

# Chemometric approach to data processing in hydrogeological research

Jana Stojković, Marina Ćuk, Maja Todorović, Jovana Milosavljević, Jakov Andrijašević



Дигитални репозиторијум Рударско-геолошког факултета Универзитета у Београду

[ДР РГФ]

Chemometric approach to data processing in hydrogeological research | Jana Stojković, Marina Ćuk, Maja Todorović, Jovana Milosavljević, Jakov Andrijašević | XVI Kongres geologa Srbije, Gornji Milanovac, 2014 | 2014 | |

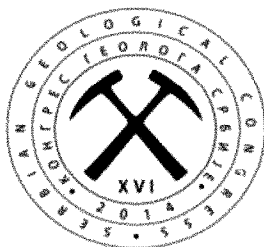
<http://dr.rgf.bg.ac.rs/s/repo/item/0005887>

Дигитални репозиторијум Рударско-геолошког факултета Универзитета у Београду омогућава приступ издањима Факултета и радовима запослених доступним у слободном приступу. - Претрага репозиторијума доступна је на [www.dr.rgf.bg.ac.rs](http://www.dr.rgf.bg.ac.rs)

The Digital repository of The University of Belgrade Faculty of Mining and Geology archives faculty publications available in open access, as well as the employees' publications. - The Repository is available at: [www.dr.rgf.bg.ac.rs](http://www.dr.rgf.bg.ac.rs)

**Српско геолошко друштво**

**Зборник радова  
XVI Конгреса геолога Србије**



**Proceedings  
of the XVI Serbian Geological Congress**

**ОПТИМАЛНО ИСТРАЖИВАЊЕ И ОДРЖИВО  
КОРИШЋЕЊЕ ГЕОЛОШКИХ РЕСУРСА**

**OPTIMAL RESEARCH AND SUSTAINABLE USAGE OF  
THE GEOLOGICAL RESOURCES**

**XVI Конгрес геолога Србије: Зборник радова**  
(Национални конгрес с међународним учешћем)

**XVI Serbian Geological Congress: Proceedings**  
(National Congress with International Participation)

**Donji Milanovac, 22-25.05.2014.**

**За издавача / For the Publisher**

Zoran Stevanović,

Председник Српског геолошког друштва / President of the Serbian Geological Society

**Главни уредник / Editor-in-Chief**

Vladica Cvetković

**Уређивачки одбор / Editorial Board**

Biljana Abolmasov, Katarina Bogičević, Meri Ganić, Rade Jelenković, Aleksandra Maran Stevanović, Vesna Matović, Vesna Ristić Vakanjac, Aleksandar Ristović, Ljupko Rundić

**Техничка припрема / Technical Preparation**

Veljko Marinović, Ognjen Jevtić, Milan Vukićević, Nemanja Krstekanić, Jelena Krstajić, Nikola Vojvodić, Bojan Kostić, Kristina Šarić

**Издавач / Publisher**

Српско геолошко друштво / Serbian Geological Society

Kamenička 6, P. Box 227, 11001 Belgrade, Serbia

<http://www.sgd.rs>; e-mail: [office@sgd.rs](mailto:office@sgd.rs)

Тираж: 300 примерака / Circulation: 300 copies

**Штампа / Printing**

Šprint, Alekse Nenadovića 28, Beograd, Serbia

ISBN 978-86-86053-14-5

**Напомена:** Аутори су одговорни за садржај и квалитет својих саопштења  
**Note:** The authors are responsible for the content and quality of their contributions

|  |                               |   |
|--|-------------------------------|---|
| XVI Конгрес геолога Србије, Доњи Милановац, 22-25 мај 2014.<br>XVI Serbian Geological Congress, Donji Milanovac, May 22 <sup>nd</sup> -25 <sup>th</sup> 2014 | Зборник радова<br>Proceedings | Проширени апстракт<br>Extended abstract |
|--|-------------------------------|---|

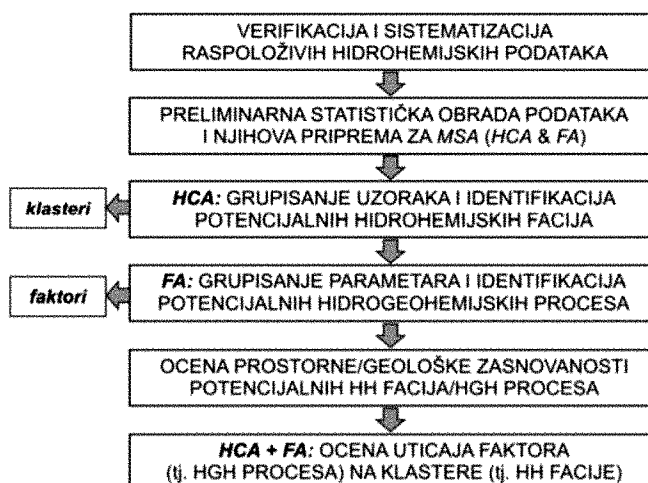
## HEMOMETRIJSKI PRISTUP OBRADI PODATAKA U HIDROGEOLOŠKIM ISTRAŽIVANJIMA

**Jana Štrbački<sup>1\*</sup>, Marina Ćuk<sup>1</sup>, Maja Todorović<sup>1</sup>, Jovana Milosavljević<sup>1</sup> i Jakov Andrijašević<sup>1</sup>**

<sup>1)</sup>Univerzitet u Beogradu, Rudarsko-geološki fakultet, Departman za hidrogeologiju, Đušina 7, Beograd, \*e-mail: jana.stojkovic@rgf.bg.ac.rs

**Ključne reči:** faktorna analiza, hidrogeohemijski procesi, hidrohemijske facije, klaster analiza, sekvencijalna metodologija

U ovom radu prikazan je tzv. hemometrijski pristup obradi hidrohemijskih podataka, prilikom rešavanja različitih hidrogeoloških problema. Suština ovakvog pristupa jeste u primeni multivarijantnih statističkih metoda, ali sekvencijalno, tj. određenim logičnim redosledom, tako da se svaki naredni korak nadovezuje na prethodni. Ovim se postiže sveobuhvatnije sagledavanje raspoloživih informacija i njihovo bolje i potpunije iskorišćavanje, a povećava se i pouzdanost dobijenih rezultata (Thyne et al., 2004). Tok razmatrane metodologije prikazan je na blok dijagramu (sl. 1).



Sl. 1 Blok dijagram sekvencijalne multivarijantne statističke metodologije. MSA – multivarijantna statistička analiza, HCA – hijerarhijska klaster analiza, FA – faktorna analiza.

Prvi korak razmatrane metodologije predstavlja verifikacija i sistematizacija raspoloživih hidrohemijskih podataka. Često je potrebno izuzeti iz dalje statističke obrade određeni broj uzoraka, odnosno fizičko-hemijskih parametara. Uzorci se najčešće odbacuju usled nekompletnosti hemijskih analiza ili u slučaju elektroneutralnosti (greške u jonskom balansu) veće od 5 %, dok jedan od glavnih razloga za izuzimanje pojedinih hidrohemijskih parametara predstavlja veliki broj vrednosti (preko 30 %) ispod nivoa detekcije (Güler et al., 2002). Takođe, neophodno je uskladiti merne jedinice, npr. sve koncentracije izraziti u mg/l.

Sledeći korak podrazumeva preliminarnu obradu podataka i njihovu pripremu za multivarijantnu statističku analizu. Najpre se određuju elementarne statističke veličine: aritmetička i geometrijska sredina, medijana, minimalna i maksimalna vrednost, koeficijent asimetrije itd. Ukoliko se ustanovi da pojedini parametri odstupaju od normalne raspodele (što je česta pojava kod hidrohemijskih veličina), potrebno je izvršiti transformaciju podataka, pri čemu najbolje rezultate daje Boks-Koks transformacija (Box & Cox, 1964). Normalnost transformisanih podataka moguće je

proveriti primenom Kolmogorov-Smirnov testa. Na kraju treba standardizovati analizirane parametre (izračunavanje tzv. z-vrednosti), kako bi svi imali jednaku težinu u daljoj statističkoj analizi (Güler et al., 2002).

Nakon formiranja konačnog skupa hidrohemijskih podataka pristupa se primeni multivarijantnih statističkih metoda. Klaster analiza (*Hierarchical Cluster Analysis, HCA*) i faktorna analiza (*Factor Analysis, FA*) imaju široku primenu u hidrohemiji. Suština obe metode jeste organizovanje i pojednostavljivanje velikih skupova podataka. HCA ima za cilj grupisanje uzoraka u tzv. **klaster**e, na osnovu sličnosti u hemijskom sastavu i fizičkim osobinama i to tako da su uzorci u okviru jednog klastera sličnih osobina, a da se razlikuju od uzoraka u drugim klasterima. Izdvojeni klasteri predstavljaju potencijalne hidrohemijske facije. S druge strane, FA služi za grupisanje višemedusobno korelisanih parametara u tzv. **faktore**, koji se mogu dovesti u vezu sa određenim hidrogeohemijskim procesima, značajnim za formiranje hemizma prirodnih voda (Cloutier et al., 2008, Helena et al., 1999, Drever 1997, Davis 1986).

Poslednji korak predstavlja verifikacija rezultata HCA i FA. Ovo podrazumeva, između ostalog, ocenu prostorne, odnosno geološke zasnovanosti dobijenih klastera, što se postiže njihovim nanošenjem na različite podloge (topografske, geološke, hidrogeološke i druge). Analizira se i uticaj izdvojenih hidrogeohemijskih procesa (tj. faktora) na razmatrane hidrohemijske facije (tj. klaster)e. Razdvajaju se faktori koji su posledica prirodnih procesa (interakcija stena – voda) od antropogene komponente (različiti izvori zagađenja) (Thyne et al., 2004).

Pored niza prednosti, opisana metodologija svakako ima i određena ograničenja. Kvalitet rezultata statističke analize zavisi od kvaliteta unetih podataka, a preciznost interpretacije uslovljena je veličinom ispitivanog područja. Takođe, pravilna interpretacija dobijenih rezultata moguća je samo u odgovarajućem geološkom, hidrogeološkom i hidrogeohemijskom kontekstu.

#### LITERATURA / REFERENCES:

- Box, G.E.P. & Cox, D.R. (1964): An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26 (2), 211-252, JSTOR, London.
- Cloutier, V., Lefebvre, R., Therrien, R. & Savard, M.M. (2008): Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *Journal of Hydrology*, 353, 294-313, Elsevier.
- Davis, J.C. (1986): *Statistics and Data Analysis in Geology*. New York: John Wiley & Sons Inc.
- Drever, J.I. (1997): *The geochemistry of natural waters*. 3rd. edition, Upper Saddle River, New Jersey: Prentice-Hall.
- Güler, C., Thyne, G.D., McCray, J.E. & Turner, A.K. (2002): Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*, 10, 455-474, Springer.
- Helena, B. A., Vega, M., Barrado, E., Pardo, R. & Fernandez, L. (1999): A case of hydrochemical characterization of an alluvial aquifer influenced by human activities. *Water, Air and Soil Pollution*, 112, 365-387, Springer.
- Thyne, G., Güler, C. & Poeter, E. (2004): Sequential Analysis of Hydrochemical Data for Watershed Characterization. *Groundwater*, 42 (5), 711-723, National Ground Water Association, Ohio, USA.