# RELATÍVNA INFORMAČNÁ ENTROPIA A GAZDOVA HYDROCHEMICKÁ KLASIFIKÁCIA

## RELATIVE INFORMATION ENTROPY AND GAZDA´S HYDROCHEMICAL CLASSIFICATION

Ján Jetel


## ABSTRACT

In classifying the chemical composition of water by Gazda´s classification, relative information entropy $H_r$ of multicomponent system (Pelto, 1954) should be used to express the typological distinctness of water (i.e. to what degree the chemical type is clean-cut). Eqs. (3) and (1) define the relative information entropy (RIE) in the n-component system, the symbol $p_i$ designating the portions of particular components ($\sum p_i = 1$). In Gazda´s classification, the components i are represented by hypothetical ionic combinations $A_2$, $S_1(Cl)$, $S_1(SO_4)$ etc.(maximum number of coexistent components n = 6). For the maximum type indefiniteness (i.e. equal proportion of all possible components) is $H_r = 1$, while $H_r = 0$ for a water containing 100% of the only component. The RIE can be also correlated with other relevant variables.



## KEY WORDS

Hydrochemical classification, information entropy, multicomponent system

## KĽÚČOVÉ SLOVÁ

Hydrochemická klasifikácia, informačná entropia, mnohozložkový systém


V návrhu klasifikácie chemizmu prírodných vôd (Gazda, 1971) prijal S. Gazda popri inom jeden princíp, ktorý v ostatných podobných klasifikáciách chýbal: rozlíšenie vôd nielen podľa najviac zastúpených zložiek, ale aj podľa vyhranenosti podielov jednotlivých zložiek. Na zohľadnenie tohto aspektu zvolil rozlíšenie typov vôd na výrazný základný, nevýrazný základný, prechodný a zmiešaný, definované podľa stupňa vyhranenosti podielu zložiek určujúcich daný typ vody. Tento postup, zohľadňujúci mieru typologickej čistoty vody, zvýšil nesporne hodnotu navrhnutého klasifikačného systému. Praktické skúsenosti s používaním tohto systému však ukázali na niektoré problémy spojené s jeho aplikáciou pri klasifikovaní vôd s nízkou vyhranenosťou – s podielom najviac zastúpenej zložky menšom ako 50 ekv.% - t.j. pri vodách prechodných a zmiešaných typov (určité narušenie logickej kontinuity medzi základným, prechodným a zmiešaným typom pri postupných kontinuálnych zmenách podielov jednotlivých zložiek a i.).

Skúsenosti s použitím Gazdovej klasifikácie ukázali, že najmä pri vodách s nižšou vyhranenosťou (s nižšou „typovou čistotou") by bolo vhodné nahradiť alebo aspoň doplniť vymedzenie jednotlivých typov v zmysle stupňov vyhranenosti racionálnejším vyjadrením stupňa vyhranenosti jediným kontinuálne sa meniacim ukazovateľom. Ako veľmi vhodný ukazovateľ tohto druhu môže slúžiť relatívna informačná entropia $H_r$.

Pri definovaní hodnôt $H_r$ vychádzame z aplikácie, ktorú pre mapovanie

**doc. RNDr. Ján Jetel, CSc.**
Šrobárova 48, 058 01 Poprad, jetel@stonline.sk

mnohozložkových systémov navrhol Pelto (1954) na základe Shannonovej definície entropie v zmysle teórie informácie (Shannon & Weaver, 1949). Informačná entropia n-zložkového systému je definovaná vzťahom

$$H_r = -\sum_i p_i \ln p_i \qquad (1)$$

kde $p_i$ = podiel i-tej zložky systému ($\sum p_i = 1$). Maximálna informačná entropia pre systém zložený z n zložiek je daná vzťahom

$$H_{max} = -\sum_{i=1}^{n} (1/n) \ln(1/n) = -\ln n \qquad (2)$$

Na základe uvedených vzťahov zaviedol Pelto (1954) pojem „relatívnej informačnej entropie", definovanej ako

$$H_r = H / H_{max} = H / \ln n \qquad (3)$$

Maximálne nevyhranený systém, v ktorom sú podiely všetkých zložiek rovnaké, má hodnotu $H_r = 1$. Naproti tomu maximálne vyhranený systém, tvorený iba jedinou zložkou z n možných, má nulovú hodnotu $H_r$

V dostupnej zahraničnej literatúre sa pojem relatívnej informačnej entropie aplikoval v mapách faciálneho vývoja sedimentov z hľadiska diverzity výskytu jednotlivých biofácií (Parker et al., 1953) a v iných obdobných sedimentologických aplikáciách (Miller & Kahn, 1958, Kullback, 1959, Griffiths, 1967).

Vyhranenosť chemického zloženia vody z hľadiska podielov jednotlivých zložiek (iónových kombinácií) definovaných v Gazdovej klasifikácii kvantitatívne vyjadríme pomocou hodnoty vnútornej relatívnej informačnej entropie $H_r$ (Jetel, 1975, 1987), ktorá predstavuje mieru nevyhranenosti typu vody. Označenie „vnútorná (interná) entropia" (IRIE) sme zaviedli vzhľadom na vyjadrenie typologickej neurčitosti vnútri určitej vzorky vody na odlíšenie od „vonkajšej (externej) entropie" ERIE, ktorou môžeme vyjadrovať diverzitu chemického zloženia vôd v určitom súbore vzoriek – pri kartografickom vyjadrení, pri hodnotení hydrochemickej homogenity určitého prostredia apod. (Jetel, 1987).

Na vyjadrenie typologickej vyhranenosti vody z hľadiska Gazdovej klasifikácie možno za podiely zložiek $p_i$ zvoliť podiely jednotlivých iónových kombinácií (zložiek) v Gazdovej klasifikácii (Jetel, 1975, 1987), t. j. $A_1$ , $A_2$, $S_1(Cl)$, $S_1(SO_4)$, $S_2(SO_4)$ atď. Keďže sa prítomnosť niektorých z týchto kombinácií vzájomne vylučuje, môže sa vo vode súčasne vyskytnúť maximálne 6 z týchto zložiek. Hore uvedené rovnice (1) - (3) preto pri hodnotení podľa Gazdovej klasifikácie aplikujeme na prípad 6-zložkového systému (n = 6). Maximálne vyhranená je voda, ktorá obsahuje 100% jedinej z Gazdových zložiek (napr. $A_2$ = 100 ekv.%); charakterizuje ju hodnota $H_r = 0$. Naproti tomu maximálne nevyhranená („maximálne zmiešaná") voda s hodnotou $H_r = 1$ obsahuje 6 zložiek s rovnakým zastúpením (s rovnakými podielmi jednotlivých zložiek) $p_i$ = 1/6 = 16,66 ekv.%.

Podiel $P_i$ každej jednotlivej zložky i charakterizovanej príslušným Gazdovým indexom a vyjadrený v ekv.% prepočítame na podiel $p_i = P_i /100$ vyjadrený ako dekadický zlomok. V prípade, že bola vo vode zistená prítomnosť menej ako šiestich Gazdových zložiek, chýbajúcim zložkám (zložkám s nulovým podielom) sa priradia fiktívne hodnoty $P_i$ = 0,01 až do $P_6$. Čiastkové hodnoty entropie pre jednotlivé prítomné zložky i = 1, 2, ...6 potom vypočítame ako

$$h_1 = -p_1 \ln p_1, h_2 = -p_2 \ln p_2 \ldots h_6 = -p_6 \ln p_6 \qquad (4)$$

Po stanovení súčtu čiastkových hodnôt

$$H = \sum h_i = h_1 + h_2 + h_3 + h_4 + h_5 + h_6 \qquad (5)$$

vypočítame relatívnu informačnú entropiu zmysle rovnice (3) ako

$$H_r = \sum h_i / \ln 6 = \sum h_i / 1{,}7918 \qquad (6)$$

Ako príklad môžeme uviesť výpočet vnútornej relatívnej informačnej entropie pre vodu s nasledujúcimi hodnotami Gazdových zložiek (hypotetických iónových kombinácií) $P_i$ (v danom konkrétnom prípade ide v Gazdovej klasifikácii o vodu zmiešaného typu s najvyšším zastúpením zložky A2) :

$S_1(NO_3)$ = 5,2 ekv.%
$S_1(Cl)$ = 10,0 ekv.%
$S_1(SO_4)$ = 9,7 ekv.%
$S_2(SO_4)$ = 27,9 ekv.%
$A_2$ = 47,2 ekv.%
$A_3$ = 0,0 (0,01) ekv.%

Po prevedení uvedených hodnôt $P_i$ na podiely $p_i$ vyjadrené formou desatinných zlomkov ($p_i = P_i /100$) vyjdú pre jednotlivé zložky tieto čiastkové entropie:

$h_1 = 0,154$
$h_2 = 0,230$
$h_3 = 0,226$
$h_4 = 0,356$
$h_5 = 0,354$
$h_6 = 0,001$

Celková entropia $H = \sum h_i$ bude potom mať hodnotu 1,321, takže vnútorná relatívna informačná entropia danej vody bude mať hodnotu

$$H_r = H/1,792 = 0,737$$

Prvou nespornou výhodou racionálneho vyjadrenia miery nevyhranenosti hodnotami vnútornej relatívnej informačnej entropie v porovnaní s členením vôd na základný, prechodný a zmiešaný typ je vyjadrenie nevyhranenosti kontinuálne sa meniacou metrickou veličinou. Kvantifikuje sa tak odlíšenie čistých typov od prechodných a zmiešaných jediným kvantitatívnym ukazovateľom.

Ak by sa pokladalo za účelné aj určité roztriedenie vôd podľa ich vyhranenosti v nadväznosti na doteraz používané Gazdovo členenie, bolo by možné použiť nasledujúce roztriedenie vôd z hľadiska stupňa vyhranenosti podľa hodnôt $H_r$:

- vody veľmi výrazne vyhranené s $H_r < 0,33$ (odpovedá zhruba vodám s podielom hlavnej zložky nad 83 ekv.%),
- vody výrazne vyhranené s $H_r = 0,33 - 0,53$ (odpovedá zhruba vodám s podielom hlavnej zložky 67 - 83 ekv.%),
- vody nevýrazne vyhranené s $H_r = 0,53 - 0,68$ (odpovedá približne vodám s podielom hlavnej zložky 50 - 67 ekv.%, t. j. vodám nevýrazného základného typu v Gazdovej klasifikácii),
- vody veľmi slabo vyhranené s $H_r > 0,68$ (vody s podielom hlavnej zložky pod 50 ekv.%, t.j. prechodné a zmiešané vody Gazdovej klasifikácie).

Kvantitatívne vzťahy medzi hodnotou $H_r$ a rozlíšením základných, prechodných a zmiešaných typov Gazdovej klasifikácie boli stanovené na základe výsledkov regresného rozboru vzťahu medzi podielom hlavnej zložky v Gazdovej klasifikácii a hodnotou $H_r$ na materiáli tvorenom niekoľkými stovkami analýz.

Hodnoty IRIE charakterizujú samozrejme iba samotný stupeň vyhranenosti bez konkrétnej identifikácie jednotlivých prítomných zložiek alebo konkrétneho typu vody. V mape však možno skombinovať vyjadrenie konkrétnych typov farbami alebo typom rastra (šrafy) s vyjadrením stupňa vyhranenosti napr. druhom či hustotou rastra (šrafy) alebo intenzitou farby. Hodnota IRIE preto pri kartografickom vyjadrení chemického zloženia vôd poslúži ako doplnková charakteristika.

Výhodou použitia hodnôt vnútornej (internej) relatívnej informačnej entropie (IRIE) je ďalej skutočnosť, že sa prezentuje veličina, ktorú možno použiť nielen pri kartografickom zobrazení chemizmu podzemných vôd, no aj ako jednoznačne definovanú premennú pri štúdiu najrôznejších kauzálnych aj priestorových vzťahov. Ako sme už spomenuli, hodnoty IRIE je výhodné použiť v mape na kartografické vyjadrenie typu vody súčasne s vyjadrením jeho vyhranenosti („typologickej čistoty"). Vizualizuje sa tak rozmiestnenie území čistých typov (resp. zdrojov čistých typov) a postupné miešanie medzi týmito územiami (kontinuálne zmeny IRIE v ploche). Možno tak sledovať aj priestorový vývoj znečisťovania vôd, pri ktorom spravidla rastie hodnota IRIE. Príklady použitia hodnôt IRIE pri regionálnom hodnotení chemického zloženia podzemných vôd uvádzajú Jetel (1971, 1987b, 1989, 1998), Jetel et al. (1990,1993) a Jetel & Rybářová (1991).

Jestvujú aj ďalšie možnosti praktického využitia hodnôt IRIE. Tak napr. pri opise a klasifikácii podľa pozície zobrazovacieho bodu v n-rozmernom diagrame môže IRIE vystupovať ako jedna z opisných súradníc. Vnútorná relatívna informačná entropia môže napokon slúžiť ako jedna z premenných pri štúdiu priestorových vzťahov chemizmu a riešení genézy vôd. V niektorých regiónoch môžu existovať výrazné korelačné vzťahy medzi IRIE a piezometrickým napätím (Jetel, 1971), vo vodách prameňov možno skúmať koreláciu medzi IRIE a redukovanou dĺžkou filtrácie (Jetel, 1989) alebo nadmorskou výškou.

Popri vnútornej relatívnej informačnej entropii sa v hydrogeochémii a hydrogeológii nedoceňuje

ani praktický význam vonkajšej (externej) relatívnej informačnej entropie (Jetel, 1987). Ako sme sa už zmienili, vonkajšia relatívna informačná entropia ERIE vyjadruje typologickú pestrosť (diverzitu) zastúpenia jednotlivých typov vôd v skúmanej množine vzoriek – napr. v určitej priestorovej jednotke. Zložkami sú tu jednotlivé možné typy (pritom však treba vopred definovať počet možných typov – t.j. počet zložiek systému). Vonkajšia relatívna informačná entropia tak môže v mape charakterizovať stupeň generalizácie: čím vyššia je vonkajšia relatívna informačná entropia, tým menej informácie o výskytu určitého typu poskytuje pre danú plochu generalizované zobrazenie, priraďujúce tejto ploche iba jeden charakteristický (najviac zastúpený) typ – tým nižšia je teda informačná jednoznačnosť mapy v danej ploche. Použitie vonkajšej relatívnej informačnej entropie umožní skúmať aj typologickú diverzitu vôd ako funkciu druhu horninového prostredia, pozície v morfológii terénu, klimatických pomerov a iných faktorov (Jetel, 1987b).

## LITERATÚRA

GAZDA, S., 1971: Modifikácia Palmerovho klasifikačného systému. Hydrogeol. Ročenka 1969-1970, Praha, 122-126.

GRIFFITHS, J. G., 1967: Scientific methods in analysis of sediments. McGraw-Hill Book Comp., New York - Sydney, 421 pp.

JETEL, J., 1971: Hydrogeologie a hydrogeochemie podloží české křídy. Závěrečná zpráva. Manuskript - Geofond Praha. 488 s.

JETEL, J., 1975: Application of information entropy in hydrogeochemistry and in hydrogeochemical maps. Věst. Ústř. Úst. geol. (Praha), 50, 1, 33-36.

JETEL, J., 1987a: Relatívna informačná entropia v hydrogeochémii a hydrogeochemických mapách. In: Hornická Příbram ve vědě a technice (1987), Sb. sekce Matematické metody v geologii - M1, Příbram, 354-356.

JETEL, J., 1987b: Vysvětlivky k základní hydrogeologické mapě ČSSR 1:200 000 list 25 Gottwaldov. Ústř. ústav geol., Praha. 216 s.

JETEL, J., 1989: Relationship between hydrogeological characteristics of near-surface zone of rock massif and hydrodynamic conditions. Západ. Karpaty. Sér. Hydrogeol. inž. Geol. (Bratislava), 8, 67-104.

JETEL, J., 1998: Hydrogeologická mapa Ľubovnianskej vrchoviny a Pienin 1:50 000 – textové vysvetlivky. Manuskript, archív Štát. geol. ústavu D. Štúra, Bratislava. 105 s.

JETEL, J., MOLNÁR, J. & VRANOVSKÁ., A., 1990: Hydrogeologický výskum Hornádskej kotliny - záverečná správa. Manuskript, archív Štát. geol. ústavu D. Štúra, Bratislava. 288 s.

JETEL, J., NEMČOK, J. & TKÁČ, J., 1993: Hydrogeologický výskum Spišskej Magury - záverečná správa. Manuskript, archív Štát. geol. ústavu D. Štúra, Bratislava. 222 s.

JETEL, J. & RYBÁŘOVÁ, L., 1991: Hydrogeologie cenomanu labské oblasti české křídové pánve. Sbor. geol. Věd, Hydrogeol., inž. Geol., (Praha), 19, 9-63.

KULLBACK, S., 1959: Information theory and statistics. J. Wiley & Sons, New York – London. 395 pp.

MILLER, R. L. & KAHN, J. S., 1962: Statistical analysis in the geological sciences. J. Wiley & Sons, New York – London. 482 pp.

PARKER, F. I., PHLEGER, F. B. & PEIRSON, J. F., 1953: Ecology of foraminifera from San Antonio Bay and environs. Southwest Texas. Spec. Publ. No. 2, Cushman Found. for Foraminiferal Res.

PELTO, C. E., 1954: Mapping of multicomponent systems. J. Geol. (Chicago), 62, 5, 151-155.

SHANNON, C. E. S. & WEAVER, W., 1949: The mathematical theory of communications. Univ. of Illin. Press, Urbana.

# RELATIVE INFORMATION ENTROPY AND GAZDA´S HYDROCHEMICAL CLASSIFICATION

## RÉSUMÉ

In classifying the chemical composition of water by Gazda´s classification, the concept of the relative information entropy $H_r$ of multicomponent system (Pelto, 1954) should be used to express the typological distinctness of water (i. e. to what degree the chemical composition is clean-cut). The relative information entropy (RIE) in the n-component system is expressed by Eq. (3) derived from the Shannon and Weaver´s definition of the information entropy (1). The symbol $p_i$ designates the shares (portions) of particular components (expressed in decimal fractions, thus $\sum p_i = 1$). In applying this concept to Gazda´s classification, the components i are represented by hypothetical ionic combinations ($A_1$, $A_2$, $S_1(Cl)$, $S_1(SO_4)$, $S_2(SO_4)$ etc.) with maximum number of possible coexistent components n = 6. After calculation of partial entropies for individual items according Eq. (4), internal RIE of the classified water is determined by (5) and (6). The internal RIE expresses the type purity (or in some case the degree of mixing) as continuously varying single variable. The value Hr = 1 expresses the maximum type indefiniteness (i. e. equal proportion of all possible components), while $H_r = 0$ characterises the maximum distinctness (type purity): a water containing 100% of the only component and zero portions of all other components. Beside the classification, the internal RIE can be used also in studying the relations of various variables relevant to genetic processes, mixing of waters or their spatial distribution. The external RIE where $p_i$ values represent the proportions of individual types in a set of waters (in an area) could be used e.g. to express the information indefiniteness in particular areas in a map.

Recenzoval: RNDr. Dušan Bodiš, PhD.