

Zastosowanie metod chemometrycznych do eksploracji danych z kontroli jakości wody pitnej

M. Michalska, I. Stanimirova, M. Daszykowski, B. Walczak

Zakład Chemometrii, Instytut Chemii, Uniwersytet Śląski, ul. Szkolna 9, 40-006 Katowice, Polska
www.chemometria.us.edu.pl

Wstęp

Głównym celem pracy było określenie tendencji zmian parametrów wody pitnej na różnych etapach uzdatniania oraz znajdowanie zależności pomiędzy wartościami parametrów, a porą roku. Do eksploracji danych pochodzących z analizy próbek wody pobieranych z Zakładu Uzdatniania Wody w Goczałkowicach zastosowano metody chemometryczne takie jak analiza czynników głównych oraz metoda STATIS.

Analizowane dane

Analizowane dane pochodzą z analizy próbek wody z ZUW w Goczałkowicach. Próbki wody były pobierane co 2 tygodnie po każdym etapie uzdatniania (Schemat 1) i analizowane pod względem 29 parametrów (Tabela 1), czyli zgromadzone dane są przykładem danych trój-modalnych (Rys. 1).

Rys. 2 przedstawia dwa możliwe sposoby organizacji danych w zależności od stosowanej do ich analizy metody chemometrycznej.

Rys. 1 Przykład danych trój-modalnych

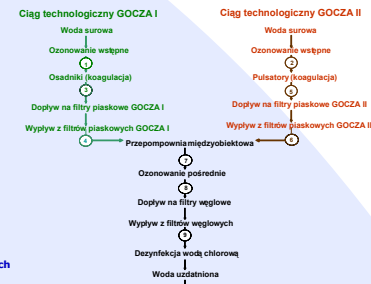
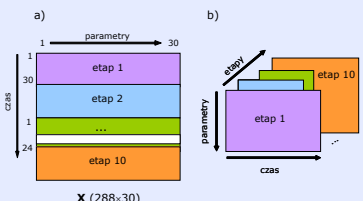


Tabela 1 Tabela przedstawiająca daty poboru próbek oraz techniki mierzenia poszczególnych parametrów

Ip	PARAMETR	SYMBOL PARAMETRÓW STOSOWANYCH W PROJEKCIACH	JEDNOSTKA	METODA POMIARU
1	barwność	bar	mg/L	Międzynarodowa
2	ciężkość właściwa	Cl	mg/L	Międzynarodowa
3	ciężkość właściwa	Zal	mg/L	Międzynarodowa
4	pH	pH	mg/L	Międzynarodowa
5	przewodność właściwa	Skal	mg/L	Międzynarodowa
6	przewodność właściwa	Skal	mg/L	Międzynarodowa
7	przewodność właściwa	Skal	mg/L	Międzynarodowa
8	K	K	mg/L	Chemograficzna
9	Ca	Ca	mg/L	Chemograficzna
10	Ca	Ca	mg/L	Chemograficzna
11	Ca	Ca	mg/L	Chemograficzna
12	Ca	Ca	mg/L	Chemograficzna
13	Mg	Mg	mg/L	Chemograficzna
14	Mg	Mg	mg/L	Chemograficzna
15	Mg	Mg	mg/L	Chemograficzna
16	Mg	Mg	mg/L	Chemograficzna
17	Mg	Mg	mg/L	Chemograficzna
18	Mg	Mg	mg/L	Chemograficzna
19	Mg	Mg	mg/L	Chemograficzna
20	Mg	Mg	mg/L	Chemograficzna
21	Mg	Mg	mg/L	Chemograficzna
22	Cu	Cu	mg/L	Atomowa
23	Mn	Mn	mg/L	Atomowa
24	Fe	Fe	mg/L	Atomowa
25	Al	Al	mg/L	Atomowa
26	Al	Al	mg/L	Atomowa
27	Al	Al	mg/L	Atomowa
28	Cu	Cu	mg/L	Atomowa
29	Mn	Mn	mg/L	Atomowa

Schemat 1 Ciągi technologiczne uzdatniania wody Zakładu Uzdatniania Wody w Goczałkowicach i miejsca poboru próbek do analizy symbolizowane przez cyfry w kółkach

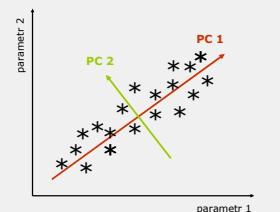


Rys. 2 Sposoby organizacji danych stosowane w metodzie a) PCA, b) STATIS

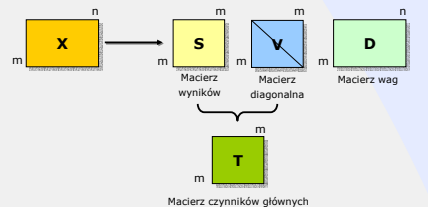
STOSOWANE METODY EKSPLORACJI DANYCH ANALITYCZNYCH

Analiza czynników głównych (z ang. Principal Components Analysis, PCA)

Analiza czynników głównych [1] to najczęściej stosowana metoda do analizy danych wielowymiarowych. Jej głównym celem jest kompresja danych, dzięki czemu wizualizacja wyników jest znacznie prostsza i przejrzystsza. Metoda ta polega na tworzeniu nowych wzajemnie prostopadłych osi, zwanych czynnikami głównymi (Rys. 3). Są one tak tworzone, by maksymalizować opis wariacji danych. Każdy kolejny czynnik jest ortogonalny do poprzednich, dzięki czemu opisuje zmienność, która nie została opisana przez poprzednie czynniki. Istnieje wiele algorytmów do wykonania analizy PCA. Jednym z nich jest Singular Value Decomposition, SVD. W jego wyniku macierz oryginalnych danych, X , jest przedstawiona jako iloczyn trzech macierzy, S , V i D (zob. Schemat 2).



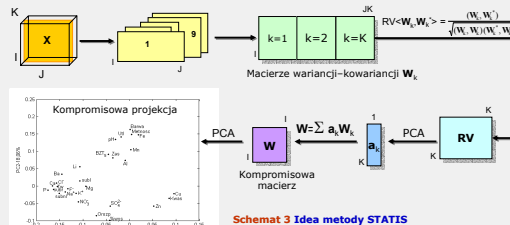
Rys. 3 Tworzenie czynników głównych PC 1 i PC 2



Schemat 2 Dekompozycja macierzy algorytmem SVD

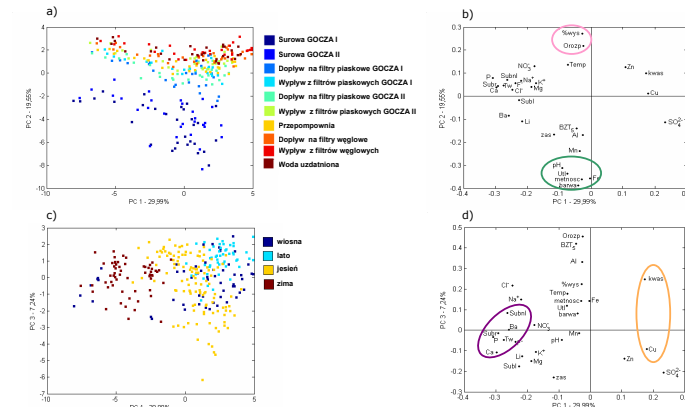
STATIS (z fr. Structuration des Tableaux A Trois Indices de la Statistique)

To jedna z metod wykorzystywanych do analizy tzw. trój-modalnych danych (Rys. 1). Główną ideę metody STATIS przedstawiono na Schemacie 3. Metoda ta umożliwia analizowanie danych jednocześnie pod względem etapów, parametrów i czasu, co pozwala na bardziej wnikliwą interpretację danych.

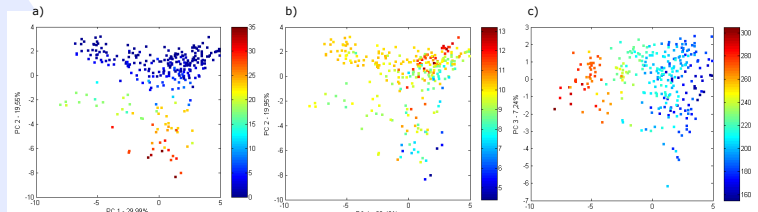


Schemat 3 Idea metody STATIS

Wyniki i dyskusja wyników

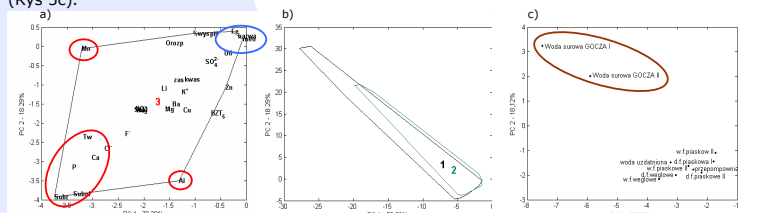


Rys. 4 a) Projekcja obiektów na płaszczyznę PC 1 i PC 2 oraz odpowiadająca im b) projekcja wag, c) projekcja obiektów na płaszczyznę PC 1 i PC 3 oraz odpowiadająca im d) projekcja wag



Rys. 5 Projekcja obiektów na płaszczyznę PC 1 i PC 2 przedstawiająca zmiany a) barwy [mgPt/L], b) zawartości tlenu rozpuszczonego w wodzie [mg tlenu/L] w kolejnych etapach uzdatniania, c) Projekcja obiektów na płaszczyznę PC 1 i PC 3 przedstawiająca zmianę przewodności właściwej [µS/cm] w kalendarzowych porach roku. Kolory punktów symbolizujących próbki są proporcjonalne do wartości danego parametru

Projekcje 4a i b pozwalają stwierdzić, że wartości takich parametrów jak barwa, mętność, zawartość żelaza (Fe), utleniałość (Utl) i pH będą małe, a wartości takich parametrów jak procent wycieszenia wody tlenem (%wty), zawartość tlenu rozpuszczonego w wodzie (Orozp) będą rosły w czasie procesu uzdatniania wody co zobrazowano na Rys 5a i b. Ponadto, z pomocą PCA potwierdzony został efekt sezonowości (Rys. 4c i d). Badane parametry można ogólnie podzielić na dwie grupy. Do pierwszej należą parametry o dużych wartościach w miesiącach chłodnych (np. zawartość wapnia (Ca) i przewodność właściwa (P)), a do drugiej, parametry o dużych wartościach w miesiącach ciepłych (np. kwasowość (kwas) i zawartość miedzi (Cu)) (Rys 5c).



Rys. 6 a) Rozkład 29 parametrów na kompromisowej projekcji dla dopływu wody na filtry piaskowe GOZCA I, b) nałożone na siebie powierzchnie zakreślone przez 29 parametrów dla (1) wody surowej GOZCA I i (2) wody surowej GOZCA II c) kompromisowa projekcja etapów uzdatniania wody

Rys. 6 przedstawia przykładowe projekcje uzyskane metodą STATIS. Tego typu projekcje pozwalają na śledzenie zmian wszystkich parametrów na poszczególnych etapach uzdatniania wody. Można wskazać parametry o wysokich (mangan (Mn), glin (Al), oraz parametry w lewym, dolnym rogu projekcji na Rys. 6a) i niskich wartościach (żelazo (Fe), barwa, mętność - Rys. 6a) na każdym etapie uzdatniania. Poprzez nałożenie na siebie powierzchni zdefiniowanych przez wszystkie parametry można porównać ze sobą różne etapy uzdatniania wody. Rys. 6b pozwala wnioskować, że wody surowe GOZCA I i GOZCA II są podobnej jakości, ale woda surowa GOZCA II jest czystsza niż woda surowa GOZCA I. Wody z tych dwóch etapów różnią się znacznie od wód z pozostałych odcinków uzdatniania co wyraźnie ilustruje Rys. 6c

WNIOSKI

Rozwój technik analitycznych powoduje, że analitycy zazwyczaj mierzą dużą liczbę parametrów dla jednej próbki. W celu eksploracji dużych zbiorów danych metody chemometryczne są bardzo użyteczne, gdyż pozwalają w prosty i szybki sposób na uwidocznienie podobieństw między próbkami i mierzonymi parametrami. Dla danych z kontroli jakości wody podejście chemometryczne pozwoliło na dogłębną analizę zależności pomiędzy kolejnymi etapami uzdatniania wody, porównanie pracy dwóch ciągów technologicznych jak i uchwycenie efektu sezonowości. Proponowana strategia eksploracji danych została także przedstawiona w [3].

PODZIĘKOWANIE

Autorzy pragną gorąco podziękować pracownikom Zakładu Chemii Analitycznej Instytutu Chemii Uniwersytetu Śląskiego w Katowicach za udostępnienie danych z kontroli jakości wody.

LITERATURA

[1] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, Amsterdam, 1998
 [2] I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, C.A. Saby, E. Di Crescenzo, STATIS, A three-way method for data analysis. Application to environmental data, Chemometrics and Intelligent Laboratory System 73 (2004) 219-233
 [3] I. Stanimirova, M. Polowniak, R. Skorek, A. Kita, E. John, F. Buhl, B. Walczak, Chemometric analysis of the water purification process data, Talanta, artykuł skierowany do druku