerickej taxonómie pôd je v štádiu rozpracovania jej základných koncepcií a kompletného vypracovania všetkých súčastí numerickej klasifikácie na vybraných modelových územiach. Predstavuje akúsi predprípravu vytvorenia všeobecných základov numerickeho klasifikačného systému pôd.

Prvé práce v ČSSR, ktoré sa zaoberali problematikou numerickej klasifikácie pôd, resp. pôdnoekologických jednotiek, sme zaznamenali u Juráňa [15], Němečka [19] a Prachárovej [20].

Prudký vývoj v taxonómii umožnil elektronický počítač a moderné štatistické postupy. Sokal, Sneath [24] opísali tento vývoj ako novú alebo numerickú taxonómiu a definovali ju ako numerické hodnotenie afinity alebo podobnosti medzi taxonomickými jednotkami a usporiadanie týchto jednotiek do taxónov na základe ich afínit. V súčasnom období existuje úsilie zachytiť expanziu numerických metód, t. j. nájsť spoločné črty a základy existujúcich numerických metód klasifikácie, navzájom ich porovnať z hľadiska ich výpočtovej a triediacej efektívnosti a nájsť možnosti ich aplikácie v rôznych vedných odboroch. Tieto a mnohé ďalšie problémy sú predmetom štúdia botriológie (z gréc. *botrios* — stravec hrozna). Botriológia je definovaná ako veda o klasifikáciách, v ktorých sa používajú matematické (nie nevyhnutne numerické) metódy [9]].

Mnohé odborné termíny sú chápané v rôznych zmysloch, a preto často dochádza k nejednoznačnosti v ich používaní. Klasifikácia, systematika a taxonómia sú ľahko zameniteľné pojmy. Podľa Simpsona [24] možno jednotlivé termíny definovať takto:

- systematika študuje druhy, ich rozmanitosť a vzájomné vzťahy medzi nimi,
- klasifikácia usporadúva druhy podľa vzťahov blízkosti a podobnosti (takto definovanú klasifikáciu chápeme ako názov procesu, ktorého výsledkom je klasifikácia),
- taxonómia teoreticky študuje základy, princípy, postupy a pravidlá klasifikácie.

Bázou numerickej taxonómie sa stali princípy Adansona [1], ktoré vo svojej základnej práci o numerickej taxonómii rozvili Sokal, Sneath [24]. V súčasnom období sa niektoré axiómy znovu modifikujú, a to buď z hľadiska výpočtových možností, alebo z vedeckých, častejšie pragmatických dôvodov, preto zároveň uvedieme aj terajšie interpretácie niektorých „problémových“ axiém.

**Axióma 1.** Taxóny sa majú vyznačovať maximálnym obsahom informácií, preto treba vychádzať z čo možno najväčšieho počtu znakov.

Interpretácia axiómy 1: V zhode s intuitívnym i matematickým ponímaním obsah informácie o taxóne bude tým väčší, čím väčší bude počet objektov v taxóne a čím väčší bude počet vybraných znakov a príslušných znakových hodnôt. V súčasnom období sa uprednostňuje narastanie počtu znakov, avšak s rešpektovaním týchto limitujúcich kritérií:

- znak by mal reprezentovať podstatné (určujúce) a nie náhodne vybrané vlastnosti skúmaných objektov,
- vybrané znaky sa nesmú logicky prekrývať (majú byť štatisticky nezávislé),
- znaky musia dostatočne variovať, teda nemožno použiť znaky invariantné voči ostatným znakom.

**Axióma 2.** Pri zostrojení taxónu sa každému znaku prísudzuje rovnaká váha.

Interpretácia axiómy 2: Axióma 2 nepripúšťa apriórne pripisovanie rôznych váh rôznym znakom. Niektorým pedológom sa tento princíp zdá nevhodný a uvádzajú Gowerov vzorec pre výpočet podobnosti [11], ktorý vedie používateľa k voľnému výberu váh  $w_k$ . Prusinkiewicz, Caliński [21] hodnotia stupeň

dôležitosti (významnosti) znakov skúmaných pôd na základe Spearmanovho poradového koeficientu a vyberajú diagnosticky najdôležitejšie znaky. Ak je pôdny profil opísaný v rovnakých intervaloch, Russell, Moore [23] navrhli váhu  $w$ , ktorá pre každý interval exponenciálne klesá s narastajúcou hĺbkou. V tomto prípade priradujú väčšiu váhu znakom vo vrchnej časti profilu, pričom poukazujú na výraznú zmenu biologickej aktivity v pôdnom profile smerom nadol. Na druhej strane Webster [25] sa domnieva, že ak existujú dostupné údaje o objektoch (pôdnych profiloch), na základe ktorých sa majú hodnotiť skutočné vzťahy medzi nimi, potom nie je žiadny dôvod pre priradenie väčšej váhy jedným znakom pred druhými. Autor zohľadňuje len tie znaky, ktoré sú všeobecne známe ako dôležité.

**Axióma 3.** Podobnosť medzi dvoma taxonomickými jednotkami odpovedá podobnosti viacerých znakov a znakových hodnôt.

Interpretácia axiómy 3: Axióma 3 vychádza z výpočtov koeficientov podobnosti, resp. vzdialenosti. Týmto problémom sa budeme zaoberať v článku II.

**Axióma 4.** Rôzne taxóny sa dajú definovať na základe komplexu znakov a znakových hodnôt.

Interpretácia axiómy 4: Matematické poznatky sústredené v záverečnej klasifikačnej schéme treba interpretovať, t. j. taxónom treba pripísať konkrétny vedecký alebo praktický význam. V zložitejších situáciách sa u vedeckého pracovníka predpokladá značná dávka skúsenosti a odbornej praxe.

**Axióma 5.** Podobnosť je hodnotená nezávisle od behaviorálnych úvah.

Interpretácia axiómy 5: Niektorí numericky založení pedológovia odmietajú akékoľvek pedogenetické predpoklady, vrátane existencie horizontov, čím sa usilujú zachovať objektívny úsudok a zabrániť mnohým subjektívnym nepresnostiam. Muir a kol. [18] založil svoje pokusy na porovnávaní pôdnych profilov tradičnými a numerickými metódami. Získal takmer identické výsledky, čo vysvetľuje takto:

- numerickým či tradičným taxonómom je spoločne prístupný fond dát o pôde,
- takmer všetky vlastnosti pôdnych profilov sa identifikujú pomocou morfológie.

Autor uvádza niekoľko zdrojov nevyhnutnej subjektivity, s ktorou treba v numerickej taxonomii počítať:

1. Samé vlastnosti nie sú úplne objektívne, ale odrážajú všeobecný vedecký názor danej etapy (kvalita a vybavenosť vedeckých prístrojov, vyškolenie vedcov atď.).

2. Ďalším problémom je výber znakov. Kým numerický taxonóm má použiť všetky znaky bez predchádzajúceho udania váhy, tradičný taxonóm riadi výber znakov podľa stupňa dôležitosti a zaužívanosti. Ak sa numerický taxonóm rozhodne, že použije znaky bez udania váhy, v skutočnosti prisúdi jednotku váhy týmto znakom a ostatným neuvažovaným znakom prisúdi nulu.

3. Výber a počet klasifikovaných objektov tak isto riadi vedecký pracovník. Pretože úplný počet objektov nie je možné analyzovať, odporúča sa pracovať so štandardami, čím sa v určitej miere objektivita stráca.

4. Numerický taxonóm musí subjektívne rozhodnúť o voľbe najvhodnejšieho koeficientu podobnosti alebo vzdialenosti a tiež o vhodnej triediacej stratégii, čo však vyplýva z povahy numerických metód klasifikácie.

Základným krokom je zozbieranie informácií o indivíduách, ktoré majú byť študované. Informácia o nich vždy existuje cez znaky a vyžaduje sa buď výber znakov a znakových hodnôt z dostupných materiálov, alebo môžu byť nainoveno zistené. V súčasnosti sa kladie hlavný dôraz na získanie presných kvantitatívnych meraní, pretože namerané hodnoty sú najinformatívnejšie a na základe ich matematického spracovania možno dôjsť k najlepším výsledkom. Dobrá kvantitatívna báza údajov je teda dôležitým predpokladom získania objektívnejších výstupov. Fond dát o pôde ČSSR rozsahom a kvalitou vyhovuje pre progresívne riadenie využívania pôd, avšak z hľadiska využívania výpočtovou technikou je na nízkej úrovni.

Prvým krokom klasifikácie je výber objektov skúmania  $o_1, \dots, o_N$ . V numerickej taxonomii má objekt predstavovať logicky uzavretú a ďalej nedeliteľnú jednotku (OTU = operational taxonomic unit). V pôdnej klasifikácii sa za základnú taxonomickú jednotku pokladá pedón (najmenšia ľubovoľne vyrezaná časť pedosféry bez priestorového rozšírenia) a polypedón (homogénny výrez pedosféry s priestorovým rozšírením, skladajúci sa z rovnakých pedónov).

Objekty sa stávajú predmetom skúmania až vtedy, keď ich rozmanitosť redukuje určitým spôsobom na ich konkrétne vlastnosti. Preto sa pojem znaku stáva základným kľúčovým pojmom pre klasifikovanie. Je vhodné si jednotlivé termíny zdefinovať. Množinu objektov  $O = \{o_1, \dots, o_N\}$  budeme skúmať cez ich vlastnosti, ktorých reprezentantmi je množina znakov  $Z = \{z_1, \dots, z_m\}$ . Konkrétnu realizáciu znakov predstavuje množina znakových hodnôt  $H_{(z)} = \{h_1, \dots, h_m\}$ .

### 3.1 Kvantifikácia a kódovanie znakov

Numericke metódy klasifikácie vyžadujú, aby znaky a ich znakové hodnoty boli najskôr matematicky a potom numericke formalizované. Analýza znakov ukázala, že znaky môžu byť kvalitatívne a kvantitatívne.

Znak  $z_i$  nazývame kvantitatívnym, ak spĺňa tieto predpoklady:

a) Množina znakových hodnôt  $H_{(z)}$  je vnorená do reálnej priamky, t. j. prvkami tejto množiny sú reálne čísla.

b) Aspoň jedna z operácií sčítania, odčítania, násobenia alebo delenia má pre objekty presne vymedzený interpretačný význam.

Kvantitatívne znaky teda predstavujú číslícovo znázornené vlastnosti fyzikálnych veličín podľa „ $G$ “ =  $\{G\} \times [G]$ , kde „ $G$ “ je fyzikálna veličina,  $\{G\}$  merané číslo a  $[G]$  merná jednotka.

Znak  $z_i$  nazývame kvalitatívnym, ak podmienka b) nie je, resp. nemôže byť splnená a podmienku a) možno splniť vhodným zakódovaním. Pri kvalitatívnom znaku rozlišujeme buď poradovú, alebo nominálnu štruktúru množiny znakových hodnôt.

Poradové (ordinálne) znaky sa dajú zakódovať na základe možnej zoraditeľnosti znakov. Možno ich znázorniť stupnicou čísel, nejde však o fyzikálne veličiny, ale o zakódovanie znakov na základe relácie rovnosti, nerovnosti a preferencie.

Nominálne (alternatívne) znaky môžeme definovať vtedy, ak množina znakových hodnôt  $H_{(z)}$  je konečná a jej prvky  $h_1, \dots, h_m$  sú nezávislé znakové

Tab. 1. Údajová matica  $X_{(o)} = [x_{ik}]$ 

Objekty	Znaky	$z_1$	$z_2$	...	$z_k$	...	$z_n$	
$O_1$		$x_{11}$	$x_{12}$	...	$x_{1k}$	...	$x_{1n}$	$\bar{x}_1$
$O_2$		$x_{21}$	$x_{22}$	...	$x_{2k}$	...	$x_{2n}$	$\bar{x}_2$
.		.	.		.		.	.
.		.	.		.		.	.
$O_i$		$x_{i1}$	$x_{i2}$	...	$x_{ik}$	...	$x_{in}$	$\bar{x}_i$
.		.	.		.		.	.
.		.	.		.		.	.
$O_N$		$x_{N1}$	$x_{N2}$	...	$x_{Nk}$	...	$x_{Nn}$	$\bar{x}_N$

Tab. 2. Matica podobnosti  $P_{(o)} = [P_{ij}]$ 

Objekty	$O_1$	...	$O_j$	...	$O_N$
$O_1$	$P_{11}$	...	$P_{1j}$	...	$P_{1N}$
.	.		.		.
.	.		.		.
$O_i$	$P_{i1}$	...	$P_{ij}$	...	$P_{iN}$
.	.		.		.
.	.		.		.
$O_N$	$P_{N1}$	...	$P_{Nj}$	...	$P_{NN}$

mená objektov. Na rozdiel od poradových znakov sú tieto nezoraditeľné a možno ich len alfanumericky zakódovať.

Dichotomické (binárne, duálne) znaky sú špecifickým prípadom nominálnych znakov. Spočívajú na situácii, keď sa celá informácia zakladá na type prítomnosť — neprítomnosť znaku. Môže sa tiež použiť označenie 1—0, (+), — [—].

#### 4. KLASICKÝ MODEL KLASTROVEJ ANALÝZY

V numerickej taxonómii pod taxometrickými metódami rozumieme metódy kvantitatívneho stanovenia ešte neznámych taxónov za účelom vytvorenia taxonomického systému. Princíp taxometrických metód možno vyjadriť dvoma pojmami: redukcia a aglomerácia.

Redukcia (projekcia, zmenšenie rozmerov) spočíva v redukovaní počtu znakov a znakových hodnôt údajovej matice a v získavaní prehľadného počtu taxometrických merných čísel. Vzťahuje sa na výpočet koeficientov podobnosti alebo vzdialenosti a na extrakciu znakov na malý počet faktorov.

Agglomerácia (agregácia, zoskupovanie, zhlukovanie, klasterizácia z angl.

*cluster* = trs, chumáč) je proces, pri ktorom sa jednotlivé objekty zahŕňajú do skupín.

Klasický model klastrovej analýzy je založený na troch predpokladoch:

1. Pre danú množinu objektov  $o_1, \dots, o_N$  máme pre znaky  $z_1, \dots, z_n$  k dispozícii množinu znakových hodnôt. Maticu  $X_{(o)}$  typu  $N \times n$  s prvkami  $x_{ik}$  nazývame údajovou maticou (maticou dát).

Vektory  $\bar{x}_1, \dots, \bar{x}_N$  sú kódmi, obrazmi objektov v priestore hodnôt. Matica  $X_{(o)}$  obsahuje v  $i$ -tom riadku vektor  $\bar{x}_i$ :

$$\bar{x}_i = (x_{i1}, \dots, x_{in}), \text{ kde } i = (1, 2, \dots, N).$$

Tento vektor opisuje vlastnosti objektu  $o_i$  a je jeho reprezentantom. Preto niekedy hovoríme o klasifikácii vektorov  $\bar{x}_1, \dots, \bar{x}_N$ .

2. Pre každú dvojicu  $\{o_i, o_j\}$  vypočítame koeficient podobnosti  $P$ , resp. koeficient vzdialenosti  $V$ . Maticu  $P_{(o)}$ , resp.  $V_{(o)}$  typu  $N \times N$  s prvkami  $P_{ij}$ , resp.  $V_{ij}$  nazveme maticou podobnosti, resp. maticou vzdialenosti. Vznikne symetrická matica podobnosti [platí  $P_{ij} = P_{ji}$ ], pričom na hlavnej uhlopriečke sa nachádzajú hodnoty totožných objektov.

3. Vyberieme určité pravidlo (stratégiu) rozhodovania o zaradení objektov do tried. Toto pravidlo každej množine objektov a každej matici podobnosti, resp. vzdialenosti jednoznačne priradí klastery  $R_1, \dots, R_r$ . Takéto pravidlo je vždy formulované v termínoch vnútrotriednej podobnosti a medzitríednych rozdielov.

V rôznych pedologických štúdiách autori použili veľké množstvo mier a koeficientov podobnosti, ako aj triediacich stratégií, preto sa touto problematikou budeme zaoberať v článku II.

## 5. ŠPECIFICKÉ PROBLÉMY POROVNÁVANIA PÔDNÝCH PROFILOV

Pôdne charakteristiky sa značne odlišujú od znakov ostatných, dosiaľ numericky klasifikovaných objektov, takže priame použitie skorších metód na pôdoznanectvo bolo nemožné. Jednou z najväčších prekážok v aplikácii numerických metód klasifikácie na pôdu je anizotropnosť pôd [Moore, Russell [17]]. Anizotropnosť pôdných profilov sa odráža v rozdelení profilov do horizontov, ktoré možno považovať za izotropné [Knox [16]]. Grígal, Arneman [12] uvádzajú, že problémy v numerickej klasifikácii pôd sú spôsobené odlišnou povahou vlastností, ktoré opisujú pôdu: kvantitatívne (napr. skutočný obsah C), poradové (napr. odstupňovanie vlhkosti), nominálne (napr. štruktúra pôdy) a dichotomické (napr. prítomnosť mangánových konkrécií). Tento problém možno riešiť dvoma spôsobmi:

a) buď prejdeme na údaje jednotného typu, avšak tento postup spôsobuje určité obmedzenia, v dôsledku čoho sa pri niektorých údajoch stráca pôvodná informácia,

b) alebo použijeme vhodný koeficient podobnosti, ktorý tento problém zohľadňuje (napr. Gowerov všeobecný koeficient podobnosti).

Väčšina pôdných profilov sa charakterizuje usporiadaním horizontov v profile a vlastnosťami horizontov. Ako teda porovnávať pôdne profile ako operatívne taxonomické jednotky navzájom?

Najjednoduchším riešením je porovnať zodpovedajúce horizonty navzájom

Tab. 3. Príklad Raynerovho porovnania horizontov

		Profil $P_1$			
		1	2	3	4
Profil $P_2$	1	84,0	82,5	80,0	77,5
	2	78,0	84,0	83,0	81,5
	3	74,5	80,5	83,0	83,5
	4	73,5	79,0	84,5	84,5

(Bidwell, Hole [3]). Ak máme tri horizonty, každú vlastnosť prvého horizontu porovnáme s tou istou vlastnosťou prvého horizontu v druhom profile, atď. Avšak takéto porovnávanie horizontov je dosť subjektívnym aktom, ktorý môže v podstatnej miere viesť k ovplyvneniu numerickej klasifikácie. Aby sme sa vyhlí budoj skutočnosti, profily sa môžu písať v niekoľkých, pevne stanovených bodoch (Moore, Russell [17]). Tento princíp sa z hľadiska numerickej taxonómie zdá byť najvhodnejší, avšak pre stanovené body je potrebné vytvoriť údajovú bázu.

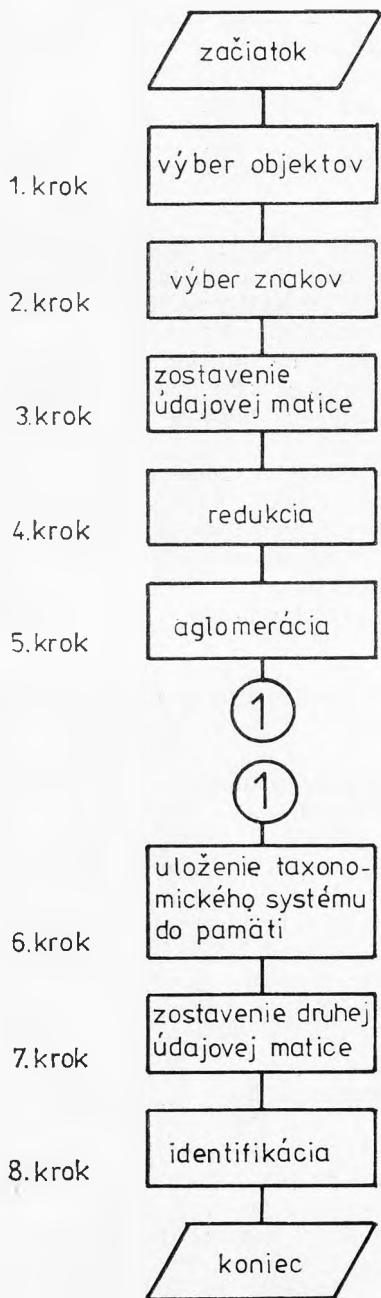
Rayner [22] navrhol postup, ktorý zabraňuje mnohým dohadom. Každý horizont v každom profile je chápaný ako jedinec a koeficienty podobnosti sa vypočítavajú pre každý pár horizontov. Podobnosť medzi dvoma profilmi  $P_1$  a  $P_2$  sa potom vypočíta z hodnôt koeficientov podobnosti medzi horizontmi. Prvý horizont v  $P_1$  je porovnaný s každým horizontom v  $P_2$ . Najpodobnejší horizont s prvým horizontom v  $P_1$  je potom vybraný ako najlepšia zhoda. Druhý horizont v  $P_1$  je ďalej porovnávaný s každým horizontom v  $P_2$  začínajúc od najzhodnejšieho horizontu pre prvý horizont a postupuje sa dovedy, až kým nemáme všetky horizonty  $P_1$  porovnané s horizontmi  $P_2$ . Úlohy  $P_1$  a  $P_2$  sa vymenia a postup sa opakuje, pričom získané hodnoty nie sú symetrické. Konečný koeficient podobnosti je vypočítaným priemerom najzhodnejších horizontov. V uvedenej tabuľke orámované koeficienty podobnosti ukazujú najzhodnejšie horizonty. Výsledný koeficient podobnosti medzi profilmi  $P_1$  a  $P_2$  bude 84, 25. Raynerova metóda má viacero predností. Nielenže umožňuje porovnávať profily v rôznych hĺbkach alebo v rôznych horizontoch, ale tiež dovoľuje automaticky vypočítať podobnosti medzi profilmi s rôznym počtom horizontov. Okrem toho berie do úvahy usporiadanie horizontov v profile, pretože napr. pri opačne obrátenom profile dostaneme rozdielne podobnosti s druhým profilom. Nevýhodou tejto metódy je pomerne vysoká časová náročnosť výpočtov, z ktorých väčšina je nepoužiteľná.

## 6. DIAGNOSTICKÝ KLÚČ A IDENTIFIKÁCIA

Ak sme pomocou klastrovej analýzy alebo faktorovej analýzy získali príslušné taxóny, je záujem určiť tie znaky, ktoré sú pre rozpoznávanie a diagnózu najdôležitejšie. Za takéto všeobecne považujeme znaky:

- ktoré sa vyznačujú stálosťou znakových hodnôt v rámci príslušného taxónu,
- ktoré vykazujú vysokú koreláciu u čo najväčšieho počtu iných znakov.

Prvý druh znakov získame utvorením priemernej hodnoty a štandardnej odchýlky v rámci príslušného taxónu. Z nich vyplývajúci variačný koeficient musí vykazovať minimálne hodnoty.



Obr. 1. Základný plán taxometrickej analýzy



Druhý druh znakov určíme z matice korelačných koeficientov medzi príslušnými znakmi. Vyberáme len znaky spojené vysokými hodnotami korelačných koeficientov.

Definovaním exemplárnych taxonomických jednotiek pre príslušné taxóny vytvoríme diagnostický kľúč. Pod identifikáciou rozumieme priradenie určitého objektu k už existujúcemu a definovanému taxónu. V literatúre sa tento problém označuje termínom *diskriminácia*, pod ktorou rozumieme vyčíslenie hodnoty funkcie diskriminácie závislej od daného objektu a posudzovaného taxónu. Priradenie neznámeho objektu k taxónu sa uskutočňuje dvoma spôsobmi:

1. Porovnávame daný objekt so všetkými taxonomickými jednotkami, ktoré vytvárajú taxón.

2. Porovnávame daný objekt s exemplárnou taxonomickou jednotkou (vychádzame teda z diagnostického kľúča).

Porovnanie sa uskutočňuje podľa tých istých koeficientov podobnosti, pomocou ktorých sa získali taxóny. Posudzovaný objekt priradíme k tomu taxónu, ku ktorému je najbližšie, t. j. maximálne hodnoty koeficientu podobnosti alebo minimálne hodnoty koeficientu vzdialenosti objektu porovnávame s príslušnými, do úvahy prichádzajúcimi taxónmi.

Kvôli prehľadnosti celej problematiky numerickej taxonómie uvedieme zjednodušený plán schému taxometrickej analýzy. V piatom kroku schémy je vhodné požadovať výstup dosiahnutých výsledkov (najčastejšie grafický). V siedmom kroku sa musí zostaviť údajová matica, ktorá sa zakladá na tých istých znakoch, pomocou ktorých sa uskutočnila redukcia a aglomerácia.

### 6.1 Vykresľovacie metódy

Najčastejšie sa používa znázornenie pomocou grafu, ktorý prehľadne vykresľuje vzťahy medzi klasifikovanými objektmi a je nevyhnutnou požiadavkou pre výstup v počítači. Opis alebo funkcia daného predmetu vo viacrozmernom znakovom priestore sa pri grafoch redukuje na dvojrozmernú rovinu, pričom smer jednej z osí nemá žiadny taxonomický význam. Príkladom takého grafu je dendrogram, ktorý znázorňuje hierarchické usporiadanie taxónov vo vertikálnom smere. Klasifikáciu s určitým počtom tried získame vyťahnutím horizontálnej čiary cez prierez toho istého počtu vertikálnych kmeňov. Kým pri dendrograme horizontálny smer nemá taxonomický význam, pri dendrografe možno v horizontálnom smere znázorniť poradie a vzdialenosť medzi taxonomickými jednotkami.

Pri vykresľovacej metóde na základe extrakcie znakov sa pri získaní dvoch faktorov vypočítajú ich hodnoty pre všetky klasifikované objekty. Jednotlivé zoskupenia taxonomických jednotiek sa znázornia v dvojrozmernej rovine faktorov. Ak sa pri extrakcii znakov získali tri faktory, v takom prípade nakreslíme dve navzájom kolmé roviny  $F_1F_2$  a  $F_1F_3$ , pričom os  $F_1$  tvorí ich priesečník. Pre faktorové hodnoty príslušných taxonomických jednotiek dostaneme dve projekcie trojrozmerného rozdelenia bodov.

## 7. ZÁVER

Teoretické poznatky v oblasti numerických metód klasifikácie nie sú zatiaľ dostatočne podložené praktickým výskumom. Preto nevyhnutnou etapou roz-

voja numerických metód klasifikácie pôd by malo byť overenie týchto metód na modelovom území SSR, na základe čoho by vznikla zaujímavá možnosť porovnania novovzniknutej klasifikácie s dosiaľ zaužívanými genetickými klasifikáciami a zároveň by sa mohli potvrdiť súčasné názory na polygenetický vývoj pôd.

#### LITERATÚRA

1. ADANSON, M.: Familles des plantes, vol. 1. Preface pp. cliv et seq., CLXIII, CLXIV. Vincent Paris, 1763. — 2. AVERY, B. W.: General soil classification: hierarchical and co-ordinate systems. 9<sup>th</sup> Int. cong. of soil sci., Adelaide 1968, ss. 169—175. — 3. BIDWELL, D. W., HOLE, F. D.: Numerical taxonomy and soil classification. Soil Sci., 97, 1964, ss. 58—62. — 4. BIDWELL, D. W., HOLE, F. D.: An experiment in the numerical classification of some Kansas soils. Soil Sci. Soc. Am. Proc., 28, 1964, ss. 263—268. — 5. BRAY, J. R., CURTIS, J. T.: An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr., 27, 1957, ss. 325—349. — 6. COLE, A. J. red.: Numerical taxonomy. Acad. Press Lond. and New York 1969. — 7. ČÍŽEK, F.: Metodologické problémy klasifikace. Filosof., 5, 1970, ss. 757—786. — 8. ELASHOFF, J. D. a kol.: On the choice of the variates in classification problems with dichotomous variables. Biometrika, 54, 1967, ss. 668—670. — 9. GOOD, I. J.: Botriologija botriologiji. Zbor. Klasifikacija i klaster, 1980. — 10. GOODALL, D. W.: Vegetational classification and vegetational continua. Angrew. Pflanzensoziologie, 1, 1954, ss. 168—182.
11. GOWER, J. C.: A general coefficient of similarity and some of its properties. Biometrics, 27, 1971, ss. 857—874. — 12. GRIGAL, D. F., ARNEMAN, H. P.: Numerical classification of some forested Minnesota soils. Soil Sci. Soc. Am. Proc., 33, 1969, ss. 433—438. — 13. GRUIJTER, J. J.: Numerical classification of soils and its application in survey. Neth. Soil Sur. Inst. Wageningen. 1977. — 14. HOLE, F. D., HIRONAKA, M.: An experiment in ordination of some soil profiles. Soil Sci. Soc. Am. Proc., 24, 1960, ss. 309—312. — 15. HORVÁTH, E.: Metódy štatistickej analýzy kvalitatívnych znakov. Časť klasifikácie. ÚVTVŠ, Bratislava 1983. — 16. JURĀN, C.: Matematický princíp triedenia pôdných jednotiek pre účely kategorizácie. VÚPVR, Bratislava 1979. — 17. KNOX, E. G.: Soil individuals and soil classification. Soil Sci. Soc. Am. Proc., 29, 1965, ss. 79—84. — 18. MOORE, A. W., RUSSELL, J. S.: Potential use of numerical analysis and Adansonian concepts in soil science. Austr. J. Sci., 29, 1966, ss. 141—143. — 19. MUIR, J. W. a kol.: The classification of soil profiles by traditional and numerical methods. Geoderma, 4, 1970, ss. 81—90. — 20. NĚMEČEK, J.: Základní diagnostické znaky a klasifikace půd ČSR. Studie ČSAV Acad., Praha 1981.
21. PRACHÁROVÁ, M.: Zoskupovanie a analýza podobnosti bonitovaných pôdnoekologických jednotiek pomocou matematicko-štatistických metód. VÚEPP, Bratislava 1982. — 22. PRUSINKIEWICZ, Z., CALIŃSKI, T.: Untersuchungen über Anwendbarkeit der statistischen Methoden in der Bodensystematik. Roczn. Gleb. PWN, 1964, ss. 285—293. — 23. RAYNER, J. H.: Classification of soil by numerical methods. Soil Sci., 17, 1966, ss. 79—92. — 24. RUSSELL, J. S., MOORE, A. W.: Comparison of different depth weighting in the numerical analysis of anisotropic soil profile data. 9<sup>th</sup> Int. cong. of soil sci. Adelaide 1968, ss. 205—213. — 25. SOKAL, R. R., SNEATH, P. H. A.: Principles of numerical taxonomy. Freeman, New York 1963. — 26. WEBSTER, R.: Quantitative and numerical methods in soil classification and survey. Monogr. on Soil Sur., Clarendon Press, Oxford 1977.

# ТЕОРЕТИЧЕСКО-МЕТОДОЛОГИЧЕСКИЕ ПРИНЦИПЫ НУМЕРИЧЕСКОЙ ТАКСОНОМИИ ПОЧВ

## 1 СОДЕРЖАНИЕ И МОДЕЛЬ НУМЕРИЧЕСКОЙ ТАКСОНОМИИ

Проблематика нумерической таксономии почв представляет одно из современных направлений обработки результатов полевых и лабораторных исследований почв в целях их классификации. Первые попытки были сделаны в 60-тых годах в США, позже в Англии, Австралии и Голландии. В настоящее время нумерическая классификация вообще, и почв в частности, развивается в ботриологии, которая является научной дисциплиной о классификациях с применением математических и нумерических методов.

Основные аксиомы нумерической таксономии исходят из принципов Адансона, а именно:

1. Таксоны должны содержать максимум информации, поэтому нужно исходить из максимального доступного количества признаков.
2. При составлении таксона любому признаку дается одинаковой удельный вес.
3. Сходство двух таксономических единиц отвечает сходству множества признаков а их величине.
4. Разные таксоны определяются на основе комплекса признаков и отвечающей им величине.
5. Сходство не зависит от генеза признака, а от его присутствия или отсутствия.

На практике отдельные аксиомы как правило полностью не соблюдаются, что обусловлено или применяемой вычислительной техникой, или другими прагматическими аргументами.

Подготовка к нумерической классификации включает:

- подбор объектов классификации  $o_1, \dots, o_M$ , которыми является как правило педон или полипедон,
- подбор признаков  $z_1, \dots, z_n$ ,
- математическо-статистическая обработка данных,
- квантификация и шифровка признаков.

По опыту авторов, наиболее важным препятствием широкого применения методов нумерической классификации почв является анизотропия почвенных профилей и различный тип признаков характеризующих почву. Для сравнения почвенных профилей оказался как самый подходящий метод по Мооре-Расселю и метод по Райнеру. В статье приводится основной план таксометрического анализа, диагностический ключ, идентификация и методы разрысовки.

Рис. 1. Основной план таксометрического анализа.

Таб. 1. Матрица данных  $X(o) = (x_{ik})$ .

Таб. 2. Матрица сходства  $P(o) = (P_{ij})$ .

Таб. 3. Пример сравнения почвенных горизонтов по Райнеру.

Перевод: Ю. Грашко

THEORETICAL-METHODOLOGICAL PRINCIPLES OF NUMERICAL SOIL TAXONOMY  
I. INPUTS AND MODEL OF NUMERICAL TAXONOMY

Questions of numerical soil taxonomy belong to the newest direction of field results processing in soil science. First attempts of the new method of classification have appeared in the U.S.A in the 1960s, later it was developed in Great Britain, Australia and the Netherlands. In the present, problems of numerical taxonomy are included into a study subject of botriology defined as classification science using mathematical and numerical methods.

Basic axioms of numerical taxonomy stem from the principles of Adanson:

1. Taxa should be characterized by maximum content of information, it is needed, therefore, to emanate from number of characters as large as possible.
2. When constructing a taxon, the same weight is assigned to each character.
3. Similarity between two taxonomic units corresponds to similarity of several characters and characters values.
4. Different taxa can be defined on the basis of complex of characters and relevant character values.
5. Similarity is weighted independently on genetic considerations. Individual axioms are modified either due to viewpoint of computer possibilities, or due to pragmatic reasons in the present time.

Preparatory steps of numerical soil classification include:

- selection of classified soil objects (pedon, polypedon)  $O_1, \dots, O_N$ ,
- selection of characters  $Z_1, \dots, Z_n$ ,
- mathematical-statistical data processing,
- character quantification and coding.

Anisotropy of soil profiles and different nature of soil describing characters are the greatest obstacles of numerical classification methods application to soil. There are several methods for soil profiles comparison, from which method of Moore—Russel and Rayner's method seem to be the most useful.

A basic plan of taxometric analysis is given for clear presentation of the whole problems of numerical taxonomy; we have briefly made reference to diagnostic key, identification and drawing methods.

Fig. 1. Basic plan of taxometric analysis.

Table 1. Data matrix  $X_{(o)} = [x_{ik}]$ .

Table 2. Similarity matrix  $P_{(o)} = [P_{ij}]$ .

Table 3. Example of Rayner's comparison of horizons.

Translated by T. Antalová