

Univerzita Pardubice

8. licenční studium chemometrie

Statistické zpracování dat při managementu jakosti

Semestrální práce Počítačová analýza vícerozměrných dat

Ing. Jan Balcárek, Ph.D.
vedoucí Centrálních laboratoří
Precheza a.s., Přerov

Přerov, 12.4.2001

Zadání

Pigmentový oxid titaničitý se vyrábí v několika druzích určených pro určité aplikace. Ve výstupní analytické kontrole se stanovuje celá řada parametrů, jejichž stanovení je více či méně časově náročné. Soubor obsahuje 75 vzorků šarží pigmentového oxidu titaničitého. Metodou hlavních komponent, faktorové analýzy a shlukovací metodou je třeba rozhodnout, které vzorky mají podobné vlastnosti. Na základě tohoto posouzení je nutno pak určit, které parametry má smysl měřit a které nikoli.

Data

Vzorek	TiO2 [%]	Těkavé látky [%]	Vodivost	pH	Barvivost	Podton	SiO2 [%]	Al2O3 [%]	Fe [%]	P2O5 [%]	K2O [%]
102588	98,04	0,37	390	7,1	1255	13	0,023	0,008	0,0018	0,375	0,275
102590	98,14	0,37	390	7,1	1260	14	0,021	0,007	0,0018	0,375	0,271
102593	98,35	0,4	410	7,2	1255	14	0,013	0,009	0,0016	0,372	0,253
102594	98,2	0,38	390	7,2	1260	14	0,019	0,014	0,0012	0,375	0,26
102623	99,11	0,33	410	7,2	1260	11	0,021	0,012	0,0066	0,307	0,236
103487	99,04	0,32	390	7	1260	9	0,029	0,011	0,0018	0,385	0,266
103942	99,26	0,35	380	7,3	1255	15	0,022	0,011	0,0018	0,303	0,244
103943	98,89	0,35	340	7,3	1240	14	0,011	0,008	0,0006	0,296	0,238
104092	98,89	0,36	340	7,1	1235	14	0,013	0,01	0,0014	0,318	0,234
103044	98,93	0,3	390	7,2	1220	7	0,021	0,007	0,0009	0,31	0,251
103647	99,01	0,33	330	7,3	1245	9	0,023	0,013	0,0004	0,326	0,225
103824	99,25	0,33	265	7,1	1230	12	0,029	0,007	0,0006	0,336	0,209
104002	99,31	0,28	360	7,2	1225	11	0,02	0,01	0,0005	0,312	0,233
104003	99,21	0,32	340	7,2	1225	11	0,025	0,008	0,0004	0,313	0,232
104004	99,12	0,31	320	7,2	1225	11	0,016	0,01	0,0009	0,307	0,23
104048	99,23	0,34	320	7,2	1225	9	0,014	0,006	0,0011	0,31	0,23
104060	98,92	0,35	330	7,1	1220	10	0,019	0,007	0,0008	0,301	0,222
104070	98,68	0,33	310	7,1	1220	11	0,02	0,009	0,0005	0,309	0,227
104071	98,94	0,31	340	7,2	1225	11	0,021	0,008	0,0009	0,304	0,231
104106	98,97	0,32	330	7,1	1235	12	0,024	0,009	0,0006	0,317	0,238
103084	98,62	0,18	570	9,1	1790	5	0,023	0,02	0,0067	0,182	0,154
103095	99,52	0,2	530	8,3	1720	5	0,01	0,023	0,0015	0,216	0,16
103096	99,25	0,25	540	8,1	1710	6	0,012	0,021	0,0011	0,22	0,167
103244	98,91	0,19	420	9,3	1760	5	0,021	0,037	0,0025	0,215	0,16
103273	98,64	0,2	430	8,7	1770	6	0,013	0,001	0,0022	0,191	0,15
103310	98,4	0,17	410	9	1760	6	0,012	0,001	0,0024	0,198	0,166
103786	98,7	0,21	420	8,9	1805	7	0,021	0,007	0,0016	0,189	0,211
104104	99,01	0,18	390	8,5	1795	9	0,028	0,01	0,0014	0,163	0,192
104109	98,79	0,18	440	8,5	1800	11	0,026	0,002	0,0013	0,163	0,191
104110	98,98	0,19	390	8,5	1810	11	0,023	0,012	0,0011	0,164	0,192
104147	99,19	0,22	460	8,5	1790	11	0,019	0,012	0,0035	0,158	0,187
104172	96,52	0,44	60	7,4	1825	11	0,202	2,308	0,0019	0,164	0,012
104191	96,2	0,46	70	7	1830	14	0,195	2,373	0,0017	0,198	0,007
104217	97,43	0,39	105	6,7	1810	13	0,145	1,913	0,0015	0,175	0,007
104222	96,89	0,43	70	7,1	1800	12	0,174	2,194	0,0013	0,174	0,012
104223	96,86	0,4	70	7,3	1805	12	0,167	2,178	0,0016	0,174	0,012
104229	96,82	0,45	95	7,1	1830	13	0,191	2,331	0,0015	0,177	0,008
104230	97,01	0,47	85	7,1	1810	12	0,21	2,367	0,0022	0,18	0,007

104231	97,03	0,45	50	8,1	1810	13	0,224	2,23	0,0017	0,199	0,015
104233	97,11	0,45	80	7,1	1800	13	0,191	2,357	0,0016	0,197	0,011
104234	96,97	0,45	75	7,1	1835	13	0,214	2,331	0,0018	0,181	0,009
104235	97,1	0,45	55	7,9	1805	13	0,225	2,232	0,0018	0,189	0,014
104093	98,19	0,25	20	8,4	1820	13	0,168	1,07	0,0022	0,172	0,01
104201	98	0,25	20	8,2	1810	13	0,167	1,211	0,0022	0,185	0,008
104211	98,21	0,23	25	8	1815	13	0,237	1,215	0,0023	0,184	0,006
104212	98,18	0,22	25	7,9	1815	13	0,197	1,187	0,0019	0,182	0,007
104213	98,61	0,23	25	7,7	1810	13	0,157	1,056	0,0014	0,181	0,005
104214	98,32	0,22	30	7,6	1810	12	0,214	1,181	0,0015	0,181	0,007
105045	96,29	0,41	60	8,6	1815	12	0,811	2,476	0,0029	0,184	0,022
105050	97,34	0,31	70	7,2	1810	14	0,498	1,715	0,002	0,179	0,013
105053	98	0,28	55	7,4	1815	14	0,27	1,417	0,0032	0,179	0,007
105054	98	0,27	35	7,8	1810	14	0,281	1,324	0,0024	0,184	0,006
105057	98,14	0,26	30	7,6	1810	12	0,236	1,294	0,0022	0,179	0,005
103740	97,19	0,39	45	8,9	1830	12	0,032	1,849	0,0021	0,191	0,023
103935	97,22	0,36	30	8,6	1820	8	0,023	1,746	0,0016	0,257	0,019
104138	97,99	0,35	65	7,2	1820	13	0,028	1,585	0,0019	0,178	0,004
104139	97,56	0,3	65	6,8	1825	14	0,037	1,604	0,0025	0,176	0,005
104144	97,89	0,3	60	7,1	1820	14	0,25	1,541	0,0026	0,179	0,005
104145	97,95	0,35	65	7,3	1820	13	0,033	1,57	0,0029	0,179	0,006
104148	98,13	0,3	55	6,9	1820	14	0,016	1,545	0,0032	0,177	0,005
104149	98,13	0,4	60	6,8	1820	13	0,021	1,585	0,0017	0,175	0,003
104159	98,31	0,38	55	7	1820	13	0,002	1,567	0,002	0,178	0,004
105069	98,76	0,25	60	6,5	1820	13	0,031	1,324	0,0021	0,193	0,004
105070	95,67	0,25	65	6,5	1850	13	0,002	1,227	0,0032	0,181	0,006
103870	95,73	0,68	65	8	1820	9	0,032	2,432	0,0018	0,205	0,018
105006	96,98	0,39	60	7,1	1820	11	0,057	1,533	0,0018	0,206	0,007
105007	97,34	0,34	60	6,9	1820	10	0,016	1,338	0,0016	0,191	0,005
105008	96,96	0,35	55	7	1820	11	0,035	1,347	0,0027	0,18	0,004
105009	97,06	0,35	55	7,1	1825	12	0,028	1,124	0,0033	0,177	0,003
105010	96,94	0,32	65	7	1825	12	0,043	1,432	0,0025	0,17	0,004
105013	96,83	0,36	100	7,6	1825	11	0,023	1,668	0,0019	0,171	0,005
105014	96,98	0,39	60	7,1	1820	11	0,057	1,533	0,0018	0,206	0,007
105015	97,34	0,34	60	6,9	1820	10	0,016	1,338	0,0016	0,191	0,005
105016	96,96	0,35	65	7	1820	11	0,019	1,347	0,0027	0,18	0,004
105017	97,06	0,35	50	7,1	1825	12	0,028	1,124	0,0033	0,177	0,003

1. Metoda hlavních komponent

Na soubor dat byla aplikována metoda hlavních komponent s použitím algoritmu NIPALS z korelační matice.

V Tabulce I jsou uvedeny vlastní hodnoty pro všech 11 komponent v původním souboru.

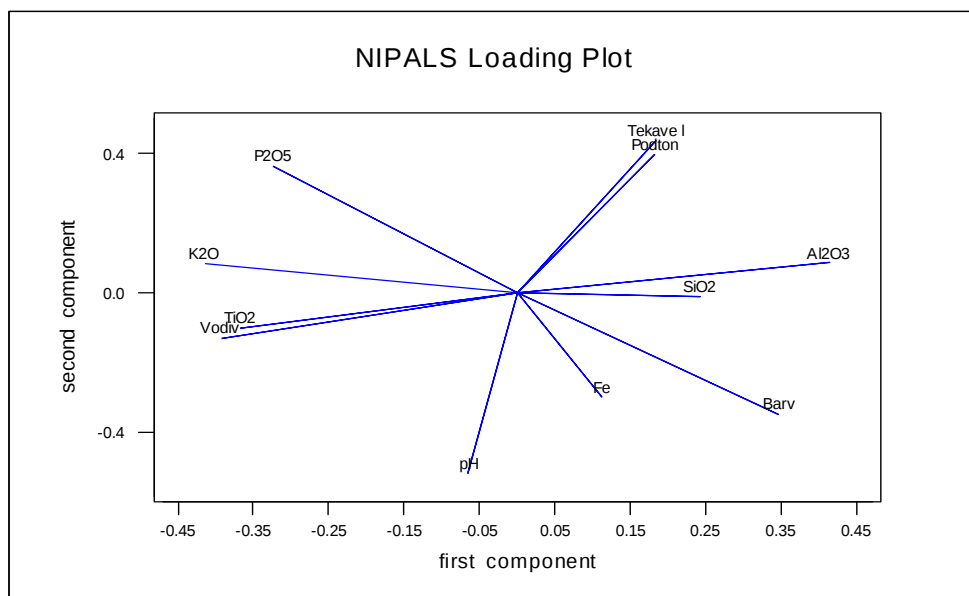
Tabulka I Tabulka vlastních hodnot

Hlavní komponenta	Vlastní hodnota	Proporčně	Kumulativně
Obsah TiO_2 1	5.3633	0.488	0.488
těkavé látky 2	2.3748	0.216	0.703
Vodivost 3	0.8781	0.080	0.783
pH 4	0.8756	0.080	0.863
Barvivost 5	0.7794	0.071	0.934
Podtón 6	0.3000	0.027	0.961
Obsah SiO_2 7	0.1844	0.017	0.978
Obsah Al_2O_3 8	0.1683	0.015	0.993
Obsah Fe 9	0.0439	0.004	0.997
Obsah P_2O_5 10	0.0221	0.002	0.999
Obsah K_2O 11	0.0100	0.001	1.000

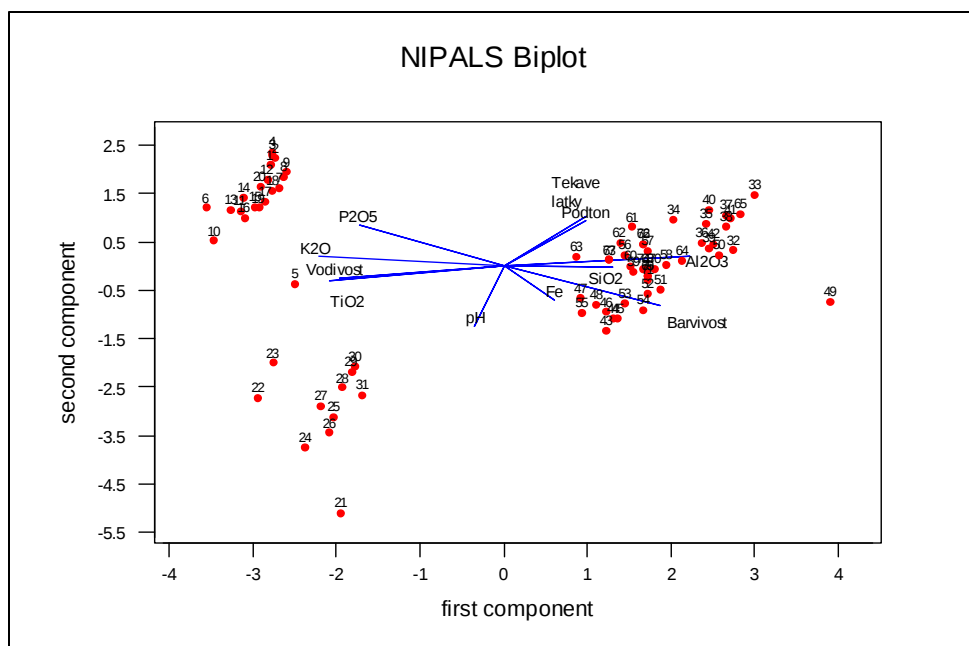
Z grafu komponentních vah (Obr. 1) a dvojného grafu (Obr.2) je patrná podobnost a tím i jejich vzájemná korelace u těchto komponent:

- obsah TiO_2 , vodivost
- podtón, těkavé látky
- obsah Al_2O_3 , SiO_2

Z první skupiny bude ponechána pro další výpočty vodivost. Obsah TiO_2 je makrosložka a jeho hodnotu lze počítat z korelačního vztahu: $\text{Obsah TiO}_2 = f(\text{vodivost})$. Ve druhé skupině bude ponechána komponenta “těkavé látky” a hodnota podtónu bude počítána z korelačního vztahu: $\text{Podtón} = f(\text{těkavé látky})$. Ze třetí skupiny bude ponechána komponenta “obsah Al_2O_3 ” a hodnota obsahu SiO_2 bude počítána z korelačního vztahu: $\text{Obsah SiO}_2 = f(\text{obsah Al}_2\text{O}_3)$.



Obr. 1 Graf komponentních vah

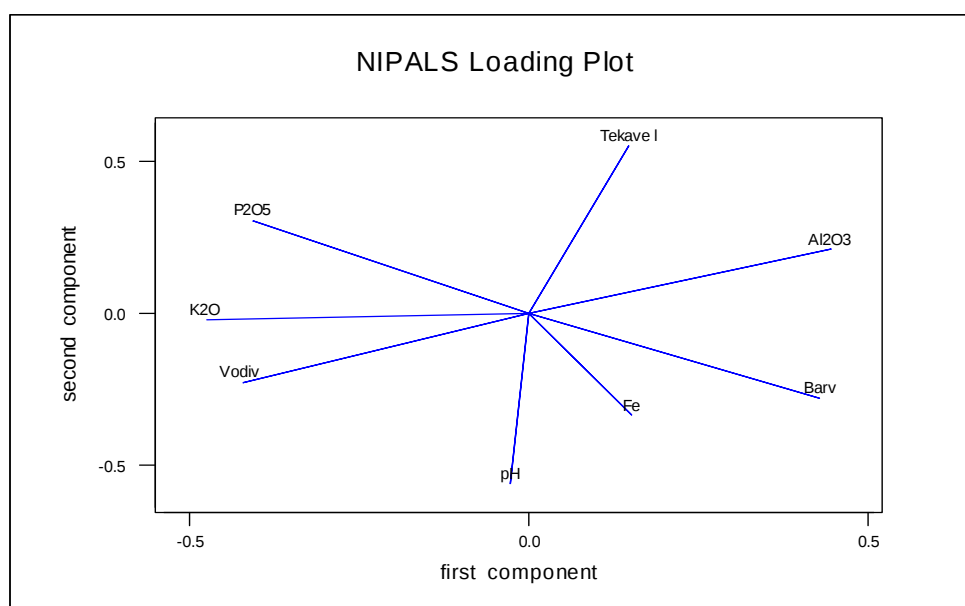


Obr. 2 Dvojný graf

Po výše uvedené redukci komponent bude situace znázorněná na obr. 3 a 4. Vlastní hodnoty jsou obsaženy v Tabulce II.

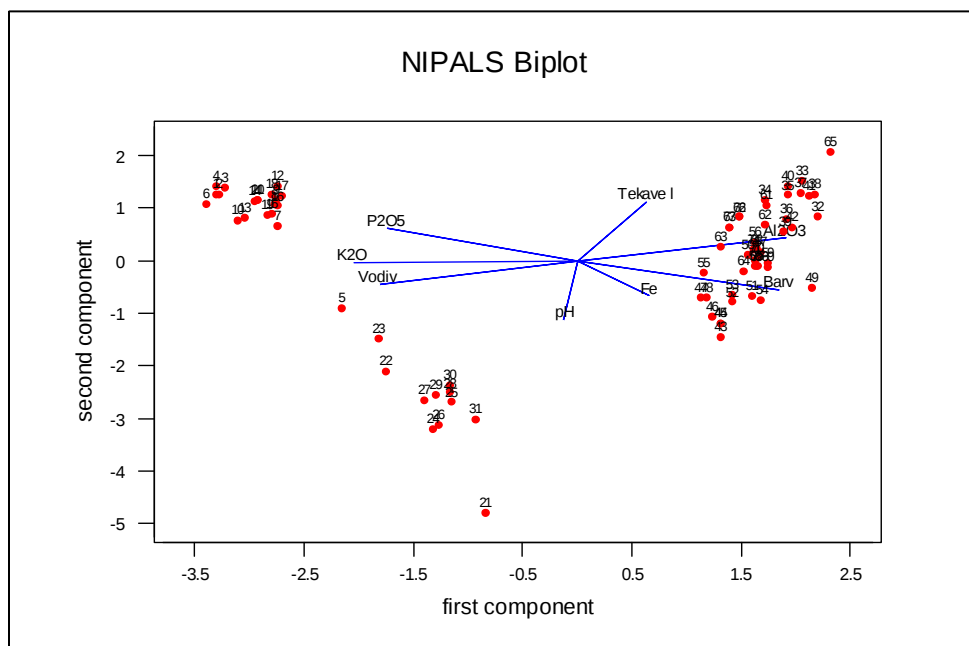
Tabulka II Tabulka vlastních hodnot po redukci komponent

Hlavní komponenta	Vlastní hodnota	Proporčně	Kumulativně
těkavé látky 2	4.2844	0.536	0.536
Vodivost 3	2.0037	0.250	0.786
pH 4	0.7928	0.099	0.885
Barvivost 5	0.6044	0.076	0.961
Obsah Al ₂ O ₃ 8	0.2131	0.027	0.987
Obsah Fe 9	0.0489	0.006	0.993
Obsah P ₂ O ₅ 10	0.0387	0.005	0.998
Obsah K ₂ O 11	0.0140	0.002	1.000

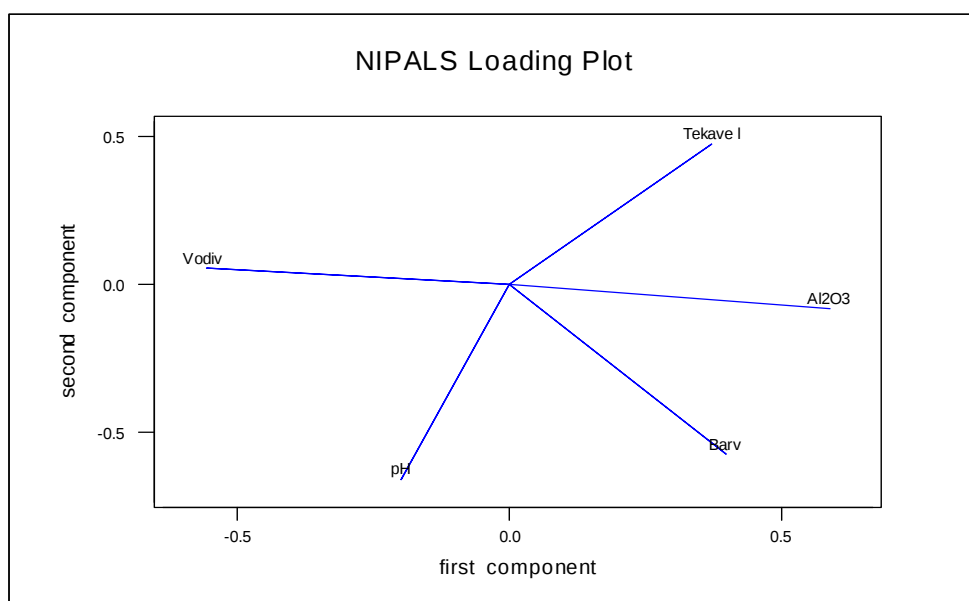


Obr. 3 Graf komponentních vah po redukci komponent

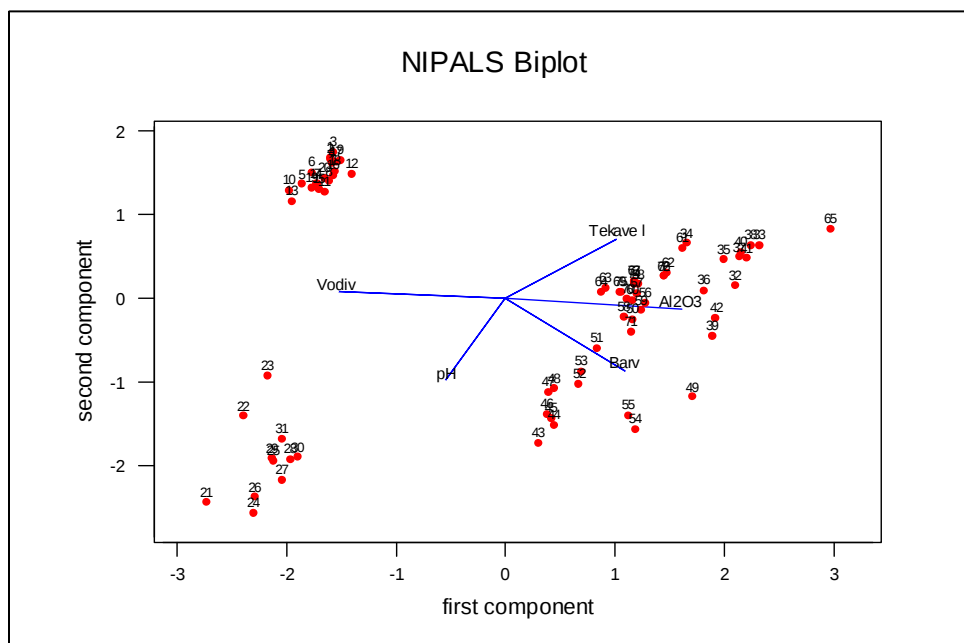
Z tabulky II plyne, že poslední tři komponenty (P₂O₅, Fe a K₂O) jsou nevýznamné a lze je zanedbat. Situace po této redukci vyplývá z obr. 5 a 6.



Obr. 4 Dvojný graf po redukcii komponent



Obr. 5 Graf komponentných vah po redukcii komponent Fe, K₂O, P₂O₅



Obr. 6 Dvojný graf po redukcii komponent Fe, K₂O, P₂O₅

Tabulka III Tabulka vlastních hodnot po redukcii komponent Fe, K₂O, P₂O₅

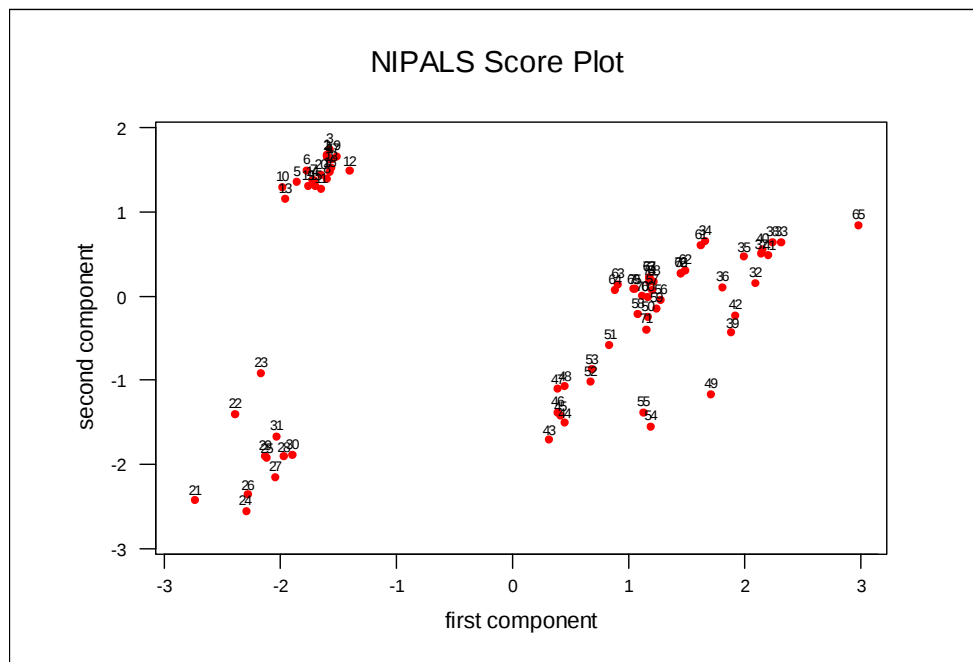
Hlavní komponenta	Vlastní hodnota	Proporčně	Kumulativně
těkavé látky 2	2.7314	0.546	0.546
Vodivost 3	1.4903	0.298	0.844
pH 4	0.5724	0.114	0.959
Barvivost 5	0.1677	0.034	0.992
Obsah Al ₂ O ₃ 8	0.0382	0.008	1.000

Po provedených redukcích se podařilo dosáhnout pro dvě latentní proměnné maxima vysvětlené variability 84,4 %. Latentní proměnné y_1 a y_2 mají následující tvar:

$$y_1 = 0,372 \cdot \text{Těkavélátky} - 0,560 \cdot \text{Vodivost} - 0,200 \cdot \text{pH} + 0,398 \cdot \text{Barvivost} + 0,591 \cdot \text{Obsah Al}_2\text{O}_3$$

$$y_2 = 0,473 \cdot \text{Těkavélátky} + 0,053 \cdot \text{Vodivost} - 0,660 \cdot \text{pH} - 0,576 \cdot \text{Barvivost} - 0,083 \cdot \text{Obsah Al}_2\text{O}_3$$

Následující obrázek ukazuje rozložení jednotlivých bodů v grafu při použití uvedených dvou proměnných:



Obr. 7 Graf závislosti y_2 na y_1 pro jednotlivé body

Z grafu je vidět, že soubor výsledků lze rozdělit dobře na tři shluky, což prezentuje tři skupiny výrobků určené pro různé aplikace.

2. Faktorová analýza

Metodou PCA se podařilo dosáhnout pro dvě latentní proměnné maxima vysvětlené variability 84,4 %. V dalším byly vypočteny faktorové zátěže a komunalita pro dva faktory. K výpočtu byla použita korelační matice. Dále byla použit metoda rotace EQUIMAX. U tohoto typu rotace je maximalizován rozptyl čtverců faktorových vah a zároveň je maximalizován součet čtvrtých mocnin faktorových zátěží. Výsledky jsou uvedeny v tabulkách IV až VII a na obr. 8, 9.

Tabulka IV Nerotované faktorové zátěže a komunalita

Proměnná	Faktor 1	Faktor 2	Komunalita
Těkavé látky	-0,616	0,577	0,712
Vodivost	0,925	0,065	0,860
pH	0,331	-0,806	0,759
Barvivost	-0,658	-0,703	0,927
Al ₂ O ₃	-0,977	-0,102	0,964
Rozptyl	2,7314	1,4903	4,2217
% Rozptylu	0,546	0,298	0,844

Tabulka V Koefficienty faktorových skóre - nerotované

Proměnná	Faktor 1	Faktor 2
Těkavé látky	-0.225	0.387
Vodivost	0.339	0.043
pH	0.121	-0.541
Barvivost	-0.241	-0.471
Al ₂ O ₃	-0.358	-0.068

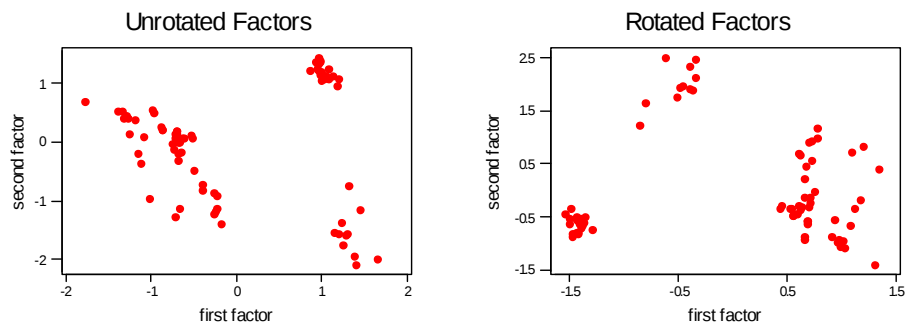
Tabulka VI Rotované faktorové zátěže a komunalita

Proměnná	Faktor 1	Faktor 2	Komunalita
Těkavé látky	0.302	-0.788	0.712
Vodivost	-0.860	0.346	0.860
pH	0.054	0.869	0.759
Barvivost	0.899	0.345	0.927
Al ₂ O ₃	0.923	-0.335	0.964
Rozptyl	2.4949	1.7268	4.2217
% Rozptylu	0.499	0.345	0.844

Tabulka VII Koeficienty faktorových skóre - rotované

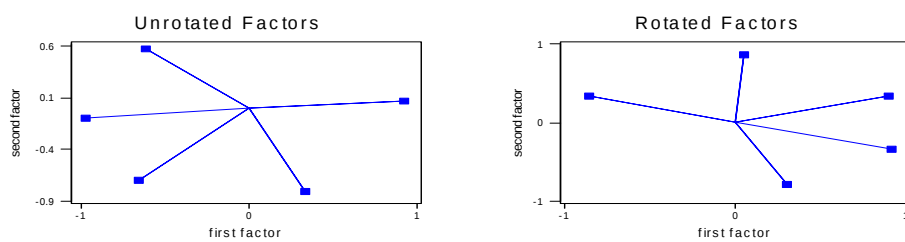
Proměnná	Faktor 1	Faktor 2
Těkavé látky	0.034	-0.447
Vodivost	-0.324	0.109
pH	0.127	0.539
Barvivost	0.423	0.319
Al ₂ O ₃	0.351	-0.095

Factor Analysis Score Plot



Obr. 8 Graf faktorových skóre před a po rotaci

Factor Analysis Loading Plot



Obr. 9 Graf faktorových zátěží před a po rotaci

Závěr

Z faktorové analýzy lze separovat tři faktorově čisté proměnné. Pro popis dat stačí použít dvou faktorů, které vysvětlují 84,4 % variability.

3. Výsledky výpočtu korelační a kovarianční matice

Výsledky výpočtu jsou uvedeny v tabulce VIII a IX.

Tabulka VIII Korelační matice

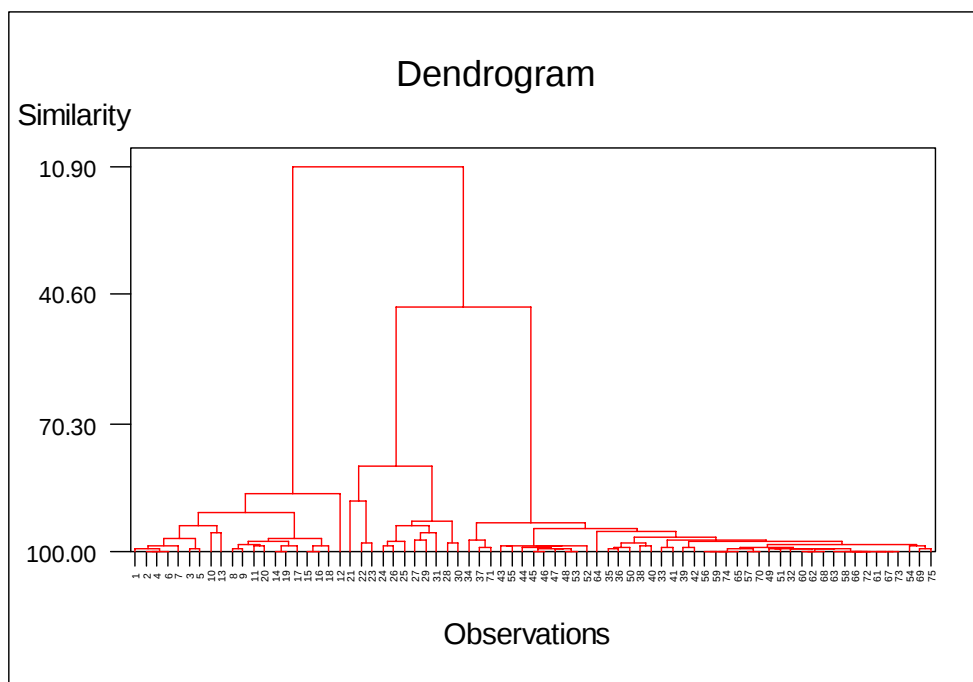
	Těkavé látky	Vodivost	pH	Barvivost
Vodivost	-0.399			
pH	-0.429	0.318		
Barvivost	-0.032	-0.612	0.267	
Al ₂ O ₃	0,590	-0,869	-0,195	0,699

Tabulka IX Kovarianční matice

	Těkavé látky	Vodivost	pH	Barvivost	Al ₂ O ₃
Těkavé látky	0,008				
Vodivost	-6.087	29217.496			
pH	-0.026	36.625	0.454		
Barvivost	-0,731	-26662.918	45.923	65037.324	
Al ₂ O ₃	0.047	-132,042	-0,117	158,474	0,791

4. Analýza shluků

Bylo použito hierarchické klastrování podle Eukleidovské vzdálenosti. Do výpočtu byly zahrnuty parametry těkavé látky, vodivost, pH, barvivost a obsah Al_2O_3 . Byl vypočten dendrogram, který ukazuje tři shluky vzorků, které jsou si podobny. Tento výpočet je v soulase s výsledkem faktorové analýzy.



Obr. 10 Dendrogram shlukové analýzy

5. Závěr

Soubor dat, který byl podroben analýze metodou hlavních komponent, faktorové a shlukové analýze lze rozdělit na tři skupiny výrobků pro různé aplikace.