



WYKORZYSTANIE METOD STATYSTYCZNYCH W IDENTYFIKACJI CHEMIZMU WÓD PODZIEMNYCH NA PRZYKŁADZIE WOJEWÓDZTWA MAŁOPOLSKIEGO

Tomasz Kotowski¹, Stefan Satora¹, Dorota Pietrucin²

¹Uniwersytet Rolniczy im. H. Kollątaja w Krakowie

²AGH Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie

THE APPLICATION OF THE STATISTICAL METHODS TO GROUNDWATER CHEMISTRY BASED ON EXAMPLE THE MAŁOPOLSKA VOIVODESHIP AREA

Streszczenie

W pracy zaprezentowano zastosowanie wybranych metod wnioskowania statystycznego, stosowanych w badaniach hydrogeochemicznych. Metody statystyczne zostały zastosowane w celu wstępnej identyfikacji procesów formowania składu chemicznego wód podziemnych w obszarze woj. małopolskiego. Jako rezultat zastosowania analizy skupień otrzymano trzy grupy (skupienia) czynników kształtujących skład chemiczny wód podziemnych (rys. 2). W wyniku analizy czynnikowej wyróżniono zbiory trzech czynników (rys. 3, tab. 1). Wszystkie zastosowane metody okazały się wrażliwe na występowanie anomalii hydrogeochemicznych. Z tego względu, w przypadku analizy skupień, niemożliwe było dokonanie w pełni poprawnego opisu chemizmu wód podziemnych w obszarze badań. Najlepsze rezultaty osiągnięto przy zastosowaniu analizy czynnikowej, ale i ta metoda jest wrażliwa na występowanie „przypadków odstających”. Zastosowane metody statystyczne okazały się pomocne, jednakże interpretacja uzyskanych rezultatów, szczególnie dla zróżnicowanych danych, jest utrudniona.

Summary

The application of selected methods of the hydrogeochemical statistical analysis is presented in the paper. Statistical methods were applied to initially identify of the formation processes of water chemical composition of groundwater within the Lesser Poland Voivodeship area. As results of the cluster analysis three clusters' set of forming the chemical composition of groundwater's was distinguished (fig. 2). A set of three factors forming the chemical composition of groundwaters was distinguished as a result of the factor analysis (fig. 3, tab. 1). All statistical methods used were found out to be susceptible upon hydrogeochemical anomaly. Taking into account the observed susceptible of cluster analysis its application makes it impossible to describe enough correctly the formation of the chemical composition of groundwaters within the study area. Using the factor analysis attained the best results but this method is also susceptible upon anomaly cases. The statistical methods used were found out as helpful. However, the interpretation of research results without the additionally factors is intractable, especially for the case of hydrogeochemical anomaly.

Key words: *groundwaters, hydrogeochemistry, cluster analysis, factor analysis, Lesser Poland Voivodeship*

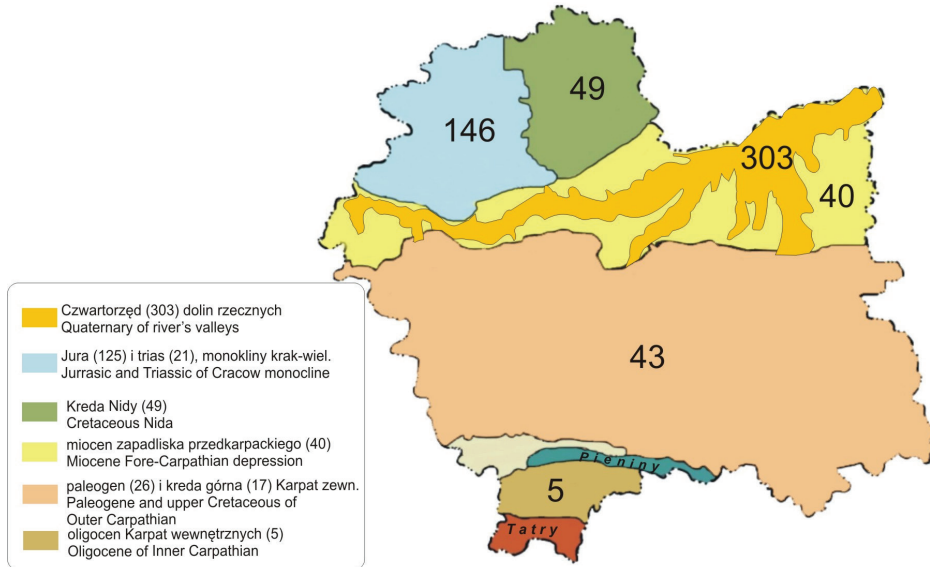
WSTĘP

W celu wstępnej identyfikacji chemizmu wód podziemnych i określenia czynników naturalnych i antropogenicznych wpływających na kształtowanie stężeń wybranych jonów oraz pH w tych wodach dla analizowanego obszaru obejmującego województwa małopolskie przeprowadzono wielowymiarowe wnioskowanie statystyczne przy wykorzystaniu hierarchicznej analizy skupień i analizy czynnikowej.

Obszar badań obejmuje teren województwa małopolskiego. W granicach omawianego obszaru występują następujące jednostki morfologiczne: Wyżyna Śląsko-Krakowska, Wyżyna Miechowska, Płaskowyż Proszowicki, Kotlina Oświęcimska, Brama Krakowska, Kotlina Sandomierska, Pogórze Wielickie, Pasma Beskidów, Pieniński Pas Skałkowy, Kotlina Podhalańska i Tatry.

Ośrodki wodonośne typu porowego i szczelinowo-porowego występują głównie w klastycznych osadach czwartorzędu, miocenu, paleogenu i górnej kredy (fliszu), permu i karbonu. Ośrodki wodonośne szczelinowo-porowego i szczelinowo-krasowego występują głównie w węglanowych warstwach kre-

dy Nidy, jury, triasu i dewonu. Obszerna charakterystyka analizowanych pięter wodonośnych została przedstawiona w pracy Satory [2008]. Lokalizację poszczególnych pięter wodonośnych i liczbę studzien występujących w ich obrębie przedstawiono na poniższym rysunku 1.



Rysunek 1. Lokalizacja pięter wodonośnych i liczba studzien występujących w ich obrębie

Figure 1. The localization of multiaquifer formations and number of wells

Z uwagi na zmienność geometrii pustek skalnych w ośrodkach wodonośnych można wydzielić główne typy środowisk hydrogeochemicznych obejmujące:

- czwartorzęd i miocen zapadliska przedkarpackiego – skały klastyczne, syplikie z dominującą zawartością materiału detrytycznego, niekiedy przewarstwione skałami spóistymi (ośrodki porowe, lokalnie szczelinowo-porowe);
- paleogen (w tym oligocen Karpat wewnętrznych) oraz kreda Karpat zewnętrznych – skały fliszowe lite spękane z dominującą zawartością materiału detrytycznego (ośrodki szczelinowo-porowe);

- kreda Nidy oraz jura i trias monokliny śląsko-krakowskiej – skały węglanowe spękane, miejscami z silnie rozwiniętymi procesami krasowymi (ośrodki szczelinowo-krasowe).

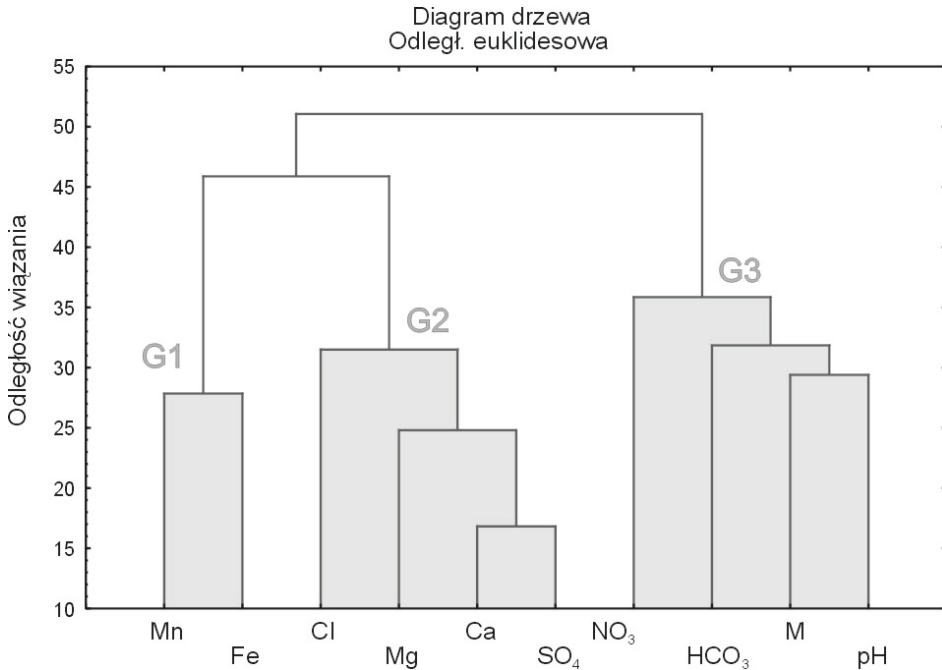
METODYKA

Wnioskowanie statystyczne prowadzono dla zbioru danych zawierającego wyniki analiz fizyko-chemicznych wód podziemnych na obszarze województwa małopolskiego. Zbiór danych obejmuje wyniki archiwalnych oznaczeń stężeń jonów HCO_3 , SO_4 , Cl , NO_3 , Ca , Mg , Fe , Mn , pH (pochodzących z baz danych „banku Hydro”) oraz zmienną M , która została przyjęta dla potrzeb wnioskowania statystycznego. Zmienna M nie jest zmienną porządkową i charakteryzuje zróżnicowanie środowisk hydrogeochemicznych. W dalszej analizie statystycznej założono, iż zmienna M może przyjmować trzy dopuszczalne stany: A – (czwartorzęd i miocen zapadliska przedkarpackiego), B – (paleogen, w tym oligocen Karpat wewnętrznych) i C – (kreda Nidy oraz jura i trias monokliny śląsko-krakowskiej).

Jako metodę pomocniczą podczas analizy głównych procesów kształtujących stężenia wybranych jonów zastosowano hierarchiczną analizę skupień. Zastosowanie hierarchicznej analizy skupień umożliwia rozdzielanie zbioru danych na skupienia (grupy) z uwagi na podobieństwo zmiennych występujących w utworzonych grupach. Metoda ta jest bardzo wrażliwa na obecność w zbiorze danych „przypadków odstających” tj. wyników analiz chemicznych, które wyraźnie różnią się od pozostałych. Jednakże pominięcie takich wyników analiz chemicznych prowadzi do zniekształcenia obrazu zmienności chemizmu wód podziemnych. Do analizy zbioru dziesięciu zmiennych o liczebności 581 przypadków wykorzystano metodę Warda. W metodzie tej wykorzystuje się do oszacowania odległości pomiędzy skupieniami analizę wariancji i jest określana, jako bardzo efektywna [Stanisz 2007].

Z uwagi na wrażliwość hierarchicznej analizy skupień dodatkowo do interpretacji procesów kształtujących zmienność stężeń wybranych jonów zastosowano analizę czynnikową. Test Barletta [1954] i analiza współczynnika Kaisera [1958] wykazały zasadność prowadzenia analizy czynnikowej dla analizowanego zbioru danych. Zastosowanie analizy czynnikowej polega m. in. na dekompozycji badanych zmiennych w nowy zbiór nieskorelowanych zmiennych

i następnie wyodrębnienie czynników wspólnych i swoistych dla takiego zbioru [Stanisz 2007]. Wykorzystanie analizy czynnikowej w hydrogeologii, wady i zalety tej metody, były przedmiotem licznych publikacji [Macioszczyk 1975, Ruiz i in. 1990, Jayakumar, Siraz 1997, Güler i in. 2002, Kotowski 2011]. Liczebność analizowanego zbioru wyników analiz fizykochemicznych wynosi 581 przypadków, co można określić jako znacznie przekraczającą wartość wystarczającą do prowadzenia analizy czynnikowej, zalecaną w literaturze [Tabachnik, Fidell 1996]. Dla zbioru danych zastosowano standaryzację i następnie transformację metodą składowych głównych. Analizowano zestaw 10 zmiennych, analogiczny jak w przypadku hierarchicznej analizy skupień. Liczbę badanych podczas analizy czynników określono w oparciu o test osypiska [Carte 1966] oraz pomocniczo kryterium Kaisera [1960]. W celu uzyskania czytelnego obrazu układu wartości czynnikowych, uzyskane wyniki poddano rotacji varimax (znormalizowanej).



Rysunek. 2. Diagram wyników hierarchicznej analizy skupień
Figure 2. The hierarchical cluster analysis result's diagram

INTERPRETACJA WYNIKÓW

Graficzną prezentację uzyskanych wyników przy wykorzystaniu hierarchicznej analizy skupień przedstawiono na rys. 2. Skupienie G3, w którym zostały zgrupowane jony HCO_3 , NO_3 , zmienna M oraz pH charakteryzuje wpływ zmienności środowiska hydrogeochemicznego na kształtowanie stężeń jonów HCO_3 oraz wartość pH. Obecność jonów NO_3 w tym skupieniu nie charakteryzuje bezpośrednio zmienności środowiska hydrogeochemicznego, lecz jest związana z wrażliwością hierarchicznej analizy skupień na występowanie w zbiorze danych przypadków odstających (anomalnych).

Zgrupowanie w skupieniu G2 jonów Ca, Mg i SO_4 oraz częściowo Cl można interpretować jako związane z rozpuszczaniem minerałów oraz procesami modyfikującymi skład chemiczny wód podziemnych np. ascencją z podłoża wód o wysokiej mineralizacji lub dopływem wód mocno zanieczyszczonych antropogenicznie. Lokalnie w osadach mioceńskich możliwe jest także rozpuszczanie gipsu, co może powodować występowanie podwyższonych stężeń jonów Ca i SO_4 . Widoczny związek pomiędzy jonami Ca, Mg i SO_4 może też wskazywać na występowanie procesów intensyfikujących rozpuszczanie minerałów węglanowych, szerzej przedstawionych w dalszej części pracy. Włączenie do skupienia G2 jonów Cl jest związane z wrażliwością analizy skupień na występowania w zbiorze danych przypadków odstających. Jony Cl nie mają bezpośredniego wpływu na rozpuszczalność minerałów, ale często współwystępują w wysokich stężeniach wraz z jonami SO_4 i Ca w wodach o wysokiej mineralizacji i/lub znacznie zanieczyszczonych antropogenicznie.

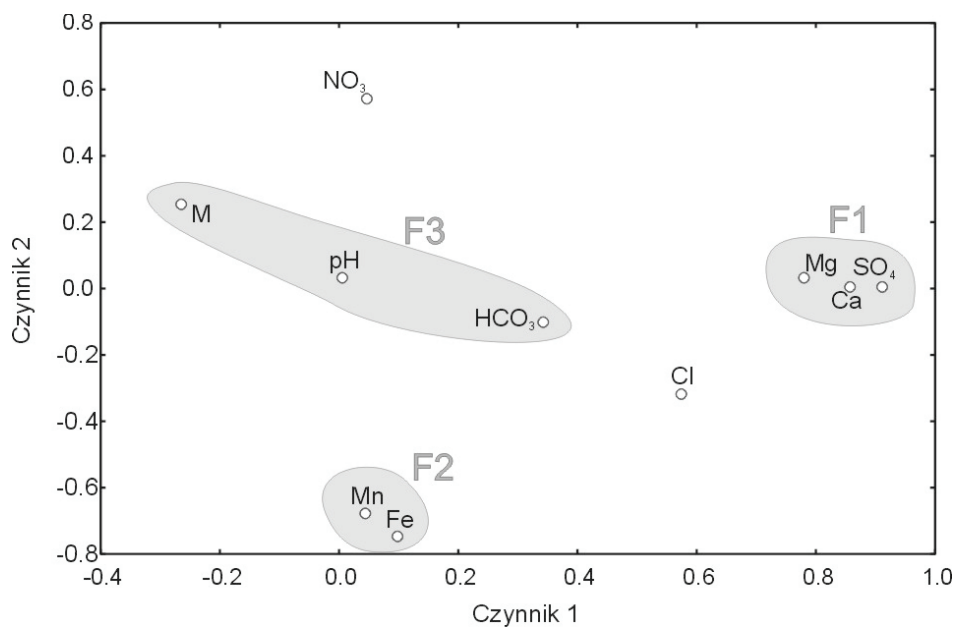
Skupienie G1 obejmujące jony Fe i Mn wskazuje, że czynnikiem odpowiedzialnym za związek tych jonów jest najprawdopodobniej zróżnicowanie warunków oksydacyjno-redukcyjnych występujące niezależnie od typu środowiska hydrogeochemicznego.

W wyniku zastosowania analizy czynnikowej otrzymano trzy czynniki, które określono odpowiednio, jako F1, F2 i F3. Wyodrębnione czynniki charakteryzują w 56, 4 % zmienność wariancji analizowanego zbioru danych. Wartości własne wyodrębnionych czynników wynoszą odpowiednio: $F1=2,78$; $F2=1,72$ i $F3=1,14$. Wyniki analizy czynnikowej w formie wykresu ładunków czynnikowych przedstawiono w tabeli 1 oraz na rys. 3.

Tabela 1. Wartości ładunków czynnikowych**Table 1.** The factor loadings

	Czynnik 1	Czynnik 2	Czynnik 3
pH	0,003	0,034	-0,645
HCO ₃	0,341	-0,099	-0,652
SO ₄	0,911	0,008	0,114
Cl	0,572	-0,314	0,069
Ca	0,856	0,007	-0,059
Mg	0,780	0,037	-0,134
Fe	0,098	-0,746	0,124
Mn	0,042	-0,675	0,275
NO ₃	0,044	0,575	0,227
M	-0,267	0,255	-0,635

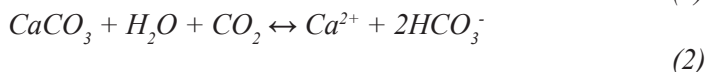
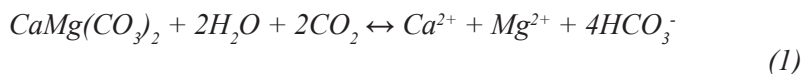
Pogrubieniem wyróżniono ładunki czynnikowe o wartości bezwzględnej powyżej 0,6.

**Rysunek. 3.** Wykres ładunków czynnikowych**Figure 3.** The plot of factor loadings

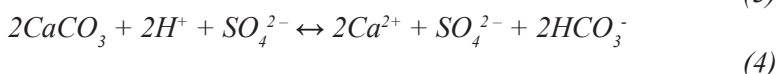
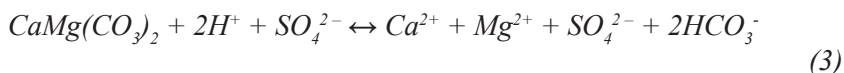
Czynnik F3 odzwierciedla, analogicznie jak omówione wcześniej skupienie G3, wpływ zmienności środowiska hydrogeochemicznego na stężenie jonów HCO_3 oraz wartość pH, która jest bezpośrednio związana z równowagą węglanową w wodach podziemnych. Nieobecność jonu NO_3 w tym czynniku wynika zapewne ze znacznie większej odporności analizy czynnikowej na występowanie przypadków anomalnych. Zmienna M (zróżnicowanie środowiska hydrogeochemicznego) została wyraźnie, przy stosunkowo wysokiej wartości ładunku czynnikowego ($>0,63$), zgrupowana w analizowanym czynniku F3.

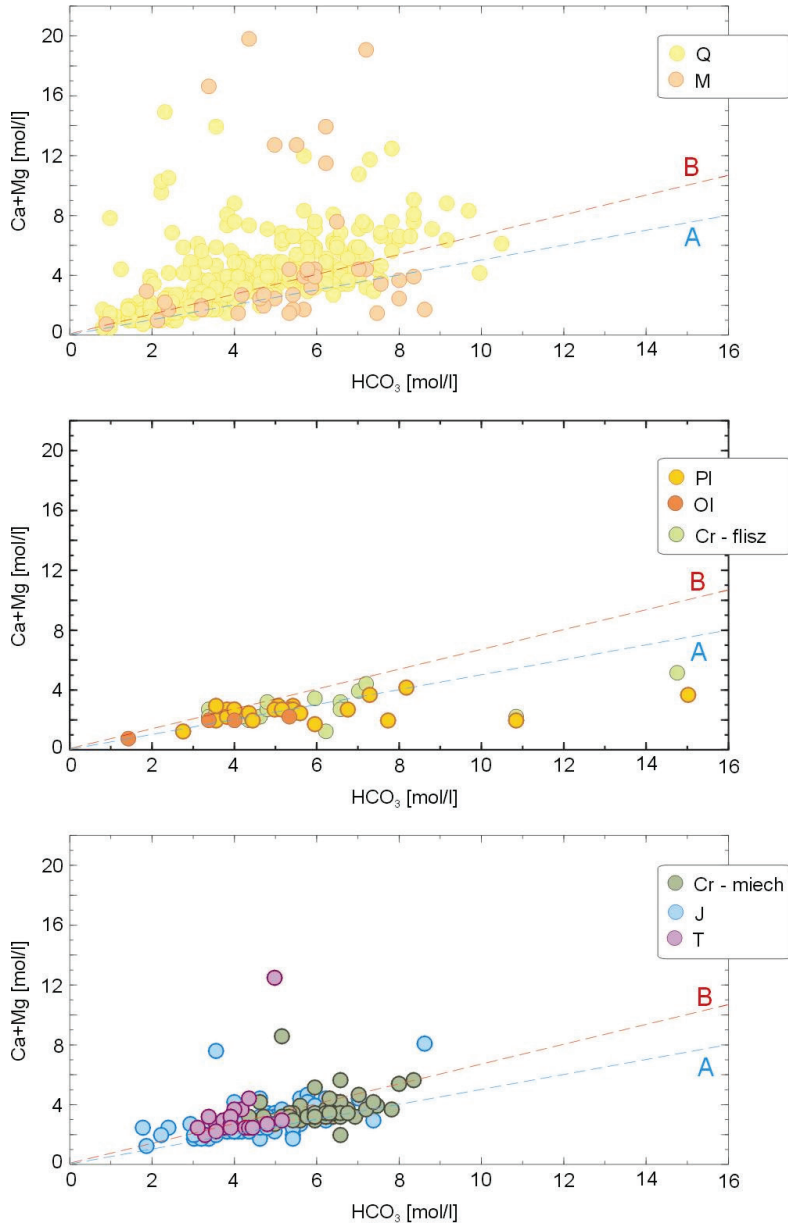
Wyodrębniony czynnik F2 zawierający jony Fe, Mn charakteryzuje zróżnicowanie warunków oksydacyjno-redukcyjnych. Wysoka wartość ładunku czynnikowego dla jonu NO_3 (tab. 1) może wskazywać na istotny wpływ procesów nityfikacyjnych i denityfikacyjnych na zmienność Eh środowiska i tym samym na kształtowanie stężeń jonów Fe i Mn, które są bardzo wrażliwe na zmiany tego parametru.

Zgrupowanie w czynniku F1 jonów Ca, Mg i SO_4 wskazuje, analogicznie do skupienia G1, na procesy rozpuszczania minerałów, głównie kalcytu i lokalnie także gipsu (osady mioceńskie). Jednakże w przypadku analizy czynnikowej jeszcze wyraźniej jest widoczna silna korelacja pomiędzy jonami Ca, Mg i SO_4 (rys. 3). Może to wskazywać na zaburzenie równowagi siarczanowej w wyniku dopływu wód o podwyższonej zawartości jonów SO_4 , co w rezultacie powoduje intensyfikację rozpuszczania minerałów węglanowych. Istotnym źródłem jonów SO_4 były w przeszłości i są obecnie zanieczyszczenia antropogeniczne (np. nawożenie, kwaśne deszcze itp.). Procesy intensyfikujące rozpuszczanie węglanów zachodzą równoległe z powszechnie występującymi procesami rozpuszczania węglanów, które są opisane przez reakcje 1 i 2 (krzywa A na rys. 4).



Intensyfikacja procesów rozpuszczania minerałów węglanowych zachodzi zgodnie z reakcjami 3 i 4 (krzywa B na ryc. 3).





Rysunek 4. Zawartość jonów Ca, Mg i HCO₃ w odniesieniu do teoretycznych linii rozpuszczania węglanów

Figure 4. The concentrations of Ca, Mg and HCO₃ ions in relation to theoretical curve of carbonate solubility

Współwystępowanie wymienionych wyżej procesów dobrze ilustruje rys.4, na którym przedstawiono wykresy teoretycznej rozpuszczalności węglanów sporządzone w oparciu o przedstawione powyżej zależności stechiometryczne w równaniach reakcji 1÷4.

Znaczna część punktów (ryc. 4) znajduje się w pobliżu teoretycznej linii rozpuszczania węglanów (krzywa A). Wskazuje to na istotne znaczenie procesów rozpuszczania minerałów opisanych reakcjami 1 i 2 w kształtowaniu składu chemicznego tych wód. Jednakże wyraźnie jest także widoczne występowanie punktów o wysokich stężeniach jonów Ca i Mg. Tak wysokie stężenia tych jonów nie mogły powstać jedynie w wyniku rozpuszczania węglanów zgodnie z reakcjami 1 i 2. W znacznej części punkty o wysokich stężeniach jonów Ca i Mg znajdują się w pobliżu krzywej teoretycznej B, która została skonstruowana w oparciu o założenia współwystępowania reakcji 1, 2, 3 i 4 pozostających ze sobą w stanie równowagi. Krzywa B uwzględnia dopływ wód o wysokich zawartościach siarczanów i wywołaną tym intensyfikację rozpuszczania węglanów. Jednakże w przypadku punktów znajdujących się wyraźnie ponad krzywą B skład chemiczny wód reprezentowanych przez te punkty jest zapewne w znacznej mierze modyfikowany przez dopływ wód zawierających wysokie stężenia jonów Ca i Mg. Dodatkowym źródłem jonów Ca i Mg może być zarówno ascenzja wód o wysokiej mineralizacji, jaki i dopływ wód silnie przekształconych antropogenicznie, co pośrednio potwierdza wysoka wartość ładunku czynnikowego¹ 0,57 określona dla jonów Cl. W przypadku osadach mioceńskich możliwe jest także rozpuszczanie gipsu, co może powodować występowanie podwyższonych stężeń jonów Ca i SO₄.

WNIOSKI

Zastosowanie metod statystycznych pozwoliło na wstępne określenie głównych procesów i czynników kształtujących skład chemiczny wód podziemnych w badanym obszarze. Rezultaty uzyskane w wyniku zastosowania analizy czynnikowej różnią się w niewielkim stopniu od wyników uzyskanego przy użyciu metody hierarchicznej analizy skupień. Wynika to głównie z odmiennej wrażliwości obu metod statystycznych występowanie w analizowanym zbiorze danych tzw. przypadków odstających.

¹Ładunek czynnikowy, czyli współczynnik korelacji pomiędzy wskaźnikami, a zmiennymi ukrytymi, wyliczany w trakcie analizy czynnikowej.

Zarówno w przypadku zastosowania metody hierarchicznej analizy skupień jak i analizy czynnikowej został wyróżniony wpływ zmienności środowiska hydrogeochemicznego na kształtowanie stężeń jonów HCO_3 , wartość pH oraz a także częściowo jonów NO_3 (skupienie G3). Z uwagi na fakt, iż analiza czynnikowa jest znacznie bardziej odporna na występowanie przypadków odstających można przyjąć, iż czynnik F3 wskazuje w sposób bardziej wiarygodny na procesy oddziaływujące dla całego zbioru danych tj. występujące w skali regionalnej dla całego obszaru badań.

Zgrupowanie w skupieniu G2 i czynnika F2 jonów Fe i Mn a także częściowo jonów NO_3 może wskazywać, że czynnikiem wpływającym na stężenia jonów Fe i Mn jest zmienność warunków oksydacyjno-redukcyjnych, od których głównie zależy np. stabilność poszczególnych form występowania związków żelaza w matrycy skalnej i/lub wodach podziemnych. W skali lokalnej istotny wpływ na stężenia tych jonów może mieć dopływ wód zanieczyszczonych antropogenicznie azotanami.

Zgrupowanie jonów Ca, Mg i SO_4 występuje niezależnie od wybranej metody analizy statystycznej. Najbardziej istotnym czynnikiem kształtującym zmienność stężeń jonów Ca i Mg jest dopływ wód o wysokich stężeniach jonów SO_4 . Dopływ takich wód może spowodować intensyfikację rozpuszczania minerałów węglanowych. Innym źródłem wysokich stężeń jonów Ca i SO_4 może być ascenzja, rozpuszczanie gipsu lub dopływ wód znacznie przekształconych antropogenicznie. Wpływ procesów antropogenicznych na stężenia jonów Ca i Mg z pewnością zaznacza się lokalnie, ale nie jest czynnikiem dominującym w całym analizowanym obszarze. Zanieczyszczenia wód podziemnych jonami SO_4 i Cl ma w większości wypadków charakter punktowy.

Generalnie, analiza czynnikowa okazała się mniej wrażliwa na występowanie wartości anomalnych w analizowanym zbiorze wyników analiz fizyko-chemicznych. Określona w analizie czynnikowej relatywnie niska wartość wyjaśnionej zmienności wariancji (56,4 %) wynika głównie z występowania licznych przypadków odstających w analizowanym zbiorze danych. Jednakże pominięcie takich wyników zniekształciłoby naturalny i charakterystyczny obraz zmienności składu chemicznego wód podziemnych w obszarze badań.

Zastosowanie analizy czynnikowej pozwoliło na wykrycie struktury i ogólnych prawidłowości występujących w zbiorze danych i tym samym na określenie głównych czynników kształtujących stężenia analizowanych jonów w skali regionalnej.

Podsumowując należy jednakże podkreślić, iż zastosowanie tylko metod statystycznych nie pozwala na wykrycie i zidentyfikowanie wszystkich procesów, które niekiedy w istotny sposób kształtują skład chemiczny wód podziemnych. Dlatego uzyskane wyniki należy traktować jako orientacyjne, głównie z uwagi na bardzo zmienne warunki hydrogeologiczne i geochemiczne oraz rozległy obszar badań.

BIBLIOGRAFIA

- Bartlett M.S., (1954) – *A note on the multiplying factors for various chi-square approximations*. Journal of the Royal Statistical Soc., (Series B) 16:296-298.
- Cattell R.B., (1966) – *The scree test for the number of factors*. Multivariate Behavioral Research, 1:245-276.
- Güler C., Thyne G.D., McCray J.E., Tuner A.K., (2002) – *Evaluation of graphical and multivariate statistical methods for classification of water chemistry data*. Hydrogeol. J., 10:455-474.
- Jayakumar R., Siraz L., (1997) – *Factor analysis in hydrogeochemistry of coastal aquifers – a preliminary study*. Environm. Geol., 31:174-177.
- Kaiser H.F., (1958) – *The varimax criterion for analytic rotation in factor analysis*. Psychometrical, 23:187-200.
- Kaiser H.F., (1960) – *The application of electronic computers to factor analysis*. Educational and Psychological Measurement, 20:141-151.
- Kotowski T., (2011) – *Porównanie wybranych statystycznych metod analizy danych hydrogeochemicznych na przykładzie kenozoicznych poziomów wodonośnych w obszarze zlewni Gwdy*. Biul. Państw. Inst. Geol. 447: 15-24
- Macioszczyk A., (1975) – *Zastosowanie analizy czynnikowej, sposób R do interpretacji danych hydrogeochemicznych*. Biul. Geol. Uniwersytetu Warszawskiego, 20:47-68.
- Ruiz F., Gomis V., Blasco P., (1990) – *Application of factor analysis to the hydrogeochemical study of a coastal aquifer*. J. Hydrology, 119:169-177.
- Satora S. (2008) – *Wpływ środowiska geologicznego na konstrukcje studni wierconych i reżim hydrogeologiczny ujętych wód podziemnych*, ZN UR 447. Rozprawy, z.324 : 127-160.
- Stanisz A., (2007) – *Przystępny kurs statystyki z zastosowaniem Statistica PL na przykładach z medycyny. Analizy wielowariantowe*. Wyd. StatSoft Kraków, t. III, ss.500.
- Tabachnik B.G., Fidell L.S., (1996) – *Using multivariate statistics*. Happer&Row, New York, 3th edition, ss.880.

Dr inż. Tomasz Kotowski)
Dr hab. inż. Stefan Satora prof. UR
Katedra Inżynierii Sanitarnej i Gospodarki Wodnej
Uniwersytet Rolniczy im.H.Kołłątaja
Al. Mickiewicza 21
31-120 Kraków
rmsatora@cyf-kr.edu.pl
kotowski.tomasz@gmail.com

Mgr inż. Dorota Pietrucin
Katedra Hydrogeologii i Geologii Inżynierskiej
Akademia Górniczo-Hutnicza im. St. Staszica
al. A. Mickiewicza 30
30-059 Kraków