

PRIMJENA GIS-A I STATISTIČKIH METODA U USPOSTAVLJANJU POVEZNICA IZMEĐU GEOKEMIJSKIH KARAKTERISTIKA OKOLIŠA I GEOGRAFSKE DISTRIBUCIJE POBOLA STANOVNIŠTVA

APPLICATION OF GIS AND STATISTICAL METHODS IN ESTABLISHMENT OF THE RELATIONSHIPS BETWEEN GEOCHEMICAL CHARACTERISTICS OF THE ENVIRONMENT AND GEOGRAPHICAL DISTRIBUTION OF DISEASE INCIDENCE RATE

ANTE KUTLE¹, JASMINA OBHOĐAŠ², DARIO PERKOVIĆ³, JOSIP GREGAC¹, DIJANA PETROVIĆ⁴,
FRANKICA KAPOR³

¹Nevladina udruga „Lijepa naša“, Palmotićeveva 17a, Zagreb, Hrvatska

²Institut Ruđer Bošković, Bijenička c. 54, Zagreb, Hrvatska

³Rudarsko-geološko naftni fakultet Sveučilišta u Zagrebu, Pierottijeva 6, Zagreb, Hrvatska

⁴Centar „Vinko Bek“, Kušlanova 59a, Zagreb, Hrvatska

Ključne riječi: Geokemija okoliša, pobol stanovništva, GIS, statistika, otok Krk, delta Neretve

Key words: Environmental geochemistry, public health, GIS, statistics, Island of Krk, Neretva delta

Sažetak

Geokemijski okoliš može utjecati na zdravlje ljudi izazivajući kronične zdravstvene probleme vezane za dugotrajnu izloženost malim dozama toksičnih tvari kao što su primjerice neki elementi u tragovima s kojima dolazimo u dodir direktno inhalacijom prašine iz zraka ili indirektno kroz prehrambeni lanac te koristeći vodu iz lokalnih izvora za piće, kuhanje, osobnu higijenu i rekreaciju. Kronični zdravstveni problemi povezani s karakteristikama geokemijskog okoliša mogu nastati i zbog trajne izloženosti deficitu za ljude esencijalnih elementa. U ovom radu ćemo na nekoliko primjera prikazati na koji se način mogu neke geografske distribucije pobola dovesti u vezu s karakteristikama geokemijskog okoliša uz primjenu GIS-a i statističkih metoda. Također ćemo prikazati kako u geokemijskom okolišu možemo razlikovati prirodnu distribuciju elemenata od antropogenog doprinosa što je važno u primjeni potrebnih mjera za smanjivanje rizika po ljudsko zdravlje.

Abstract

Geochemical environment can influence human health causing chronic medical problems related to long-term, low-level exposures to toxic agents such as trace elements. Humans can be directly exposed to toxic substances by inhalation of air dust or indirectly through food chain or by consumption of local water for drinking, cooking, personal hygiene and recreational purposes. Chronic medical problems related to geochemical characteristics of the environment can also be caused by chronic deficit of chemical elements essential for humans. In this paper we will present several applications of the GIS and statistical methods for relating the geographical distribution of diseases with geochemical characteristics of the environment. In addition, we are presenting methods applied for distinguishing natural distribution of elements from anthropogenic contribution, which is important information for establishing protective measures necessary for decreasing the health risk.

Uvod

Studije utjecaja geokemijskog okoliša na zdravlje ljudi vrlo su kompleksne, uglavnom stoga što je potrebno prikupiti veliki broj georeferenciranih podataka iz različitih segmenata okoliša, a najčešće uključuju podatke o koncentracijama kemijskih supstanci u zraku, vodi za piće, tlu, sedimentu i bioti koji se uspoređuju s incidencijama bolesti (broj dijagnoza unutar jedne godine) u

stanovništvu, za koje se može pretpostaviti povezanost s dugotrajnim izlaganjem toksičnom učinku ili pak deficitu neke tvari iz okoliša. Krajnji cilj takvih studija je ustanoviti uzroke i izvore toksičnosti, područje i način njihova širenja te potrebne zaštitne mjere za smanjivanje izloženosti ljudi zdravstvenom riziku. Integracija i analiza tako velikog broja podataka bila bi praktički nemoguća bez upotrebe Geografskog informacijskog sustava (GIS) i statističkih metoda. Ove metode primijenjene zajedno

omogućuju međusobno uspoređivanje podataka različitih izvora i kvalitete te otkrivanje odstupajućih vrijednosti i klastera prikrivenih velikim brojem podataka, pridružujući im pritom i prostornu dimenziju

Postoji velik broj bolesti za koje je dokazano da su povezane s geokemijskim obilježjima okoliša (Valković, 1980; Medical Geology, 2010). Većina studija o uzročno-posljedičnim vezama između geokemije okoliša i pobola stanovništva započinje s identificiranjem područja povećanih incidencija bolesti (Cao et al., 2004; Dissanayake, and Chandrajith, 1999; Masironi, 1987; Tan et al., 2002). Uobičajena praksa je da se u razmatranje uzimaju administrativne jedinice poput općina i gradova uglavnom stoga što se zdravstveni podatci prikupljaju prema tim područjima (Cromley and McLafferty, 2002). Također treba voditi računa o tome da se u obzir uzmu čimbenici poput migracije stanovništva, prehrane, socijalnog statusa i sl.

S obzirom da je ovo područje geokemije kod nas još prilično nepoznato, u ovom ćemo radu prikazati neke rezultate istraživanja poduzetih posljednjih godina na području otoka Krka i delte Neretve. Pri tome će naglasak biti na prikazu primijenjenih GIS i statističkih metoda koje se mogu koristiti i za druge aplikacije u geokemiji i zaštiti okoliša uz diskusiju o mogućim zamkama na koje se može naići prilikom njihova korištenja, a koje se ponekad previde kao što se može primijetiti u nekim znanstvenim publikacijama. Od statističkih metoda opisane su metode deskriptivne statistike, analiza varijance (Kruskal-Wallis test), klaster analiza, linearna regresijska analiza, faktorska analiza glavnih komponenti i analiza glavnih komponenti između grupa. Primjera sličnih istraživanja u kojima se koriste različite statističke metode za povezivanje oboljenja s drugim parametrima ima mnogo. Mather et al. (2004) i Barr et al. (2006) daju pregled statističkih metoda koje se mogu koristiti za evaluiranje rizika po zdravlje uslijed izloženosti zagađivačima iz okoliša. Spominju se deskriptivna analiza, korelacijske metode i metode redukcije. Palesh et al. (2007) koristi Kruskal-Wallis analizu u dokazivanju povezanosti raka dojke i izloženosti velikom stresu ili traumi, dok Sun et al. (2011) koristi klaster analizu za otkrivanje veza između oboljenja i genskih klastera. Naglasak u ovom članku je na praktičnoj primjeni umjesto teorijskih objašnjenja metoda s krajnjim ciljem jednostavne interpretacije dobivenih rezultata.

Metode

Podaci o pobolu otoka Krka za razdoblje 1997.-2001. prikazani u ovom radu prikupljeni su putem Zavoda za javno zdravstvo grada Rijeke. Podaci o pobolu delte Neretve za razdoblje 1999.-2008. prikupljeni su od Hrvatskog zavoda za javno zdravstvo. **Analize geokemijskih uzoraka** napravljene su u Laboratoriju za nuklearne analitičke metode Instituta Ruđer Bošković. Prikazani su rezultati analize vode za piće na otoku Krku te rezultati analiza tla i sedimenta na području delte Neretve. Na otoku Krku go-

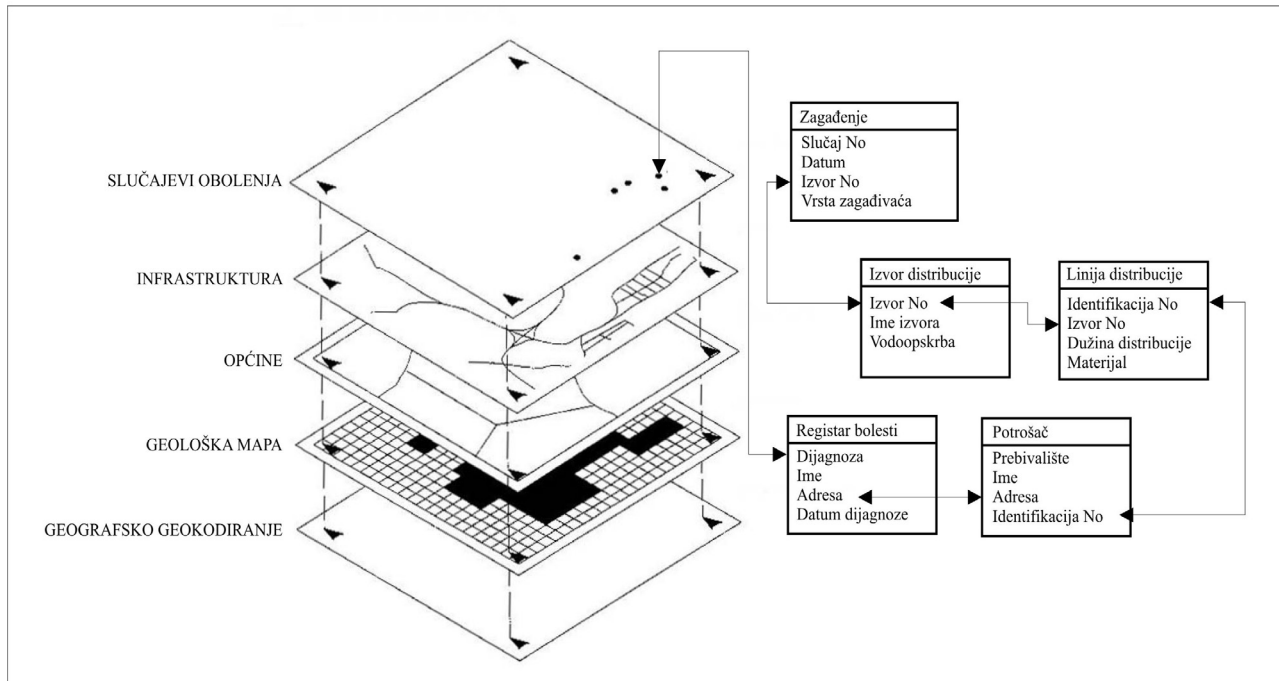
dine 1999. ukupno je sakupljeno 83 uzoraka iz nasumično odabranih kućanstava na način da su bila zastupljena sva naselja. Izmjerene su koncentracije elemenata Mn, Fe, Ni, Cu, Zn, As **metodom energetske disperzivne fluorescencije x-zraka** (engl. «Energy Dispersive X-ray Fluorescence», EDXRF) na tankim metama dobivenim filtriranjem 100 ml uzorka koji je prethodno prekoncentriran s APDC kelatom pri pH = 3.5 i uz 20 min. miješanja. Za EDXRF analizu korištena je rendgenska pobuda (W anoda) i Mo sekundarna meta, a spektri su izmjereni pomoću Si(Li) detektora (FWHM = 175 eV za 5.9 keV) u pravokutnoj geometriji. **Za obradu spektara korišten je QXAS program, metoda direktne usporedbe, a kao referentni materijali korišteni su Flukini standardi koncentracije 1 g/L.** Sedimenti i tlo na području delte Neretve također su izmjereni opisanom EDXRF metodom, ali na debelim metama. Mjereni su elementi K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Ga, As, Br, Rb, Sr, Y, Zr i Pb na ukupno 318 uzoraka sakupljenih u razdoblju 2008.-2011. Analize su napravljene na frakciji $\phi < 2$ mm. Korištena je metoda direktne usporedbe i IAEA referentni materijali S7 i SL-1. Detaljniji opis metoda i rezultati mjerenja objavljeni su u literaturi (Valković et al., 2001; Kutle, 2004; Kutle et al., 2004 i 2006; Obhodas et al., 2007 i 2011).

GIS, pobol i geokemijska analiza okoliša

GIS je jednostavno rečeno bilo koji skup informacija povezan s prostorom. Nezamjenjiv je u radu s velikim brojem informacija koje moraju biti brzo dostupne i vizualno jasne. Potencijalno velika primjenjivost GIS-a u javnom zdravstvu je prepoznata i o tome postoje mnogobrojne publikacije (Atkinson i Unwin, 2002; Buckeridge et al., 2002; Koch i Denike, 2001).

Na slici 1. je primjer kako se mogu povezati slučajevi akutnog oboljenja s nekim incidentom zagađenja. Pobol povezan s dugotrajnom izloženosti malim koncentracijama toksičnih tvari je puno teže ustanoviti.

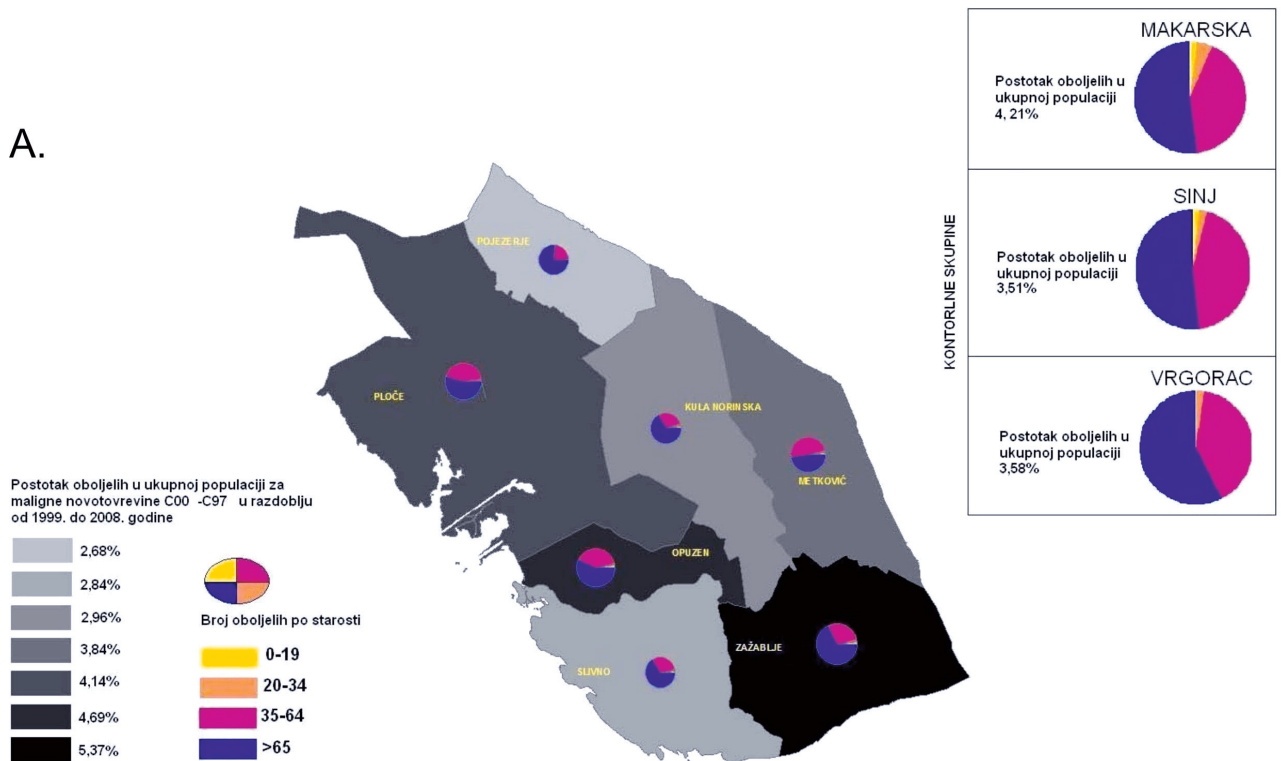
Na slici 2. je prikazana GIS analiza pobola za ukupna sijela raka (C00.0-C97.0), maligne novotvorine debelog crijeva (C18.0-C21.8) i maligne novotvorine mjehura (C67.0-C67.9) za administrativne jedinice u delti Neretve. U tablici 1 prikazani su podatci za ukupan broj stanovništva prema popisu stanovništva 2001. i koeficijent starosti (udio starijih od 65 godina) za administrativne jedinice na području delte Neretve. Primjećuje se neobična distribucija za maligne novotvorine mjehura koja se ne može dovesti u vezu sa sociološkim faktorima kao što su starosna dob ili ruralno-gradski način života što ukazuje na potrebu razmatranja geokemije okoliša kao mogućeg etiološkog faktora bolesti.

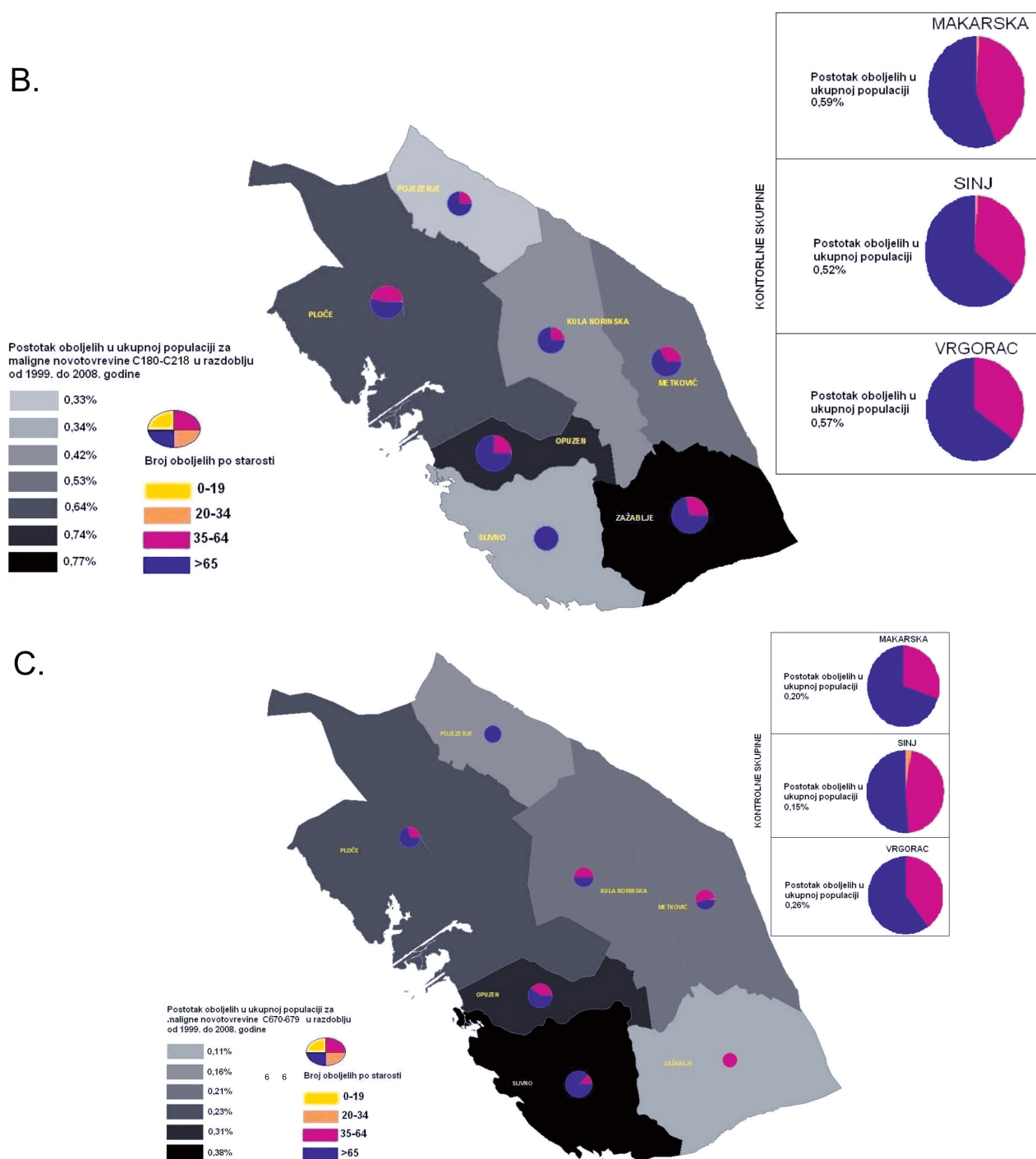


Slika 1. Shematski prikaz preklapanja tematskih mapa i pridruživanje baze podataka koje je moguće povezati preko zajedničkih polja. Primjer kako možemo dovesti slučajeve akutnog oboljenja u vezu s nekim incidentom zagađenja (prema Cromley and McLafferty, 2002)

Figure 1. A schematic presentation of the thematic maps overlapping and joining of the data bases by using mutual fields. An example of relating the acute disease cases to a pollution incident (after Cromley and McLafferty, 2002)

A.





Slika 2. Prikaz GIS analize učestalosti pobola među stanovništvom delte Neretve za razdoblje 1999.-2008. A. ukupno sijelo raka (C00.0-C97.0), B. maligne novotvorine debelog crijeva (C18.0-C21.8) i C. maligne novotvorine mjehura (C67.0-C67.9)

Figure 2. Presentation of the GIS analysis of disease incidence rates in the population of the Neretva delta for the period 1999-2008. A. Malignant neoplasms (C00.0-C97.0), B. Malignant neoplasms of colon (C18.0-C21.8), C. Malignant neoplasms of bladder (C67.0-67.9)

Identificiranje područja povećane incidencije i neobične distribucije bolesti može biti povod za provođenje opsežnih geokemijskih istraživanja, međutim, potrebno je ispitati da li su uočene razlike u pobolu između administrativnih jedinica statistički značajne. Na isti je način

potrebno analizirati i prikupljene geokemijske podatke. Ako se primjerice mjerene koncentracije elemenata u trgovima iz okoliša statistički značajno ne razlikuju za analizirane administrativne jedinice, tada uzrok treba tražiti drugdje.

Tablica 1. Ukupan broj stanovništva prema popisu 2001. i koeficijent starosti (udio starijih od 65 godina) za administrativne jedinice na području delte Neretve

Table 1. Population in municipalities of the Neretva delta, Census 2001, and age coefficient (rate of population above age of 65)

| Administrativne jedinice | Broj stanovnika | Koeficijent starosti (%) |
|--------------------------|-----------------|--------------------------|
| Metković | 15384 | 10.90 |
| Ploče | 10834 | 13.24 |
| Opuzen | 3242 | 14.84 |
| Kula Norinska | 1926 | 17.60 |
| Slivno | 2078 | 20.84 |
| Zažablje | 912 | 21.49 |
| Pojezerje | 1233 | 15.70 |

Deskriptivna statistika

Prije poduzimanja statističke obrade podataka specifičnog cilja poput utvrđivanja statističke značajnosti među unaprijed zadanim grupama, uobičajeno se provodi deskriptivna (osnovna) statistička analiza. Ona, kao što joj ime govori, opisuje neki skup podataka procjenjujući centralne tendencije i varijabilnost. Deskriptivna statistika se koristi za utvrđivanje odstupanja od uniformnosti ili za određivanje parametara sa svrhom usporedbe različitih skupova podataka. Pri tome treba imati na umu da geokemijska mjerenja često daju podatke koji odudaraju od normalne raspodjele. Ti su podatci uglavnom pozitivno asimetrični (engl. positively skewed data), a ponekad i jako pozitivno asimetrični (engl. highly skewed data). Naime, u prirodnom uzorku podataka o okolišu nema negativnih vrijednosti koncentracija, većina ih ima niske vrijednosti uz pojavu značajno odstupajućih vrijednosti (engl. outliers), odnosno vrijednosti koje bitno odstupaju od prosjeka. U takvim slučajevima, značajno odstupajuće vrijednosti snažno utječu na srednju vrijednost (\bar{X}) i standardnu devijaciju (SD) te manjak ili dodatak čak i samo jedne takve vrijednosti uzrokuje lošu procjenu mjera centralnih tendencija i varijabilnosti. Za pozitivno asimetrične raspodjele podataka, iznad srednje vrijednosti može se

nalaziti manje od 50%, ili čak manje od 25% podataka. U tom slučaju srednja vrijednost nije dobra procjena podataka. Isto tako, odstupajuće će vrijednosti precijeniti standardnu devijaciju, uzrokujući veću varijabilnost nego što pokazuje većina dobivenih podataka.

Alternativne mjere za centralnu tendenciju i varijabilnost za asimetrične raspodjele podataka su medijan i interkvartilni raspon (IQR). Prema definiciji, medijan je vrijednost u sredini serije, odnosno postoji 50% izmjerenih vrijednosti iznad i 50% ispod medijana. Odstupajuće vrijednosti ne utječu snažno na medijan kao što utječu na srednju vrijednost, pa prema tome medijan predstavlja bolju procjenu za asimetrične raspodjele podataka. Za simetrične raspodjele, medijan je jednak srednjoj vrijednosti. Iz istih razloga često se računa geometrijska srednja vrijednost. To je srednja vrijednost logaritama izračunatih iz izmjerenih podataka i predstavlja procjenu medijana (za nepreračunatu jedinicu).

Kao ni medijan, ni IQR nije pod utjecajem najnižih i najviših vrijednosti. Izračunava se kao interval između vrijednosti koja predstavlja 75% podataka (gornji kvartil) i vrijednosti koja predstavlja 25% podataka (donji kvartil) te obuhvaća 50%-ni raspon podataka. IQR je jednak 135% standardne devijacije za normalnu razdiobu. Za asimetrične raspodjele IQR je često znatno manji od standardne devijacije i predstavlja bolju procjenu varijabilnosti. Često se koristi i za opis raspodjele koju čini mali broj podataka (<30).

Da bi se usporedile varijacije različitih varijabli uvodi se koeficijent varijacije (CV) koji se računa kao $CV = SD/x_{sr}$, odnosno kao standardna devijacija u odnosu na srednju vrijednost. Ako je koeficijent varijacije jednak ili manji od pogreške mjerenja tada govorimo o homogenom uzorku.

U slučaju da je distribucija podataka asimetrična odnosno podaci odstupaju od normalne raspodjele potrebno je pažljivo pristupiti daljnjoj statističkoj obradi podataka jer se neke statističke metode u tom slučaju ne mogu primijeniti. Podatke je tada preporučljivo transformirati nekom od metoda logaritmiranja ili primijeniti neparametarske statističke metode. U tablici 2 prikazana je osnovna statistika za izmjerene koncentracije elemenata u vodi otoka Krka koja se koristi za piće.

Tablica 2. Osnovna statistika za izmjerene koncentracije elemenata u vodi otoka Krka koja se koristi za piće ($\mu\text{g/L}$)

Table 2. Descriptive statistics for analyzed concentrations of elements in drinking water of the Island of Krk ($\mu\text{g/L}$)

| Element | Srednja vrijednost | Median | Minimum | Maksimum | Standardna devijacija | Asimetričnost |
|---------|--------------------|--------|---------|----------|-----------------------|---------------|
| Mn | 125.8 | 96.0 | 0 | 863.0 | 139.4 | 2.9 |
| Fe | 99.4 | 45.0 | 4 | 1252.0 | 191.5 | 4.6 |
| Ni | 63.7 | 36.0 | 1 | 444.0 | 81.6 | 2.5 |
| Cu | 14.2 | 8.0 | 0 | 147.0 | 21.2 | 4.0 |
| Zn | 773.1 | 239.0 | 2 | 6880.1 | 1427.3 | 2.9 |
| As | 1.0 | 0.8 | 0 | 7.0 | 1.2 | 1.9 |

Tablica 3. Osnovna statistika za izmjerene koncentracije elemenata u tlima i sedimentima na području delte Neretve**Table 3.** Descriptive statistics for analyzed concentrations of elements in the Neretva delta soil/sediment data-set

| | Valjani broj uzoraka | Srednja vrijednost | Median | Minimum | Maksimum | Donji kvartil | Gornji kvartil | 95% percentil | Standardna devijacija | Asimetričnost | Zaobljenost |
|---------|----------------------|--------------------|----------|---------|----------|---------------|----------------|---------------|-----------------------|---------------|-------------|
| K mg/L | 313 | 14412.00 | 16090.00 | 1276.00 | 34120.00 | 10700.00 | 18880.00 | 21550.00 | 6157.000 | -0.47 | -0.26 |
| Ca % | 319 | 13.10 | 9.10 | 1.00 | 58.50 | 6.40 | 14.20 | 44.30 | 11.500 | 2.05 | 3.76 |
| Ti mg/L | 318 | 2899.00 | 2974.00 | 109.00 | 8587.00 | 1749.00 | 4023.00 | 5252.00 | 1658.000 | 0.45 | 0.56 |
| V mg/L | 309 | 72.80 | 72.80 | 5.90 | 264.00 | 47.50 | 92.80 | 128.00 | 39.400 | 0.96 | 2.62 |
| Cr mg/L | 165 | 107.20 | 79.00 | 12.80 | 2962.00 | 55.70 | 109.00 | 203.00 | 229.000 | 11.97 | 149.85 |
| Mn mg/L | 317 | 576.00 | 569.00 | 22.30 | 2399.00 | 249.00 | 804.00 | 1168.00 | 386.000 | 1.14 | 2.86 |
| Fe % | 318 | 2.85 | 2.94 | 0.03 | 7.77 | 1.72 | 3.85 | 5.33 | 1.553 | 0.23 | -0.09 |
| Ni mg/L | 308 | 63.80 | 63.80 | 9.70 | 269.00 | 39.90 | 79.80 | 117.00 | 33.700 | 1.58 | 6.95 |
| Cu mg/L | 318 | 37.00 | 26.70 | 8.10 | 273.00 | 19.60 | 39.50 | 93.30 | 33.100 | 3.43 | 14.90 |
| Zn mg/L | 317 | 119.00 | 114.00 | 21.00 | 672.00 | 89.10 | 138.00 | 203.00 | 64.400 | 3.39 | 21.85 |
| Ga mg/L | 279 | 12.20 | 11.50 | 1.20 | 52.80 | 8.20 | 15.30 | 22.00 | 5.700 | 1.62 | 8.26 |
| As mg/L | 316 | 9.20 | 9.10 | 0.70 | 69.10 | 6.60 | 11.00 | 17.20 | 5.500 | 4.33 | 44.71 |
| Br mg/L | 318 | 425.00 | 316.00 | 4.90 | 4367.00 | 90.00 | 646.00 | 1127.00 | 484.900 | 3.66 | 23.32 |
| Rb mg/L | 316 | 90.30 | 95.40 | 7.90 | 190.00 | 66.40 | 120.00 | 150.00 | 41.000 | -0.34 | -0.62 |
| Sr mg/L | 316 | 216.00 | 199.00 | 78.10 | 1556.00 | 168.50 | 245.00 | 325.00 | 121.400 | 7.27 | 70.69 |
| Y mg/L | 318 | 31.70 | 31.60 | 2.00 | 73.10 | 23.50 | 40.10 | 57.40 | 14.700 | 0.11 | -0.24 |
| Zr mg/L | 313 | 168.00 | 168.00 | 1.10 | 507.00 | 123.00 | 212.00 | 309.00 | 81.100 | 0.84 | 2.10 |
| Pb mg/L | 319 | 29.10 | 27.70 | 1.00 | 359.00 | 16.90 | 35.80 | 60.00 | 24.500 | 8.02 | 103.19 |

U tablici 3 prikazana je osnovna statistika za izmjerene koncentracije elemenata u tlima i sedimentima na području delte Neretve.

Analiza Varijance

Nakon analize uniformnosti podataka možemo pristupiti analizi varijance što je statistička metoda koja se primjenjuje kad se želi utvrditi postoje li značajne razlike u varijabilnosti između grupe podataka. Tim se postupkom ispituje da li je varijabilitet između grupa na temelju kojih su izračunate pojedinačne aritmetičke sredine, značajno veći od varijabiliteta unutar grupa.

Treba imati na umu da se takva analiza pomoću ANOVA, koja se najčešće koristi za tu svrhu, smije primijeniti samo za grupe čija je raspodjela podataka normalna. U slučaju da su raspodjele asimetrične, podatke je potrebno transformirati (obično izrazom $\ln(x+1)$) odnosno približiti normalnoj raspodjeli ili koristiti neparametarsku ANOVA, primjerice Kruskal-Wallis test.

Ako je varijabilitet među grupama značajno veći od varijabiliteta unutar grupa, onda su to grupe koje u statističkom smislu ne pripadaju istoj razdiobi. Kao rezultat analize varijance dobiju se F vrijednosti koje označuju omjer varijance između grupa (MS_{ig}) i varijance unutar grupa (MS_{ug}):

$$F = MS_{ig} / MS_{ug}$$

Ako je izračunata F vrijednost za odgovarajući broj stupnjeva slobode i razinu značajnosti (na primjer $p=0,05$) veća od teorijske F vrijednosti, onda se ne prihvaća nulta hipoteza (H_0) o nepostojanju razlika među varijancama

pretpostavljenih grupa i obrnuto, ako je izračunata vrijednost manja od teorijske, onda se H_0 prihvaća.

Da bismo provjerili da li se administrativne jedinice statistički značajno razlikuju s obzirom na incidencije bolesti C00.0-C97.0, C18.0-C21.8 i C67.0-C67.9 za razdoblje 1999.-2008. upotrijebljena je neparametarska ANOVA - Kruskal-Wallis (KW). Broj oboljelih u administrativnim jedinicama delte Neretve na tisuću stanovnika i rezultati KW analize dati su u tablici 4. S obzirom da su za incidencije bolesti C00.0-C97.0 i C67.0-C67.9 vrijednosti $p < 0,05$, usvaja se hipoteza da te incidencije pobola ne pripadaju istoj distribuciji. Za maligne novotvorine C18.0-C21.8 treba usvojiti hipotezu da pretpostavljene grupe (incidencije bolesti prema administrativnim jedinicama) pripadaju jedinstvenoj raspodjeli.

Klaster analiza

Klaster analiza daje vizualni smisao procjenama povezanosti multidimenzionalnih točaka. Prvi stupanj klaster analize je generiranje različitosti matriksa (odnosno udaljenosti), izračunavanjem koeficijenta različitosti između svih parova uzoraka. Koristi se mnoštvo algoritama za dobivanje koeficijenta, a najčešći su Euklidska udaljenost, kvadratna Euklidska udaljenost, Manhattan udaljenost, Chebychev metrična udaljenost, potencija, postotak neslaganja, 1-Pearson-ov r, itd.

U početnom koraku analize pretpostavljamo da svaka točka (uzorak) označuje zaseban klaster. Izborom algoritma računamo udaljenost između dva klastera, da bi se dva najbliža klastera spojila u jedan, novi klaster.

Tablica 4. Incidencija bolesti C00.0-C97.0, C18.0-C21.8 i C67.0-C67.9 (normalizirano na 1000 stanovnika) za razdoblje 1999.-2008. i analiza varijance između administrativnih jedinica delte Neretve

Table 4. Disease incidence rate for disease groups C00.0-C97.0, C18.0-C21.8 and C67.0-C67.9 (normalized to 1000 citizens) for the period 1999-2008 and results of the Variance analysis for the Neretva delta municipalities

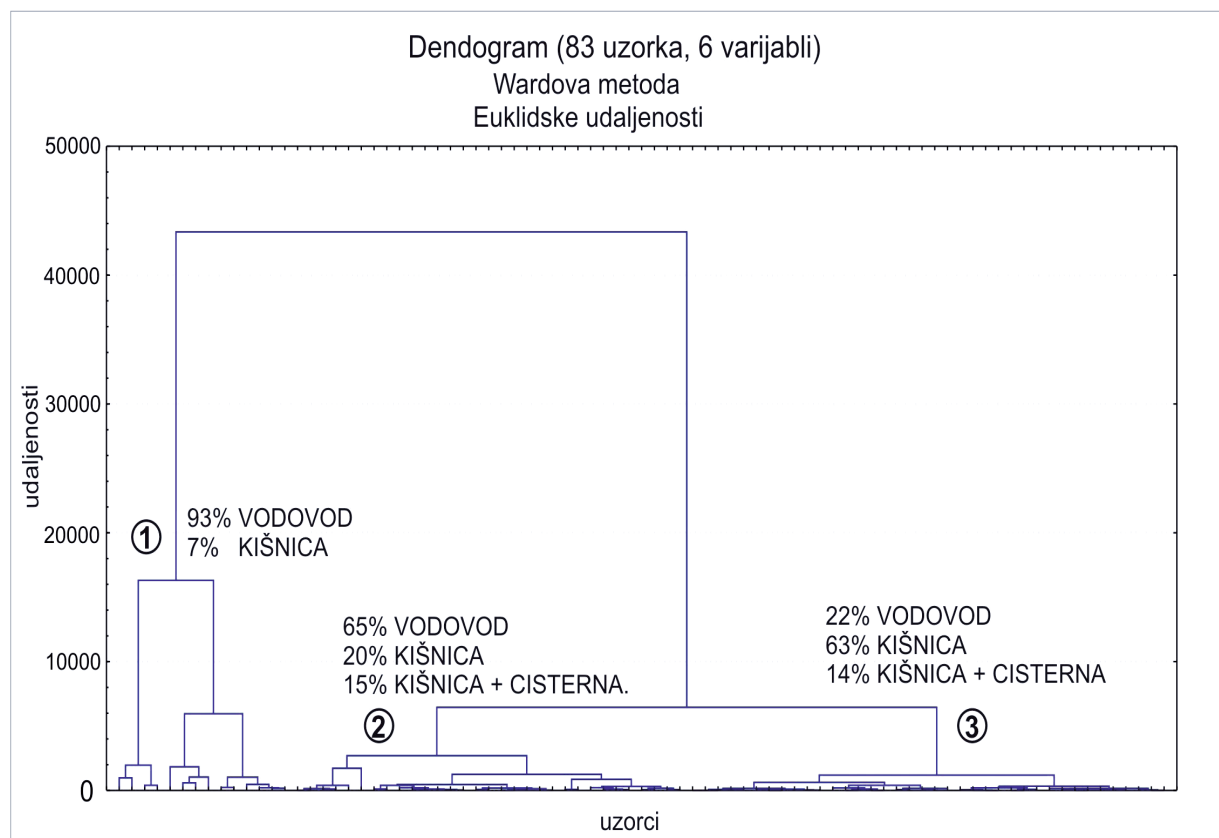
| Maligne novotvorine (C00-C99) H= 20,84 p = 0,002 | | | | | | | |
|--|----------|----------|----------|----------|----------|----------|-----------|
| Godina | Metković | Ploče | Opuzen | Kula N. | Slivno | Zažablje | Pojezerje |
| 1999 | 4.225169 | 4.245893 | 6.785935 | 2.596054 | 2.406160 | 7.675439 | 0.81103 |
| 2000 | 3.575143 | 3.968987 | 4.009870 | 1.038422 | 4.812320 | 5.482456 | 0.81103 |
| 2001 | 3.380135 | 3.507476 | 2.467613 | 1.557632 | 2.406160 | 7.675439 | 5.67721 |
| 2002 | 3.315133 | 4.892007 | 3.701419 | 4.672897 | 3.368624 | 8.771930 | 4.05515 |
| 2003 | 3.575143 | 4.615101 | 3.392967 | 4.672897 | 1.443696 | 2.192982 | 1.62206 |
| 2004 | 3.575143 | 3.876685 | 3.084516 | 2.076843 | 3.368624 | 5.482456 | 0.81103 |
| 2005 | 4.940198 | 3.415174 | 5.860580 | 3.115265 | 2.887392 | 2.192982 | 2.43309 |
| 2006 | 3.835153 | 4.615101 | 5.860580 | 2.596054 | 0.962464 | 4.385965 | 3.24412 |
| 2007 | 3.770151 | 3.968987 | 4.935225 | 3.634476 | 3.368624 | 6.578947 | 4.05515 |
| 2008 | 4.225169 | 4.338195 | 6.785935 | 3.634476 | 3.368624 | 3.289474 | 3.24412 |
| Maligne novotvorine (C18.0-C21.8) H=7,43 p=0,283 | | | | | | | |
| 1999 | 0.455018 | 0.646114 | 0.925355 | 0 | 0 | 3.289474 | 0 |
| 2000 | 0.325013 | 0.461510 | 1.233806 | 0 | 0.481232 | 1.096491 | 0.81103 |
| 2001 | 0.260010 | 0.461510 | 0.308452 | 0 | 0.481232 | 2.192982 | 0.81103 |
| 2002 | 0.455018 | 0.923020 | 0.308452 | 0.519211 | 0.481232 | 0 | 0.81103 |
| 2003 | 0.715029 | 0.646114 | 0.308452 | 1.038422 | 0 | 0 | 0.81103 |
| 2004 | 0.520021 | 0.646114 | 0.925355 | 0 | 0.962464 | 1.096491 | 0 |
| 2005 | 0.845034 | 0.646114 | 2.159161 | 0.519211 | 0 | 0 | 0 |
| 2006 | 0.455018 | 0.738416 | 0.308452 | 0 | 0 | 0 | 0 |
| 2007 | 0.585023 | 0.461510 | 0.308452 | 0 | 0 | 0 | 0 |
| 2008 | 0.715029 | 0.738416 | 0.616903 | 2.076843 | 0.962464 | 0 | 0 |
| Maligne novotvorine (C67.0-C67.9) H=13,53 p=0,035 | | | | | | | |
| 1999 | 0.260010 | 0.276906 | 0 | 0 | 0.481232 | 0 | 0 |
| 2000 | 0.325013 | 0.184604 | 0.308452 | 0 | 0.481232 | 0 | 0 |
| 2001 | 0.455018 | 0.184604 | 0.616903 | 0 | 0 | 0 | 0 |
| 2002 | 0.065003 | 0.646114 | 0.308452 | 0.519211 | 0.962464 | 1.096491 | 0 |
| 2003 | 0.130005 | 0.276906 | 0.308452 | 0.519211 | 0 | 0 | 0 |
| 2004 | 0.260010 | 0 | 0.308452 | 0 | 0 | 0 | 0 |
| 2005 | 0.130005 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2006 | 0.260010 | 0.276906 | 0.616903 | 0 | 0.481232 | 0 | 0 |
| 2007 | 0.130005 | 0.092302 | 0.308452 | 1.038422 | 0.962464 | 0 | 0.81103 |
| 2008 | 0.130005 | 0.369208 | 0.308452 | 0 | 0.481232 | 0 | 0.81103 |

To rezultira povezivanjem sve većeg i većeg broja točaka u sve veće klustere sve dok ne ostane samo jedan. Kod obrade podataka koji potječu iz različitih populacija (grupa) analizom će se dobiti jasno definirani klasteri u obliku dendograma. Na slici 3. prikazana je klaster analiza za uzorke vode za piće na otoku Krku. U analizi je korištena Euklidska udaljenost za izračunavanje razlika (udaljenosti) između uzoraka. Varijable uzete u obzir su

Mn, Fe, Ni, Cu, Zn, As. Na otoku Krku kao i na brojnim drugim našim otocima u kućanstvima se još uvijek u značajnoj mjeri koristi kišnica. Klaster analiza jasno pokazuje 3 grupe podataka određene tipom vode. U prvoj grupi nalaze se uzorci koje uglavnom čini vodovodna voda, u drugoj grupi su pomiješani tipovi vode (primjerice gusterne koje se pune kišnicom i vodovodnom vodom), a u trećoj su uzorci koje većinom čini kišnica. S obzirom da su

tri tipa vode jasno grupirana, u daljnjem koraku smo pomoću regresijske analize ispitali da li korištenje različitog tipa vode ima utjecaj na zdravlje otočnog stanovništva. Naime, još iz 1950-tih je poznato da regije u kojima se u kućanstvima troši meka voda imaju povećanu incidenciju

i smrtnost od bolesti krvožilnog sustava (I00-I99), (Kobayashi 1957; Schroeder, 1960a, 1960b i 1969). Uzrok je najvjerojatnije u tome što meka voda u sebi sadrži znatno manju količinu otopljenih soli, poglavito kalcija i magnezija.



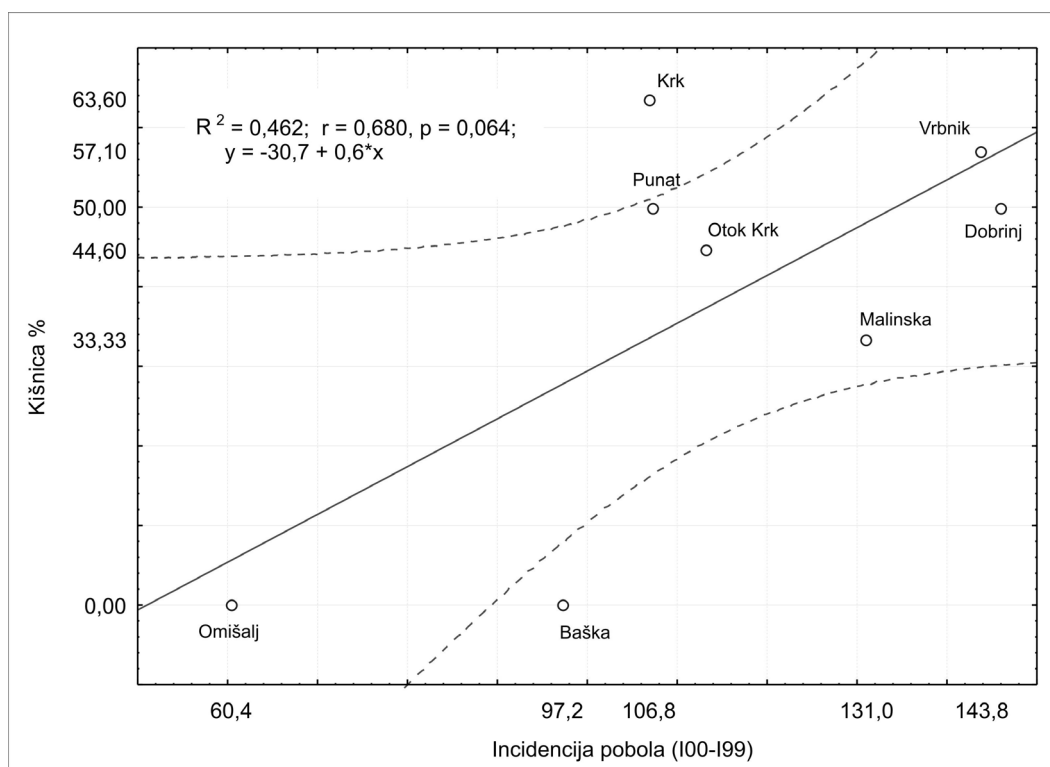
Slika 3. Klaster analiza za uzorke vode za piće na otoku Krku. Ukupan broj uzoraka = 83, Ward metoda, Euklidska udaljenost (Kutle et al., 2006)

Figure 3. Cluster analysis of drinking water samples collected at the Island of Krk. Total number of samples = 83, Ward method, Euclidean distance (Kutle et al., 2006)

Regresijska analiza

Regresijska analiza uključuje tehnike u kojima je fokus na funkcijskoj vezi između dvije ili više varijabli. Mogu se ispitivati razne regresijske funkcije pri čemu se teži onoj funkciji koja može dovesti u vezu dvije distribucije s čim manje parametara. Broj parametara određuje i broj stupnjeva slobode. U linearnoj regresijskoj analizi određuju se primjerice parametri a i b koji se odnose na nagib i odsječak linearne funkcije ($Y=aX+b$), a broj stupnjeva slobode je $N-2$ gdje je N ukupan broj podataka. Regresijskom analizom se određuje još i Pearson-ov korelacijski koeficijent (r) koji je indikator međusobne

ovisnosti dviju distribucija te koeficijent determinacije (R^2) koji procjenjuje frakciju opisane varijance u linearno regresijskim funkcijama. Statistička značajnost korelacije je različita za različite nivoe pouzdanosti (pri čemu se najčešće uzimaju nivoi 0,90, 0,95 ili 0,99) te za različiti broj uzoraka. Na slici 4 prikazana je regresijska analiza za distribuciju bolesti krvožilnog sustava (I00-I99) na otoku Krku i udjela kišnice u upotrebi vode u kućanstvima. Iako na granici statističke značajnosti, regresijska analiza ukazuje na moguću povezanost deficita elemenata Ca-a i Mg-a u populaciji i povećane incidencije bolesti krvožilnog sustava.



Slika 4. Linearna regresijska analiza za incidenciju bolesti krvožilnog sustava (100-199) na 1000 stanovnika za administrativne jedinice otoka Krka, nivo konfidencije 0,95. Za $N=8$ i $N-2$ stupnja slobode kritična vrijednost za $r = 0,621$. Koeficijent starosti (udio stanovništva starijeg od 65 god.) prema administrativnim jedinicama: Omišalj 12,8%, Malinska 23,9 %, Krk 21,4%, Dobrinj 32,7%, Vrbnik 31,9%, Punat 26,8%, Baška 31,1%, otok Krk 25,8%.

Figure 4. Linear regression analysis of the circulatory system diseases incidence rates (100-199) for the municipalities of the Island of Krk, confidence level 0.95. Statistical significance critical value of r for $N=8$ and $N-2$ degrees of freedom is 0.621. Age coefficient (rate of population above age of 65) after municipalities: Omišalj 12,8%, Malinska 23,9 %, Krk 21,4%, Dobrinj 32,7%, Vrbnik 31,9%, Punat 26,8%, Baška 31,1%, otok Krk 25,8%.

Faktorska analiza i analiza glavnih komponenti

Faktorska analiza (eng. Factor Analysis, FA) i analiza glavnih komponenta (eng. Principal Component Analysis, PCA) su najčešće korištene statističke metode za smanjenje obima podataka, maksimalno zadržavajući što je više moguće varijabilnosti podataka.

Za razliku od FA, PCA je metoda koja kombinira uzorke tako da traži nekolerirane linearne transformacije zadržavajući maksimalnu varijancu bez unaprijed zadanog modela. Kod PCA prva glavna komponenta predstavlja jednu linearnu kombinaciju originalnih varijabli koja opisuje najveći dio varijance podataka (odnosno prolazi kroz podatke s najmanjom kvadratnom greškom). U tom smislu je prva glavna komponenta PCA metode najinformativnija. Svaka sljedeća komponenta opisuje sve manji sadržaj informacija u podacima. Kod FA to nije nužno tako (Jolliffe, 2002).

Obje metode se vrlo često koriste za odvajanje prirodnih i antropogenih izvora varijacija u okolišu. Često se prvo provodi PC analiza kako bi se odredile glavne komponentne osi koje definiraju podprostor zadržavajući većinu varijance početne matrice i potom FA da bi se redefinirale osi ortogonalnom rotacijom ("Varimax raw"

metoda) ili nekim drugim modelom rotacije kako bi dobili faktore jednostavnije za tumačenje. Ova vrsta analize je široko dostupna u komercijalnim programskim paketima, ali se često pogrešno referencira kao FA odnosno PCA jer je u biti kombinacija ovih dviju metoda.

Broj osnovnih komponenti (PC) podvrgnut rotaciji najčešće se izabire u skladu s Kaiser-ovim pravilom prema kojemu se izabiru faktori koji objašnjavaju varijancu veću od 1 jer nema smisla interpretirati faktore koji objašnjavaju varijancu manju od izvorne varijable. Treba napomenuti da je za PCA, koja se vrlo često pogrešno poistovjećuje s FA, razina od 0,7 puno primjerenija (Jolliffe, 2002).

U mnogim primjerima koji se mogu pronaći u literaturi, rezultirajuće FA faktore ili PCA komponente lako je tumačiti, no ponekad je interpretacija nejasna. Nemogućnost tumačenja faktora može ukazivati na nepravilno korištenje metoda ili na lošu kvalitetu podataka. Poteškoće u interpretaciji faktora su najčešće kod skupova podataka sa statistički nedovoljnim brojem redaka (uzoraka, slučajeva) za zadani broj varijabli. Kao opće pravilo trebalo bi biti najmanje 10 redaka po varijabli, no preporuča se 20 i više. Analiza s 5 redova po varijabli neće biti relevantna. To se događa stoga što su kanonske jednadžbe i odgovara-

jući koeficijenti u pojedinim multivarijantnim statističkim metodama (FA, PCA, diskriminantna analiza, višestruka regresijska analiza, itd.) nestabilne ako je broj statističkih jedinica nedovoljan s obzirom na varijable. Za razliku od većine multivarijantnih statističkih metoda, uvijek je moguće pokrenuti analizu glavnih komponenti između grupa (eng. Between Group – Principal Component Analysis, BG-PCA) za bilo koji broj varijabli, skupina ili podataka prema grupi (ADE-4, 1997).

Neki autori koriste transformaciju podataka (normalizacija unosa, Box-Cox, Gaussian Anamorphosis Modelling i sl.) prije primjene multivarijantne statistike kako bi dobili normalno distribuirane podatke (Kribek et al., 2010., Sollitto et al., 2010). Normalizacija podataka ili bilo kakve eksplicitne distribucijske pretpostavke zapravo su nepotrebne. Naprotiv, normalno distribuirani podatci manje su interesantni s obzirom da su odstupanja od normalnosti indikatori skrivenih struktura unutar podataka, kao što su klasteri ili odstupajuće vrijednosti (engl. outliers). Asimetrija i zaobljenost su parametri opisne statistike koji se obično koriste za utvrđivanje odstupanja od uniformnosti podataka. Oni se mogu koristiti ako je potrebno izabrati podskup varijabli zbog stabilnosti kanonskih jednadžbi. Transformacija podataka prije primjene FA ili PCA su ipak neophodne ako su varijable izmjerene na različitim skalama, a međusobno se uspoređuju. Ukoliko se koristi korelacijska matrica, podatci će biti normalizirani na sljedeći način:

$$(x_{ij} - m_j) / s_j$$

gdje je x_{ij} - i vrijednost u stupcu j , m_j - srednja vrijednost j stupca, a s_j - standardno odstupanje od j stupca.

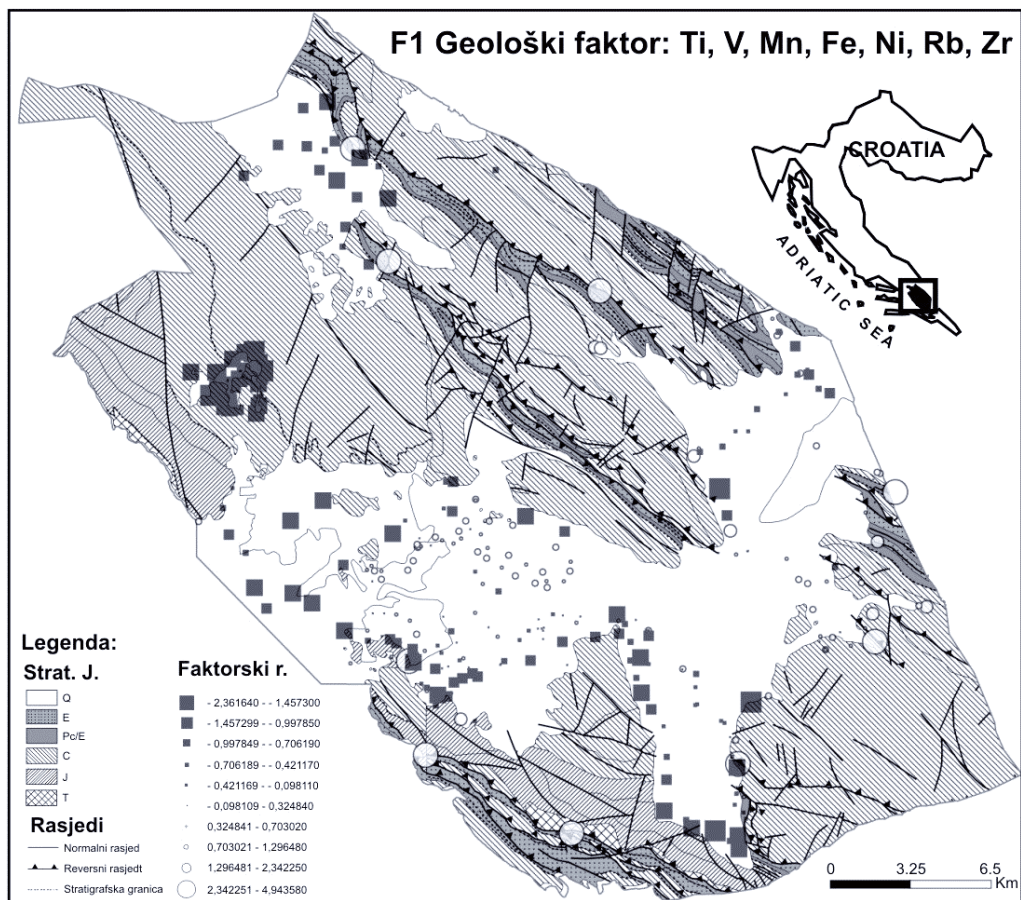
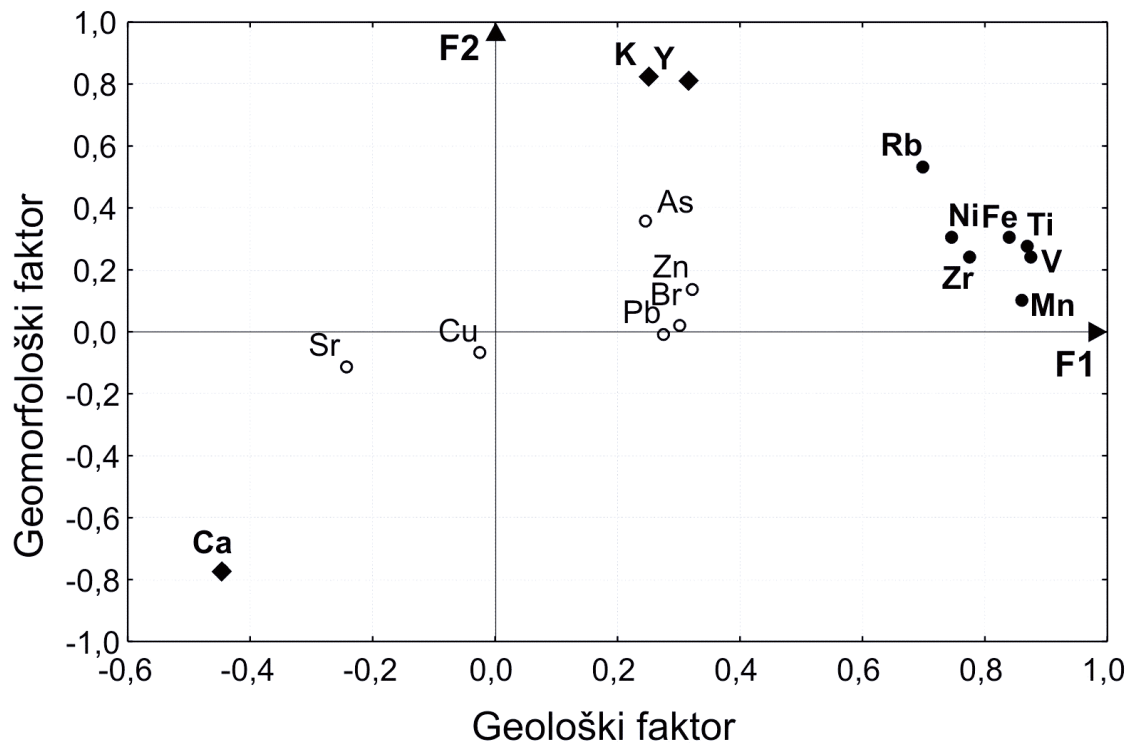
Normalizacija daje svakoj varijabli srednju vrijednost nule i varijancu jednaku jedinici. Za varijable izmjerene na istoj skali, analiza se može provesti na kovarijantnoj matrici. Potrebno je napomenuti da će kovarijantne matrice dati veću težinu varijablama koje sadrže veće vrijednosti s obzirom da je njihova varijanca veća. Dodatne transformacije podataka su neophodne i za vrijednosti koje nedostaju unutar analizirane matrice što je vrlo česta pojava u velikom skupovima podataka o okolišu (vrijednosti ispod granice detekcije, analitičke pogreške, izgubljeni uzorci, itd.). Obično se bira između opcije zanemarivanja redaka koji sadrže vrijednosti koje nedostaju, što je zadovoljavajuće ukoliko nedostaje mali broj vrijednosti ili opcije zamjene sa srednjim vrijednostima, što nije sasvim zadovoljavajuće za vrijednosti ispod minimalne granice detekcije. Stoga je za usporedbe različitih analiza iznimno važna informacija o tome da li se je koristila korelacijska ili kovarijantna matrica i kako su tretirane vrijednosti koje nedostaju.

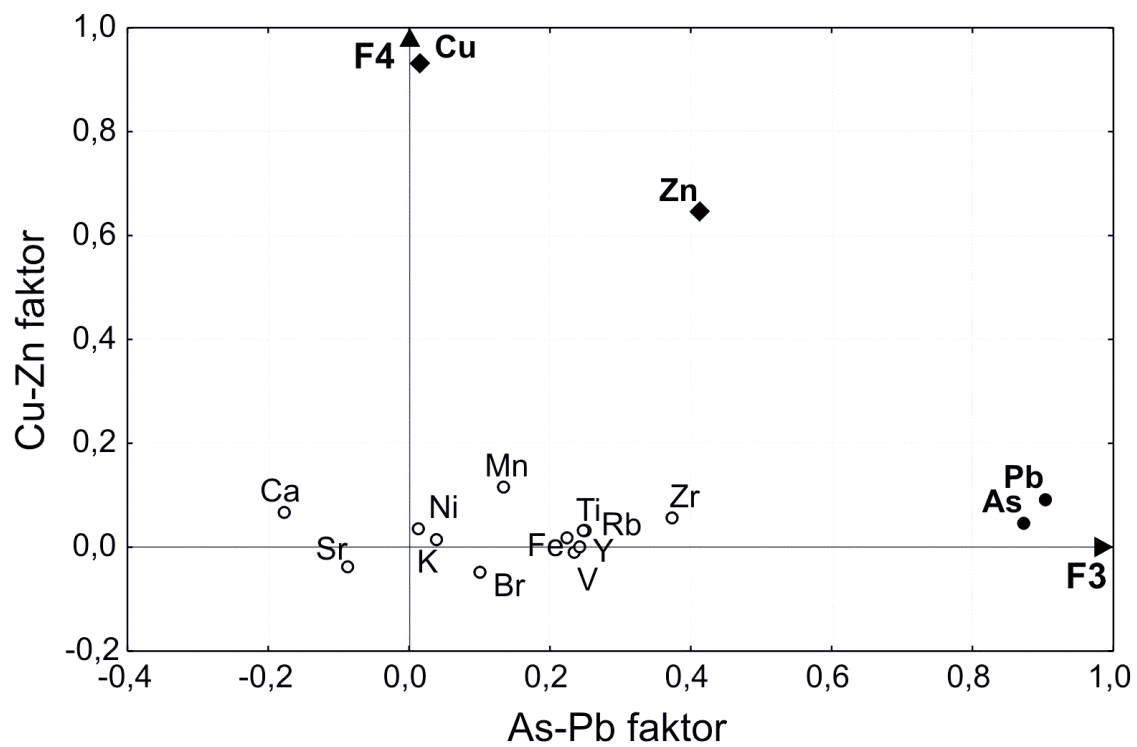
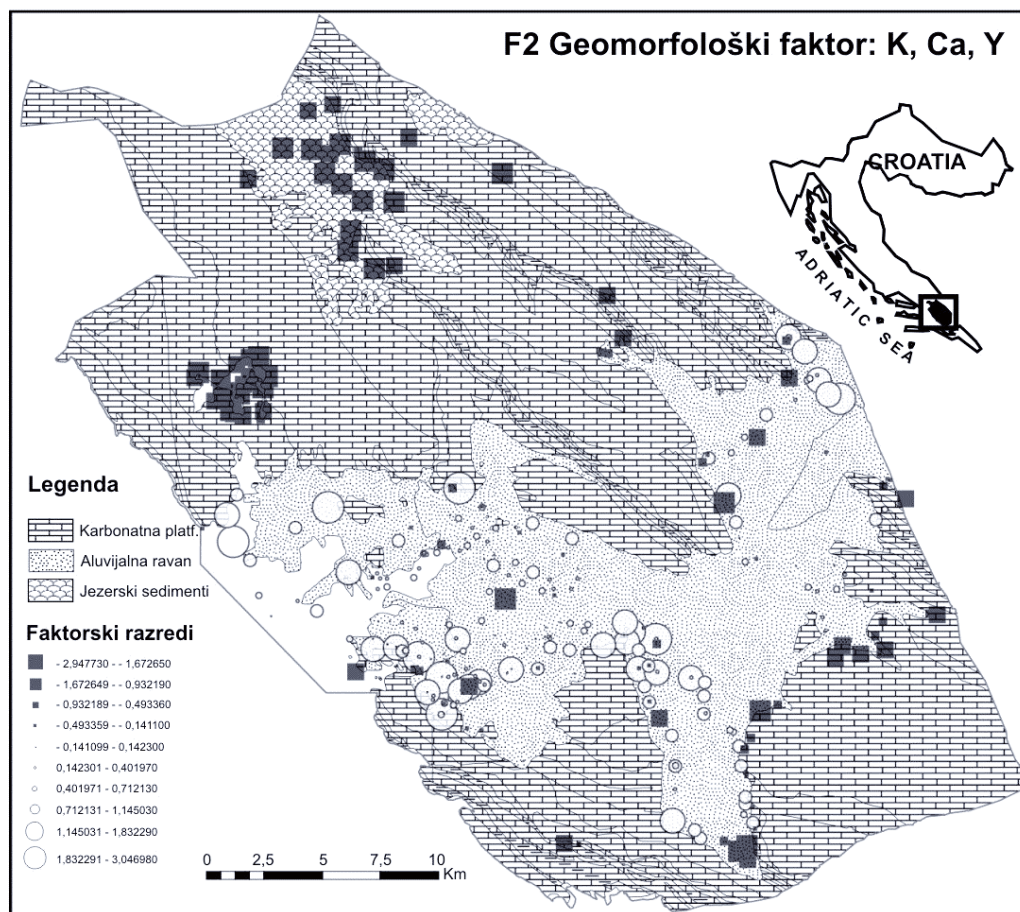
U tablici 5 prikazani su rezultati faktorske analize glavnih komponenti za sedimente i tlo delte Neretve, a na slici 5 rezultirajuće faktorske karte koje su dobivene projekcijom graduiranih vrijednosti faktorskih doprinosa za pojedinačne uzorke na različitim tematskim regionalnim kartama. Tematska karta koja najbolje opisuje i vizualizira prostorne odnose između klastera ili odstupajućih vrijednosti odabrana je kao pozadinska karta za prikaz faktorskih karata. Vizualizacijom prostornih odnosa odvajaju se prirodne raspodjele koncentracije elemenata u tlu i sedimentu od raspodjela koje sadrže antropogene doprinose.

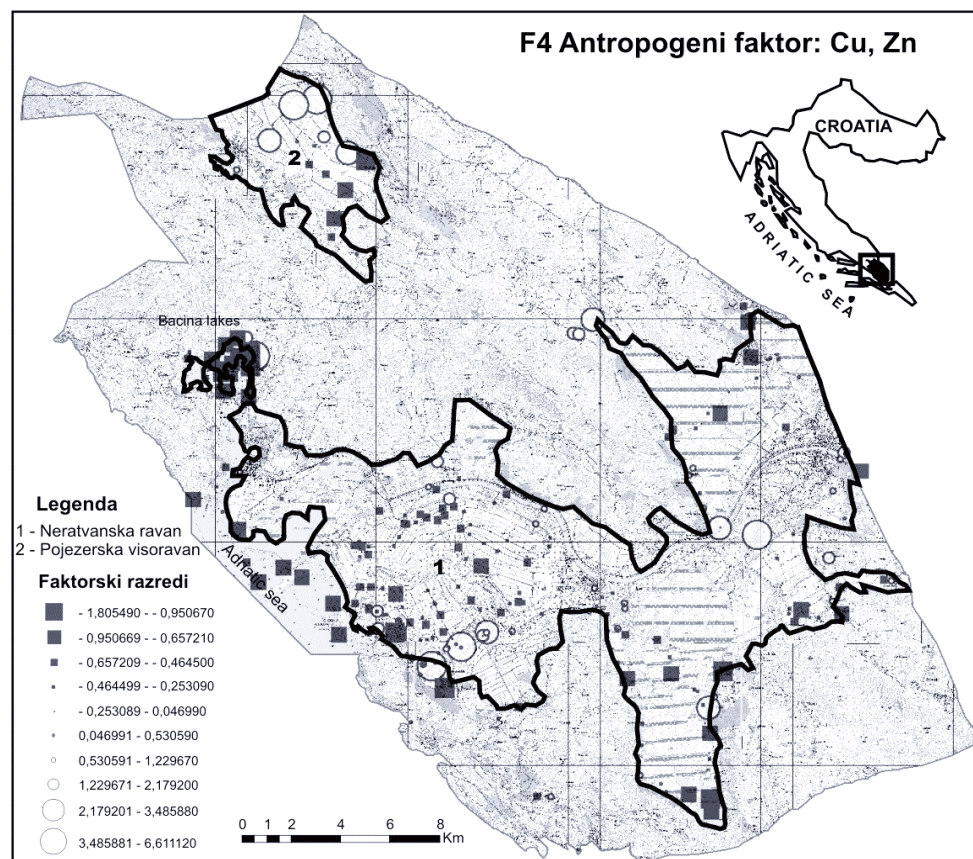
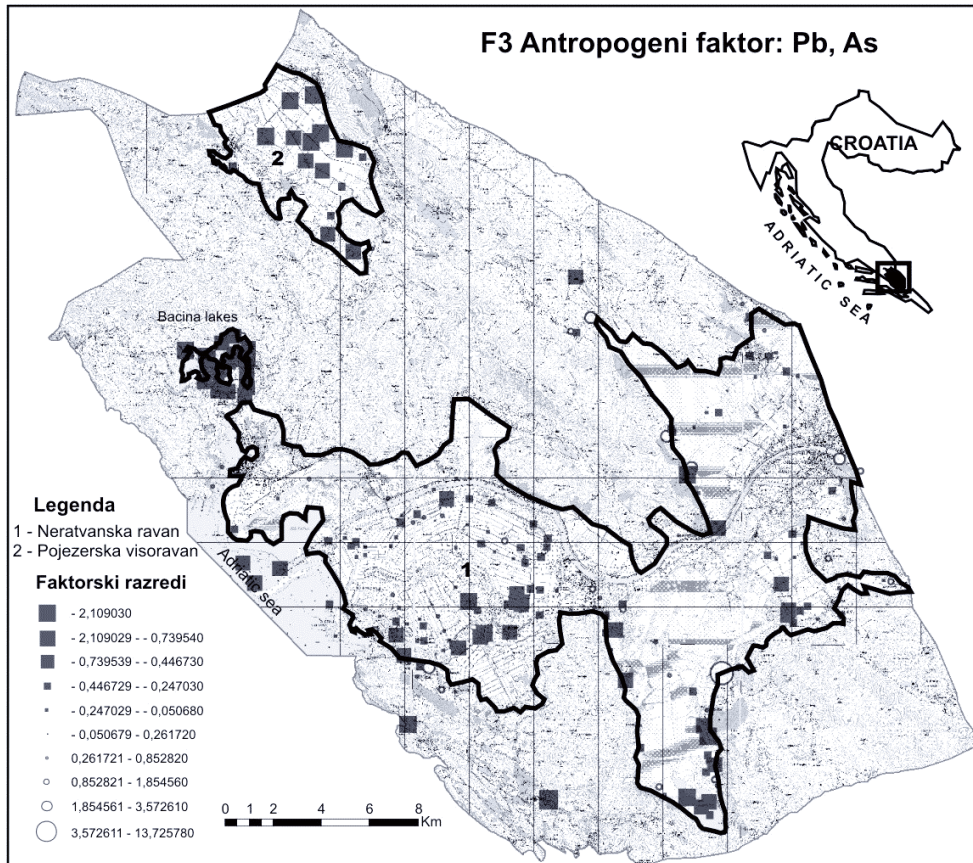
Tablica 5. Faktori i doprinosi varijabli nakon početne ekstrakcije glavnih osi. „Varimax raw“ rotacija (podcrtana opterećenja >0,65).

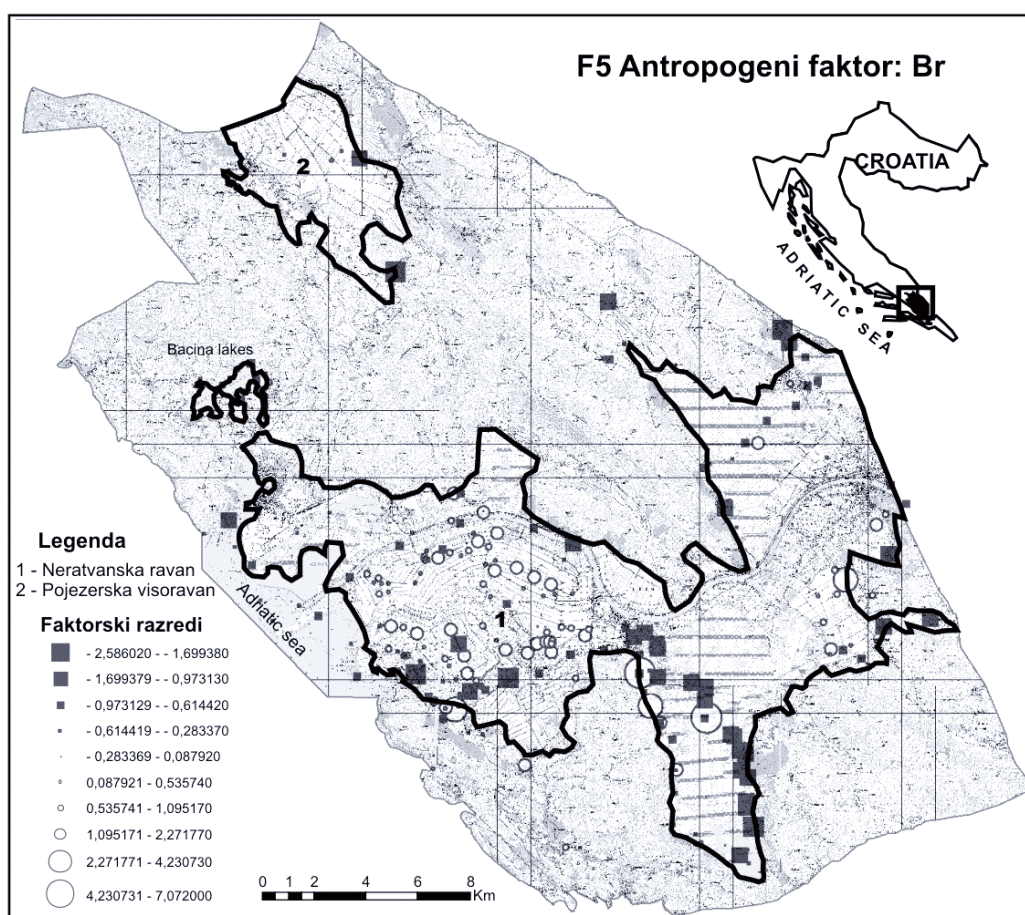
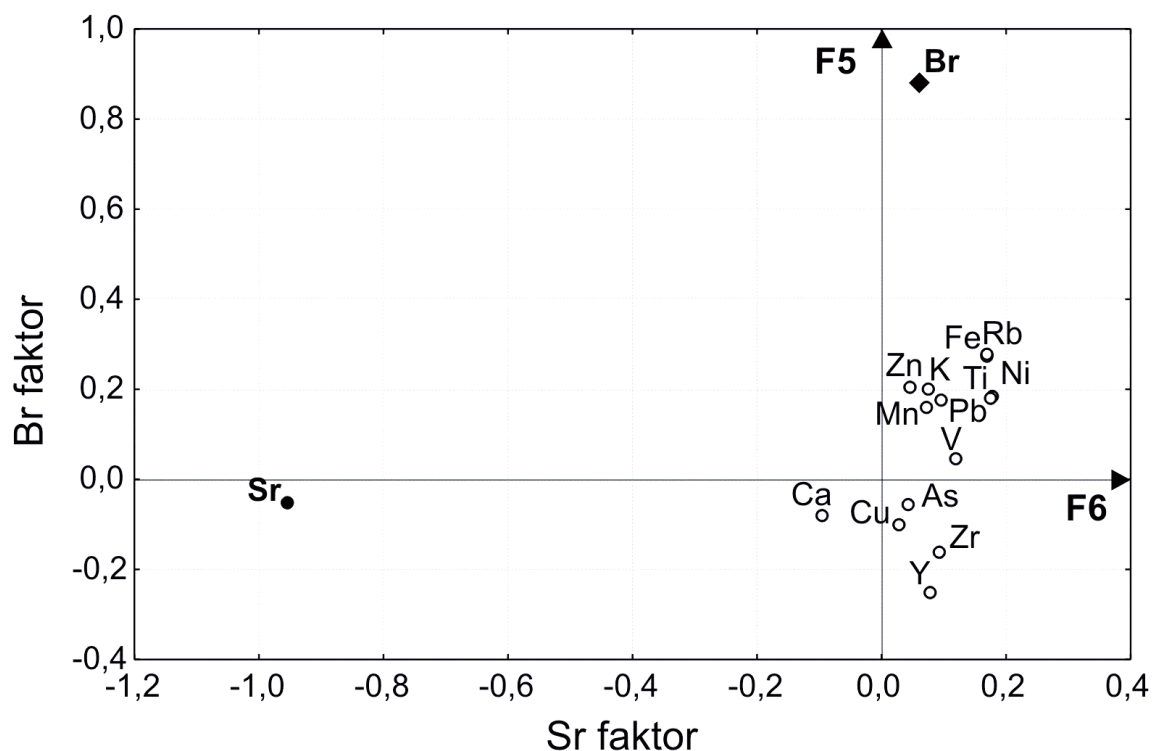
Table 5. Factors and their variable loadings after initial extraction of the Principal component axes. Method: Varimax raw rotation (marked loadings > 0,65)

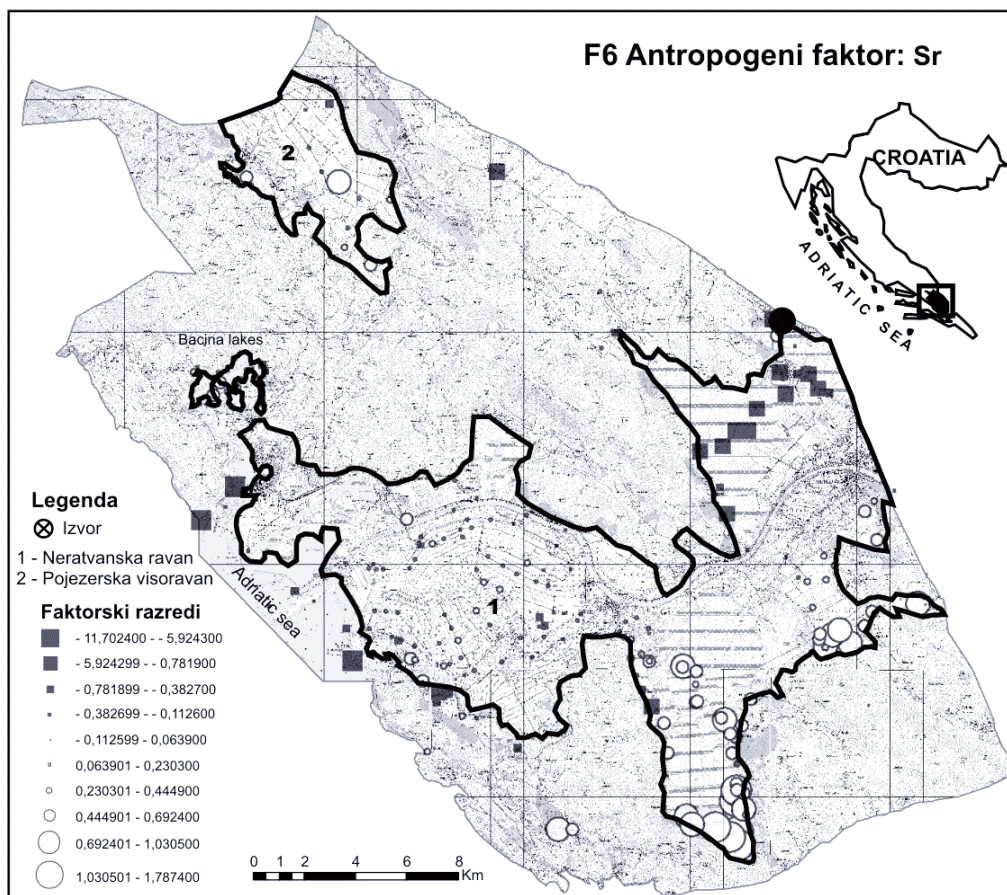
| | Prirodni faktori | | Antropogeni faktori | | | |
|---------------------------------|------------------|----------------------|---------------------|-----------------|-----------------|-----------------------|
| | F1 Geologija | F2 Geomorfologija | F3 As-Pb | F4 Cu-Zn | F5 Br | F6 Sr salinizacija |
| K ppm | 0.250057 | <u>0.827567</u> | 0.037246 | 0.016545 | 0.203204 | 0.074443 |
| Ca % | -0.447404 | <u>-0.765408</u> | -0.179414 | 0.068249 | -0.079138 | -0.098129 |
| Ti ppm | <u>0.868161</u> | 0.277202 | 0.249716 | 0.033036 | 0.184203 | 0.175727 |
| V ppm | <u>0.873175</u> | 0.245512 | 0.233964 | -0.007725 | 0.048427 | 0.116961 |
| Mn ppm | <u>0.857565</u> | 0.103339 | 0.133327 | 0.118031 | 0.163298 | 0.069928 |
| Fe ppm | <u>0.839289</u> | 0.306635 | 0.223176 | 0.020101 | 0.275599 | 0.168885 |
| Ni ppm | <u>0.745227</u> | 0.310440 | 0.011136 | 0.037378 | 0.184003 | 0.174833 |
| Cu ppm | -0.026331 | -0.064280 | 0.015012 | <u>0.935468</u> | -0.098602 | 0.026247 |
| Zn ppm | 0.320957 | 0.140885 | 0.411088 | <u>0.646391</u> | 0.206121 | 0.044477 |
| As ppm | 0.244437 | 0.362508 | <u>0.871728</u> | 0.047477 | -0.053757 | 0.039851 |
| Br ppm | 0.301037 | 0.024912 | 0.100029 | -0.046609 | <u>0.884129</u> | 0.058984 |
| Rb ppm | <u>0.695921</u> | 0.534659 | 0.246951 | 0.034488 | 0.278348 | 0.167108 |
| Sr ppm | -0.242697 | -0.112325 | -0.087312 | -0.035627 | -0.048812 | <u>-0.955704</u> |
| Y ppm | 0.315115 | <u>0.813603</u> | 0.241751 | 0.003667 | -0.251257 | 0.077237 |
| Zr ppm | <u>0.772456</u> | 0.245099 | 0.372137 | 0.056569 | -0.159982 | 0.089843 |
| Pb ppm | 0.274635 | -0.006544 | <u>0.901796</u> | 0.094780 | 0.177353 | 0.094612 |
| objašnjena varijanca | 5.342721 | 2.785467 | 2.236289 | 1.333819 | 1.257129 | 1.095728 |
| objašnjena varijanca (%) | 33.39 | 17.41 | 13.98 | 8.34 | 7.86 | 6.85 |











Slika 5. Doprinos varijabli pojedinim faktorima i faktorske karte (Obhodas et al., 2011)

Figure 5. Variables contribution to factors and factorial maps (Obhodas et al., 2011)

Faktori F1 i F2 opisuju najviše od ukupne varijabilnosti podataka (50,8%), a uvjetovani su geološkim i geomorfološkim obilježjima prostora. Povećane koncentracije Ti, V, Mn, Fe, Ni, Rb i Zr mogu se povezati sa stijenama paleogenske starosti u sinklinalnim strukturama dok su niže koncentracije vezane za kredne vapnence. Deltu Neretve u užem smislu obilježavaju povećane koncentracije K i Y za razliku od karbonatne platforme koja je obilježena povećanim koncentracijama Ca. Analiza je pokazala i četiri antropogena izvora koji ukupno opisuju 37,03 % varijance. F3, F4 i F5 se mogu dovesti u vezu s intenzivnom agrarnom djelatnosti u delti Neretve, a faktor F6 sa salinizacijom dubokog tla što upućuje na pretjerano crpljenje vode iz vodoopskrbnog izvora Norin. Faktori F3, F4 i F5 se mogu objasniti upotrebom insekticida, fungicida i pesticida na bazi As-Pb, Cu-Zn i Br. Nekoliko izrazito povećanih vrijednosti u prikazu As-Pb faktora može biti indicacija recentne upotrebe ovog inače zabranjenog zaštitnog sredstva. Cu i Zn su tipični sastojci Bordeaux mješavine koja se i danas uobičajeno koristi za zaštitu vinograda i voćnjaka. Klaster povećanih vrijednosti Cu i Zn-a nalazimo na pojezerskoj visoravni i na nekoliko

lokacija u delti Neretve. Zaštitna sredstva na bazi broma se koriste u zaštiti agruma. Pozitivne vrijednosti koje pokazuju antropogeni doprinos ovog elementa uglavnom pokazuju umjerenu upotrebu ovog sredstva uz nekoliko visokih, odstupajućih vrijednosti. Kod faktora F6, visoke negativne vrijednosti faktorskih doprinosa ukazuju na povećane koncentracija stroncija koji je u sediment rijeke Norin dospio najvjerojatnije kapilarnim izdizanjem morske vode iz dubokog tla. Klasteri visokih vrijednosti identični su za morske sedimente iz obalnog područja i sedimente rijeke Norin, u kojima su koncentracije stroncija i 10 puta veće nego što je to tipično za kopneni okoliš.

Analiza glavnih komponenti između grupa (BG-PCA)

BG-PCA je metoda koja omogućava ispitivanje pripadnosti uzoraka unaprijed zadanim grupama, a koristi se nakon bilo koje analize vlastitih vrijednosti da bi se ustanovila značajnost varijacija među grupama. BG-PCA metodi se daje prednost kad se razmatra velik broj varijabli. Metoda je neosjetljiva kako na broj varijabli, tako i na broj grupa i broj mjerenja unutar grupe. BG-PCA je u na-

čelu slična diskriminantnoj analizi (DA) uz tu razliku što se DA provodi uz generalno pravilo 20 uzoraka (svakako ne manje od 5) po varijabli.

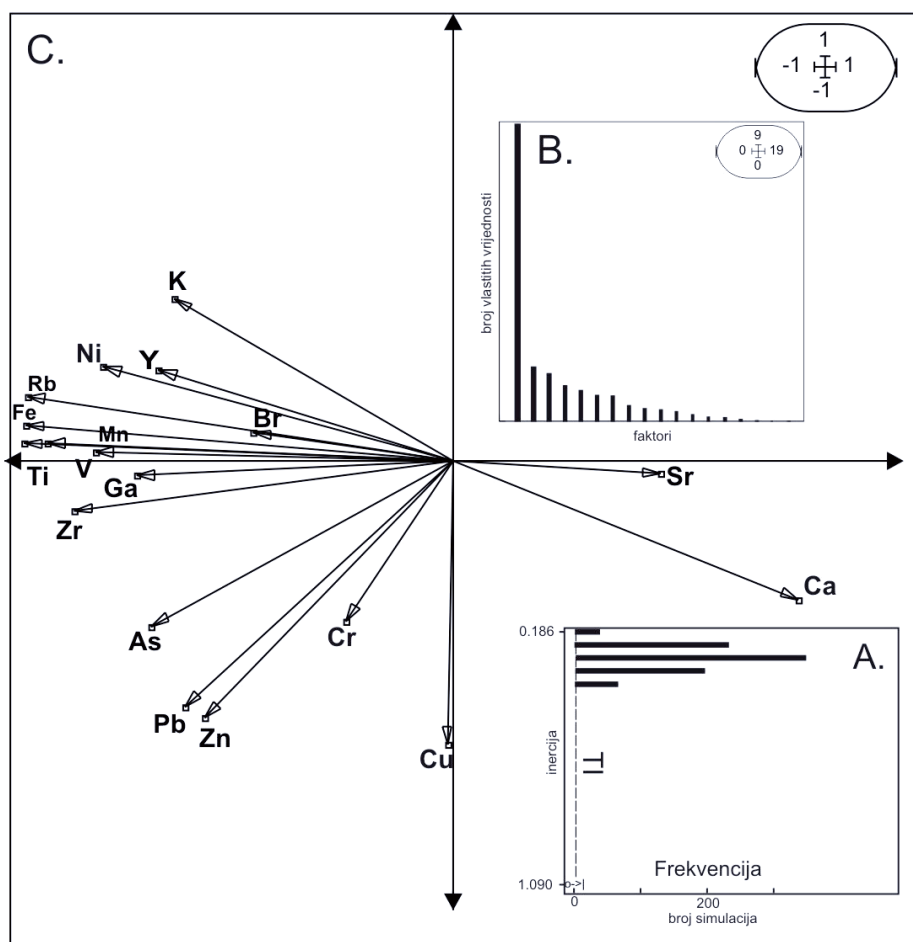
BG-PCA kao i PCA kombinira uzorke tako da traži nekolerirane linearne transformacije zadržavajući maksimalnu varijancu, samo što BG-PCA to radi s unaprijed definiranim grupama (Culhane et al., 2002; Doldec i Chessel, 1997). Za N grupa računa se $N-1$ vlastitih vektora ili komponenti koje razmještaju grupe na način da maksimiziraju varijancu između grupa. Svaki vlastiti vektor može se koristiti kao diskriminator koji odvaja jednu od ostalih grupa. Individualni uzorci razmještaju se uzduž tih osi oko centroida izračunatih za svaku grupu.

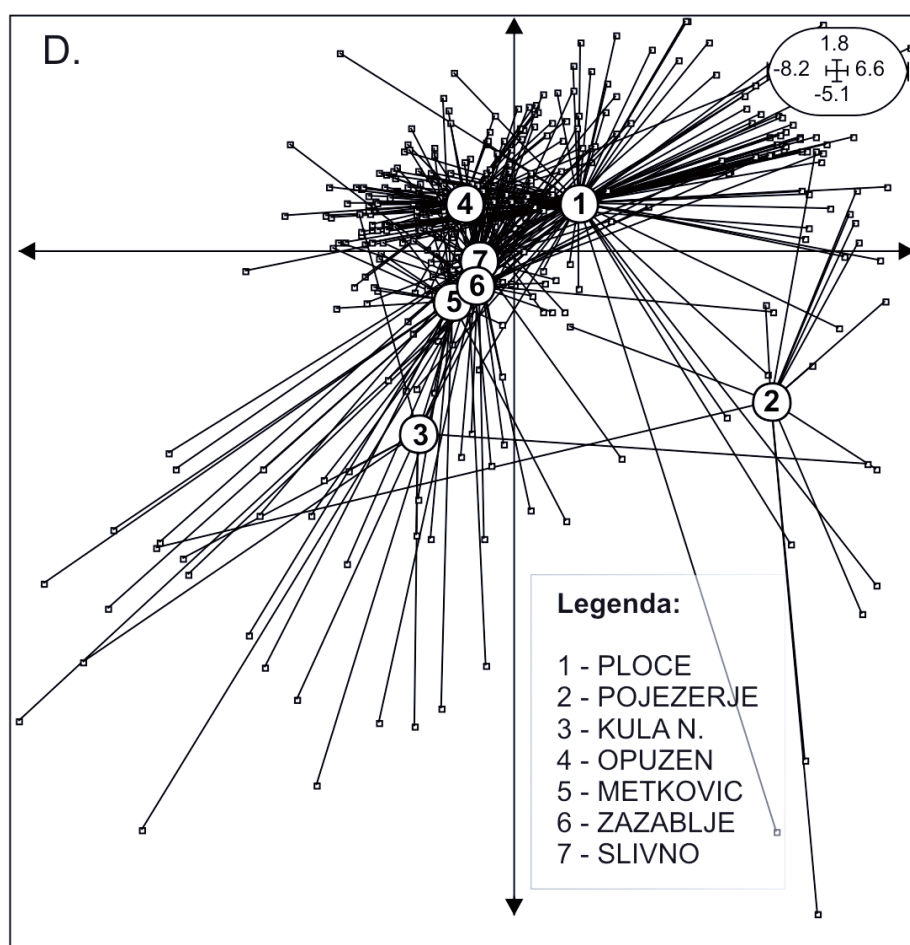
BG-PCA se primjenjuje i u svrhu utvrđivanja varijabli koje će najbolje razlučiti između dviju ili više prirodnih grupa. Za grupiranje se upotrebljavaju varijable čije se srednje vrijednosti statistički značajno razlikuju između različitih grupa. Za razlikovanje dviju grupa podataka koristimo se metodom višestruke regresije (utvrđivanje odnosa između zavisne varijable i više nezavisnih varijabli), a za određivanje više od dviju grupa kanoničkom korelacijskom analizom (utvrđivanje odnosa između dvaju skupova varijabli, npr. utvrđivanje veze između različitih faktora i grupe različitih simptoma). Kanonička analiza se općenito može provesti ako su uključene najmanje dvije

varijable i tri grupe podataka. Rezultat kanoničke analize jesu kanonički korijeni, odnosno diskriminantne funkcije. Prva dobivena funkcija (korijen) odgovara najboljoj diskriminaciji (razlučivanju) među grupama, druga funkcija sljedećoj najboljoj diskriminaciji itd. Maksimalni broj funkcija (korijenja) jednak je broju grupa umanjeno za jedan ili broju uzetih varijabli, ovisno o tome koji je od tih dvaju brojeva manji.

Kao rezultat analize dobiju se vrijednosti koje govore o vrijednosti ekstrakcije. Prva je vrijednost takozvana Wilks lambda (WL) koja može poprimiti vrijednosti od 0 kod idealne ekstrakcije do 1 prigodom preklapanja grupa, dok parcijalni lambda (PL) daje pojedinačni doprinos svake varijable diskriminaciji među grupama, a također može poprimiti vrijednosti od 0 do 1, ali je značenje obrnuto u odnosu prema WL. F i p vrijednosti pokazuju vjerojatnost doprinosa pojedine varijable razlikovanju podataka na razini značajnosti od 5% (obrnut kriterij u odnosu prema analizi varijance).

Na slici 6 prikazani su rezultati BG-PCA za varijable koje se odnose na koncentracije elemenata K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Ga, As, Br, Rb, Sr, Y, Zr i Pb u tlu i riječnom sedimentu administrativnih jedinica u delti Neretve.





Slika 6. BG-PCA za podatke o tlima/sedimentima donje Neretve grupiranim po općinama i gradovima. A) Statistički značajne varijabilnosti među grupama. TI - ukupna inercija, 1000 simulacija, „0“ - eksperimentalna ukupna inercija. Test je statistički značajan ukoliko je većina TI iznad „0“. B) Dijagram vlastitih vrijednosti. Opisuje varijance koje se odnose na osi. C) Grafički prikaz varijabli. D) Pojedinačni uzorci i gravitacijski centri (Obhodas, 2011)

Figure 6. BG-PCA for the Neretva region soil/sediment data set grouped by municipalities. A. Statistical significance of the variability among groups. TI – total inertia, 1000 simulations, „0“ – experimental total inertia. Test is significant if the most of the simulated TI is above „0“. B. Eigenvalues diagram. It describes axes related variance. C. Graphical display of variables. D. Scores and centers of gravity

BG-PCA pokazuje da su varijacije među administrativnim jedinicama statistički značajne, a vizualnim pregledom može se ustanoviti koje varijable najviše doprinose varijabilnosti i kako koreliraju. Može se zaključiti da Metković, Zažablje i Slivno imaju slične geokemijske značajke dok se ostale četiri administrativne jedinice prilično razlikuju. Rezultati pokazuju 4 grupe varijabli. Prvu čine Ca i Sr. Druga grupa je negativno korelirana s prvom, a uključuje K, Ti, V, Mn, Fe, Ni, Rb, Ga, Y i Zr. Treća grupa se sastoji od Cr, As, Zn i Pb, a četvrta sadrži samo Cu. Br nije detektiran kao značajan doprinositelj varijabilnosti među grupama. Razmještaj varijabli u usporedbi s gravitacijskim centrima može se interpretirati na sljedeći način. Kula Norinska je izložena povišenom riziku za zdravlje povezano s povišenim koncentracijama toksičnih elemenata As, Pb, Zn, Cr i Cu. Pojezerje ima povišeni rizik obzirom na izloženost povišenim koncentracijama Cu. Stanovništvo gradova Ploče i Opuzena je manje izloženo toksičnim elementima, dok je populacija Metkovića, Za-

žablja i Slivna izložena umjerenom riziku. S obzirom da se navedena analiza odnosi samo na tlo i sediment, nije moguće potvrditi niti opovrgnuti hipotezu da postoji indikacija o učinku geokemijskih obilježja okoliša na pobol stanovništva. U budućnosti će se analizirati još voda za piće, biota i kosa školske djece, a nakon kompletiranja baze podataka i usporedbe s pobolom moći će se testirati postavljena hipoteza.

Zaključak

U radu su prikazane GIS i statističke metode koje se mogu koristiti za povezivanje geografske distribucije pobola i geokemijskih karakteristika okoliša pri čemu treba imati na umu njihova ograničenja i uvjete korištenja. Uz ispravno korištenje ove metode mogu dati vrlo korisne rezultate stanovništvu, zakonodavcima i znanstvenoj zajednici.

Na primjeru otoka Krka prikazali smo kako izloženost ljudi nedostatku kalcija i magnezija u vodi može biti etiološki faktor bolesti krvožilnog sustava (I00-I99). Dokazali smo da postoje indikacije o povećanoj incidenciji krvožilnog pobola u populaciji koja u kućanstvima upotrebljava kišnicu koja se često koristi na našim otocima. Na žalost ovaj problem kod nas nije dovoljno istražen iako je u svijetu prepoznat još 1950-tih.

Na primjeru delte Neretve smo prikazali analizu pobola ukupnih sijela raka (C00.0-C97.0), malignih novotvorina debelog crijeva (C18.0-C21.8) i malignih novotvorina mjehura (C67.0-C67.9). Analiza varijance je pokazala da poboli C00-C99 i C67.0-C67.9 prema administrativnim jedinicama delte Neretve ne pripadaju istim raspodjelama što upućuje da bi mogući uzrok pobola mogao biti u karakteristikama geokemijskog okoliša. Sveobuhvatna analiza geokemijskog okoliša delte Neretve trenutno je u tijeku, a u radu su prikazane metode kojima su analizirani faktori raspodjele koncentracija nekih kemijskih elemenata u sedimentu i tlu. Primijenjene analize ukazuju na izvore i način širenja potencijalno toksičnih elementa čak i kad se radi o vrlo malim dozama koje se dugotrajnim djelovanjem akumuliraju u sedimentu i tlu te tako postaju vidljivi.

Kad je riječ o antropogenim izvorima poduzete mjere zaštite mogu uključivati edukaciju stanovništva i donošenje legislative s definiranim standardima ukoliko nedostaju i realnim granicama dozvoljenih koncentracija toksičnih tvari kako bi se smanjio rizik po zdravlje stanovništva. Da bi to bilo moguće, upravna i zakonodavna tijela bi trebala imati uvid u visokokvalitetne podatke o okolišu i u suradnji sa znanstvenom zajednicom definirati protokole za zaštitu zdravlja stanovništva.

Literatura

- ADE-4, 1997. Discrimin - Manual. <http://pbil.univ-lyon1.fr/ADE-4/home.php?lang=eng> [accessed on 13 July, 2011].
- Atkinson, P.J., Unwin, D.J. (2002): Density and local attribute estimation of an infectious disease using MapInfo Computers and Geosciences, 28, 1095-1105.
- Barr, D.B., Landsittel, D., Nishioka, M., Thomas, K., Curwin, B., Raymer, J., Donnelly, K.C., McCauley, Ryan, P.B. (2006): A Survey of Laboratory and Statistical Issues Related to Farmworker Exposure Studies, Environ. Health. Perspect. 114, 961-968.
- Buckeridge, D.L., Mason, R., Robertson, A., Frank, J., Glazier, R., Purdon, L., Amrhein, C.G., Chaudhuri, N., Fuller-Thomson, E., Gozdyra, P., Hulchanski, D., Moldofsky B., Thompson, M., Wright, R. (2002): Making health data maps: a case study of a community/university research collaboration, Social Science and Medicine 55, 1189-1206.
- Cao, Y.Q., Yang, Y.S., Hu, K.R., **Kalin, R.M. (2004): Ecological and geochemical modelling of hydrogeological system with particular connection to human health.** Ecological Modelling 174:4, 375-385.
- Cromley, E.K., McLafferty, S.L. (2002): GIS and public health, The Guilford press, 158 pp, New York.
- Culhane A.C., Perrière, G., Considine, E.C., Cotter, T.G., Higgins, D.G. (2002): Between-group analysis of microarray data. Bioinformatics, 18/12, 1600-1608.
- Dissanayake, C.B. i Chandrajith, R. (1999): Medical geochemistry of tropical environments, Earth. Sci. Rev., 47, 219-258.
- Doldec, S. i Chessel, D. (1997): Between – and within-groups principal components analysis. User manual, Ade-4 freeware, <http://pbil.univ-lyon1.fr/ADE-4-old/ADE-4.html>, [pristupljeno 28. siječnja 2007.]
- Jolliffe, I.T. (2002): Principal Component Analysis, 2nd ed, Springer, 518 pp, New York.
- Kobayashi J. (1957): On geographical relationship between the chemical nature of river water and death-rate from apoplexy. Berichte des ohara institutes fur landwirtschaftliche biologię; 11, 12-21.
- Koch, T., Denike, K. (2001): GIS approaches to the problem of disease clusters: a brief commentary, Social Science and Medicine, 52, 1751-1754.
- Kribek, B., Majer, V., Veselovsky, F., Nyambe, I. (2010): Discrimination of lithogenic and anthropogenic sources of metals and sulphur in soils of the central-northern part of the Zambian Copperbelt Mining District: A topsoil vs. subsurface soil concept, J. Geochem. Explor., 104, 69-86.
- Kutle, A., Orešćanin, V., Obhodaš, J., Valković, V. (2004): Trace element distribution in soils of the Island of Krk and its influence on the local population. J. Radioanalytical and Nuclear Chemistry, 259:2, 271-276.
- Kutle A., Obhodaš, J., Valković, V. (2006): Relation between the geochemical environment and disease incidence rate: Case study – the Island Krk in the Adriatic Sea, Croatia. Journal of Radioanalytical and Nuclear Chemistry, 269:3, 605-614.
- Kutle, A. (2004): Distribucija nekih elemenata u tragovima otoka Krka i pobol stanovništva, Ph.D. thesis, School of medicine, University of Osijek, Croatia,
- Masironi, R. (1987): Geochemistry, soils and cardiovascular diseases. Experientia, 43:1 68-74.
- Mather, F.J., White, L.E., Langlois, E.C., Shorter, C.F., Swalm, C.M., Shaffer, J.G., Hartley, W.R. (2004): Statistical Methods for Linking Health, Exposure, and Hazards, Environ. Health. Perspect. 112, 1440-1445.
- Medical Geology: A Regional Synthesis (2010): Selinus, O., Finkelman, R.B., Centeno J.A. (Eds.), Springer, 443 pp, Netherlands.
- Obhodaš J., Tucak, S., Kutle, A. and Valković, V. (2007): Indications for Synergetic and Antagonistic Effects between Trace Elements in the Environment and Human Health, Collegium Antropologicum, 31, 315-325.
- Obhodas, J. Kutle, A., Matjajac, L., Nad, K., Gregac, J., Valkovic, V. (2011): GIS-based multivariate statistics for the evaluation of the environmental degradation caused by the intensive agricultural practice, submitted to Env. Poll.
- Palesh, O., Butler, L. D., Koopman, C., Giese-Davis, J. Carlson, R. Spiegel, D. (2007): Stress History and Breast Cancer Recurrence, J. Psychosom. Res. 63/3, 233-239.

- Schroeder H. (1960a): Relations between hardness of water and death rates from certain chronic and degenerative diseases in the United States. *J. Chronic. Dis.*, 12, 586-591.
- Schroeder H. (1960b): Relationship between mortality from cardiovascular disease and treated water supplies. Variations in states and 163 largest municipalities of the United States, *J Amer. Med. Assoc.*, 172, 1902-1908.
- Schroeder H. (1969): The water factor, *N. Engl. J. Med.*, 280, 836-838.
- Sollitto, D., Romic, M., Castrignanò, A., Romic, D., Bakic, H. (2010): Assessing heavy metal contamination in soils of the Zagreb region (Northwest Croatia) using multivariate geostatistics. *Catena*, 80, 182-194.
- Sun, P.G., Gao, L., Han, S. (2011): Prediction of Human Disease-Related Gene Clusters by Clustering Analysis, *Int. J. Biol. Sci.*, 7, 61-73.
- Tan J, Zhu, W., Wang, W., Li, R., Hou, S., Wang, D., Yang, L. (2002): Selenium in soil and endemic diseases in China, *Sci. Total Environ.*, 284, 227-235.
- Valković, V. (1980): Analysis of biological material for trace element analysis using X-ray spectroscopy., CRC Press, 242 pp, Florida, Boca-Raton.
- Valković, V., Oreščanin, V., Mikulić, N., Obhodaš, J. (2001): Geochemical Map of Island Krk in the Adriatic Sea: Elements Determined by XRF, *J. Trace and Microprobe Techniques* 19:3, 393-408.

APPLICATION OF GIS AND STATISTICAL METHODS IN ESTABLISHMENT OF THE RELATIONSHIPS BETWEEN GEOCHEMICAL CHARACTERISTICS OF THE ENVIRONMENT AND GEOGRAPHICAL DISTRIBUTION OF DISEASE INCIDENCE RATE

Study of the geochemical environment influence on human health are very complex, mainly because it is necessary to collect a large number of georeferenced data from different segments of the environment. This usually include collection of data on concentrations of chemical substances in air, drinking water, soil, sediment, and biota that can be compared with the disease incidence (number of diagnoses within one year) in the population for which it may be presumed association with long-term exposure to a toxic effects or deficiency of a substances from the environment. The goal of such studies is to identify the causes and sources of toxicity, how they spread, the affected area and finally to identify the necessary protective measures to reduce human exposure to health risks. Integration and analysis of such a large amount of data would be practically impossible without the use of geographic information systems (GIS) and statistical methods.

Given that this area of the geochemistry is still fairly unknown in Croatia, this paper presents some results of research undertaken in recent years on the island Krk

and the Neretva delta. In this view, the emphasis is on examples of applied GIS and statistical methods used to connect the geographical distribution of disease and geochemical characteristics of the environment, along with a discussion of the pitfalls that may be encountered and that are sometimes overlooked as can be observed in some scientific publications. Described are methods of descriptive statistics, variance analysis, cluster analysis, linear regression analysis, factor analysis and the analysis of the main components between groups. In focus of this paper is practical application rather than theoretical explanations of the methods with the final goal of a simple interpretation of the results.

With proper use, GIS and statistical methods can provide very useful information to population, law regulators and to scientific community. Based on the island Krk example, we have showed that exposure to low concentrations of calcium and magnesium in the drinking water may be an etiological factor for cardiovascular diseases (I00-I99). We have proven that there are indications of an increased incidence of cardiovascular diseases in a population that uses rainwater in households, which is often the case on Croatian islands. Unfortunately, the problem has not been studied sufficiently in Croatia, even though recognized in the world since 1950's.

Based on the Neretva delta example, we have shown distribution analysis of disease incidence of malignant neoplasms (C00.0-C97.0), malignant neoplasm of the colon (C18.0-C21.8) and malignant neoplasm of bladder (C67.0-C67.9). Analysis of variance showed that distributions of C00.0-C97.0 and C67.0-C67.9 incidences after Neretva delta administrative units are statistically significantly different, suggesting that a possible etiological factor could be in the geochemical characteristics of the environment. Comprehensive analysis of the Neretva Delta geochemical environment is currently in progress, and paper presents methods that search for factors affecting the concentration distribution for selected chemical elements in the Neretva delta soils and sediments. Applied analysis point to the sources and spread of potentially toxic elements even in the case of very small doses which accumulate in sediment and soil over long period of time and thus become visible.

When it comes to anthropogenic sources, the protection measures can include education and legislation measures. The missing standards and realistic maximum permitted concentrations of toxic substances in the environment have to be defined in order to reduce the population health risk. To make this possible, the administrative and legislative bodies should have access to information on high quality environmental data and they have to define public health protection protocols in collaboration with the scientific community.