

Univerzita Pardubice
Fakulta chemicko-technologická

Katedra analytické chemie

Počítačová analýza vícerozměrných dat

Ing. Roman Slavík
ŽDB a.s.

V Bohumíně 14.4.2001

Příklad č.1

Monitorování kvality podzemních i povrchových vod.

Zadání :

Monitoring je prováděn na území zahrnující hutní odval (vlastník ŽDB), skládku tuhých průmyslových odpadů (vlastník ŽDB) a městskou skládku komunálního odpadu (vlastník BMS), včetně přilehlých území. Veškeré terénní, vzorkovací a laboratorní práce jsou finančně i časově velmi náročné. Rozhodněte, zda je možné redukovat počet odběrových míst a charakterizovat jednotlivé vrty menším počtem stanovovaných ukazatelů.

Data :

Bylo vzorkováno 20 vrtů a jedna voda povrchová.

VZOREK	OZN	PH	VODIVOST	-F	CL	CHSK	CR	B	BA	CU	FE	MN	ZN	SN
3438	OD 1	6,980	250,000	1,000	386,100	62,000	0,131	0,065	0,009	27,300	17,600	0,018	0,650	
3439	OD 3	7,050	364,000	2,550	772,200	49,000	3,470	0,059	0,008	6,160	1,710	0,009	0,394	
3440	OD 4	7,020	249,000	1,390	386,100	53,000	4,200	0,240	0,038	62,000	1,570	0,046	0,071	
3441	OD 5	8,110	366,000	3,600	702,000	30,000	4,100	0,069	0,014	0,617	0,087	0,007	0,400	
3442	OD 7	6,570	344,000	1,150	737,100	52,000	0,198	0,484	0,013	93,200	5,970	0,010	0,070	
3443	OD 8	7,760	351,000	2,200	666,900	68,000	2,400	0,103	0,004	3,260	3,140	0,022	0,527	
3444	PV 32	6,870	215,000	0,480	122,900	32,000	4,270	0,063	0,012	24,300	2,500	0,060	0,438	
3445	PV 102	7,700	242,000	5,750	280,800	41,000	6,130	0,068	0,029	2,410	1,580	0,023	0,192	
3446	PV 105	7,050	222,000	1,130	80,730	36,000	0,505	0,069	0,165	19,500	3,460	0,377	0,467	
3447	PV 106	6,770	271,000	0,480	280,800	170,000	2,740	0,112	0,024	74,700	14,900	0,045	0,682	
3448	BC 2	7,600	256,000	0,380	631,800	21,000	0,430	0,340	0,001	1,050	0,363	0,048	0,180	
3450	PV 1	6,710	366,000	0,250	912,600	41,000	0,428	0,071	0,054	2,190	4,130	0,045	0,061	
3451	PV 2	7,210	196,300	0,900	210,600	35,000	3,860	0,044	0,018	0,177	0,320	0,025	0,049	
3452	PV 5	7,520	219,000	0,880	421,200	12,000	0,437	0,068	0,019	0,095	11,850	0,031	0,155	
3453	PV 9	6,770	157,700	0,550	175,500	22,000	1,390	0,113	0,073	12,100	1,610	0,112	0,028	
3454	PV 10	6,340	348,000	0,290	702,000	23,000	7,340	0,080	0,027	0,643	2,690	0,073	0,097	
3455	PV 11	6,290	166,400	0,690	263,200	10,000	0,949	0,121	0,018	12,000	1,900	0,018	0,078	
3456	PV 12	5,850	375,000	4,100	1123,000	40,000	0,039	0,607	0,016	96,600	23,650	0,074	0,157	
3457	PV 40	5,670	137,100	5,100	421,200	13,000	0,114	0,509	0,024	53,500	4,000	0,028	0,047	
3458	PV 103	6,680	282,000	0,510	473,800	43,000	1,375	0,084	0,017	0,984	1,250	0,036	0,084	
3459	PV 104	6,560	161,700	0,940	98,280	40,000	0,436	0,131	0,013	34,900	3,750	0,024	0,081	

Program :

Statgraphics, Microsoft Word, Scan

Řešení :

- Metoda hlavních komponent

Cílem metody PCA je redukce počtu proměnných a zároveň získání co nejmenšího počtu hlavních komponent vysvětlujících maximální variabilitu původní matice. Metodou hlavních komponent jsme zpracovali výše uvedená data. Data byla standardizována, protože nejsou ve stejných jednotkách. V tabulce č. 1 jsou uvedeny výsledky. Ve druhém sloupci jsou uvedena procenta variability vysvětlená jednotlivými hlavními komponentami, tj. podíly rozptylů jednotlivých hlavních komponent na celkové variabilitě, v pořadí od nejvyššího k nejnižšímu. Ve třetím sloupci jsou tytéž podíly uvedeny kumulovaně a lze z nich určit optimální počet komponent. Většinou se požaduje, aby součet prvních nejvyšších komponent měl 85 až 92 %. Z tabulky č.1 vyplývá, že tento předpoklad splňuje 5 hlavní komponenta.

Na obr. č. 1 – 3 jsou znázorněny grafy metody hlavních komponent. Na obr.č.1 je znázorněn graf komponentních vah pro první dvě hlavní komponenty u jednotlivých proměnných vstupujících do analýzy. Na obr.č.2 je znázorněn bodový diagram, který zobrazuje hodnoty prvních dvou hlavních komponent u jednotlivých objektů. Tohoto grafu se využívá ke grafickému ověření předpokladu normality pro další analýzy.

Na obr.č.3 je znázorněn dvojný graf, který kombinuje oba předchozí grafy. Body představují komponentní skóre jednotlivých objektů. Přímký spojující se v bodě (0,0) představují původní proměnné, délka každé přímký je úměrná svému příspěvku k prvním dvěma hlavním komponentám.

Metody s latentními proměnnými a klasifikační metody

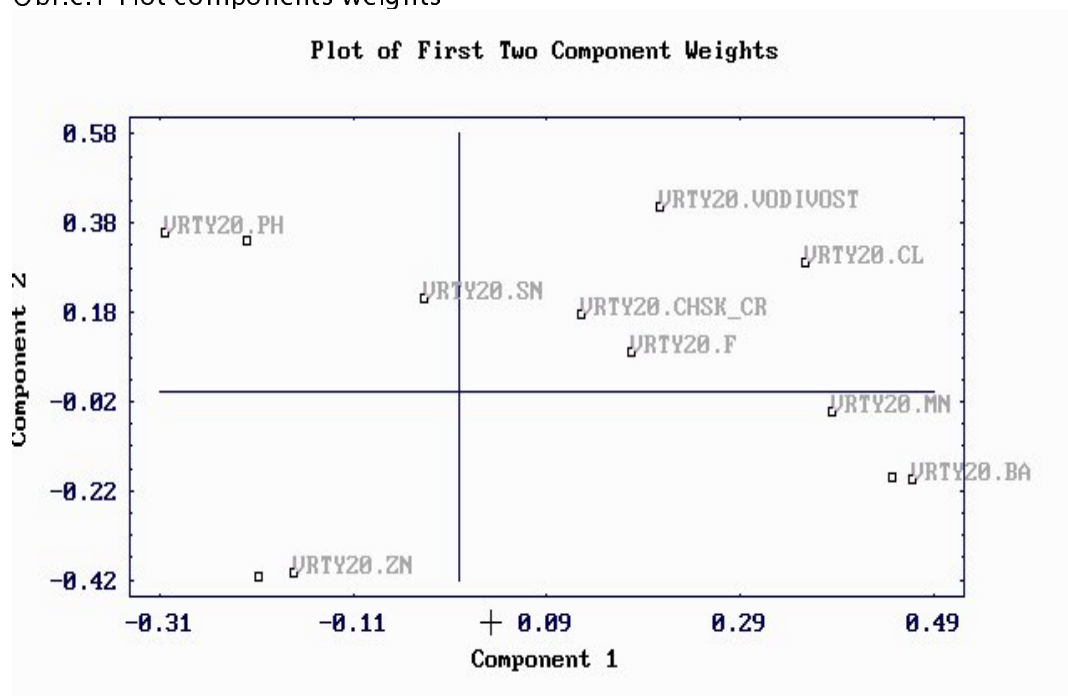
Z číselného výstupu (Tab.č1) je zřejmé, že 91,04% splňuje součet rozptylů 6 hlavních komponent. Z grafického výstupu obr.č.1-3 vyplývá, že můžeme redukovat počet proměnných na 5 (B, Zn, Fe, Cl, Vodivost). Výsledky jsme uvedli do Tabulky č.2, ze které vyplývá, že 85% splňuje součet rozptylů 3 hlavních komponent. Na obr. č.4 – 6 jsou znázorněny grafy metody hlavních komponent pro zredukováná data.

Tab. č. : 1 Analýza hlavních komponent

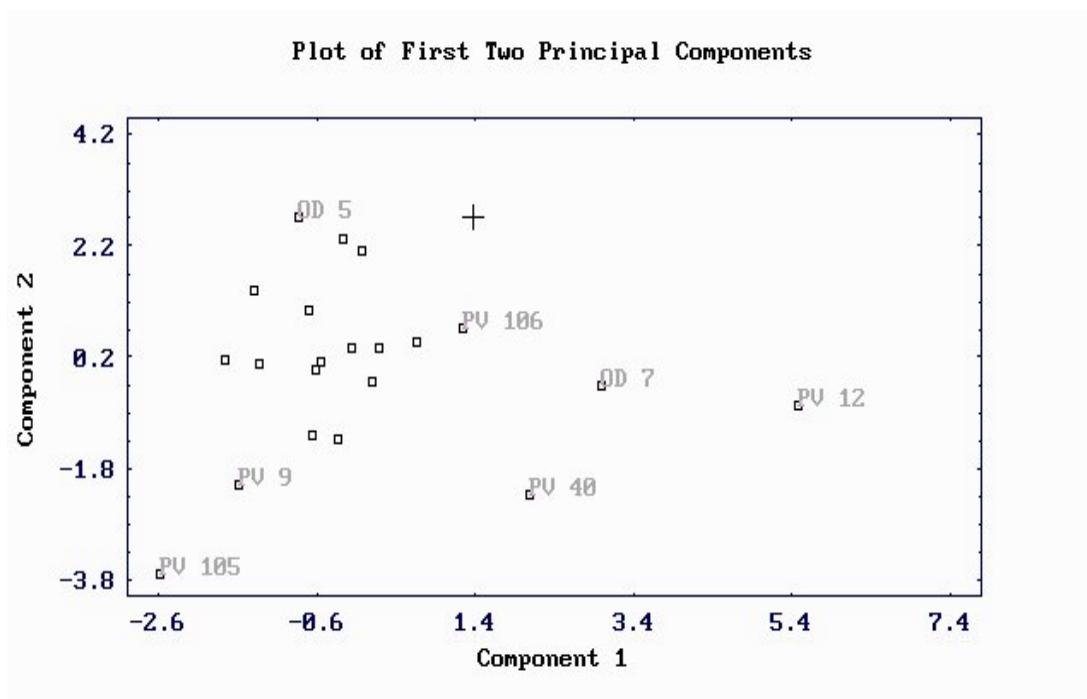
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	26.60107	26.60107
2	20.14378	46.74485
3	17.59076	64.33561
4	11.33636	75.67197
5	8.75419	84.42616
6	6.61726	91.04342
7	3.84493	94.88835
8	2.15340	97.04174
9	1.95821	98.99996
10	.72737	99.72732
11	.23827	99.96560
12	.03440	100.00000

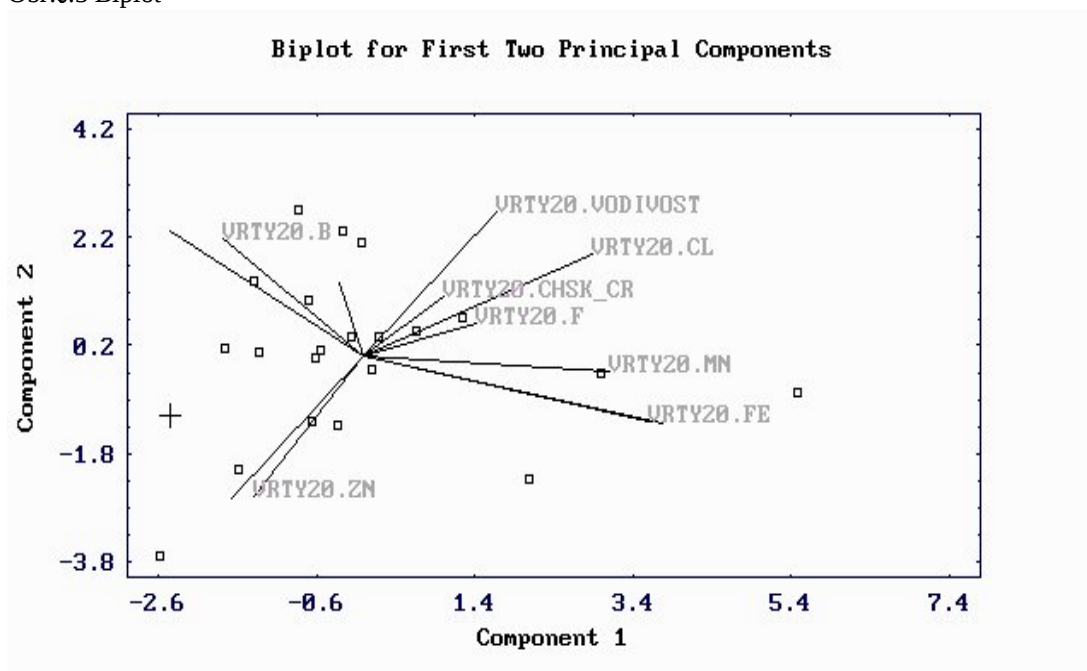
Obr.č.1 Plot components weights



Obr.č.2 Scatterplot



Obr.č.3 Biplot

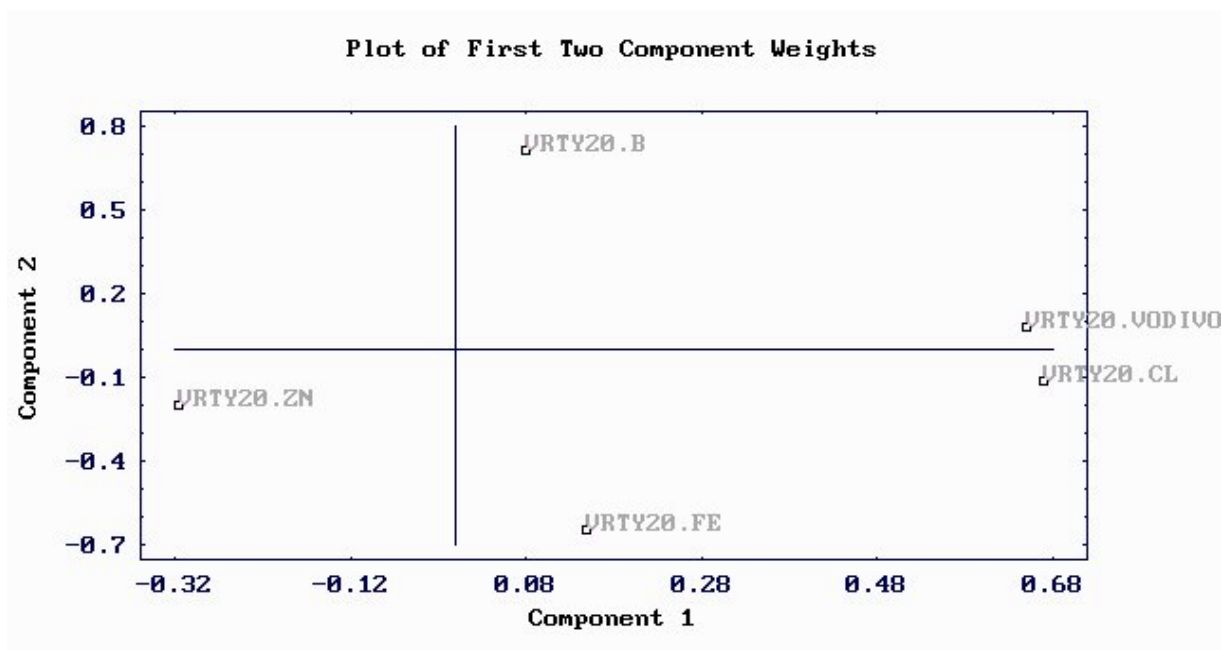


Tab. č. : 2 Analýza hlavních komponent (redukovaný počet)

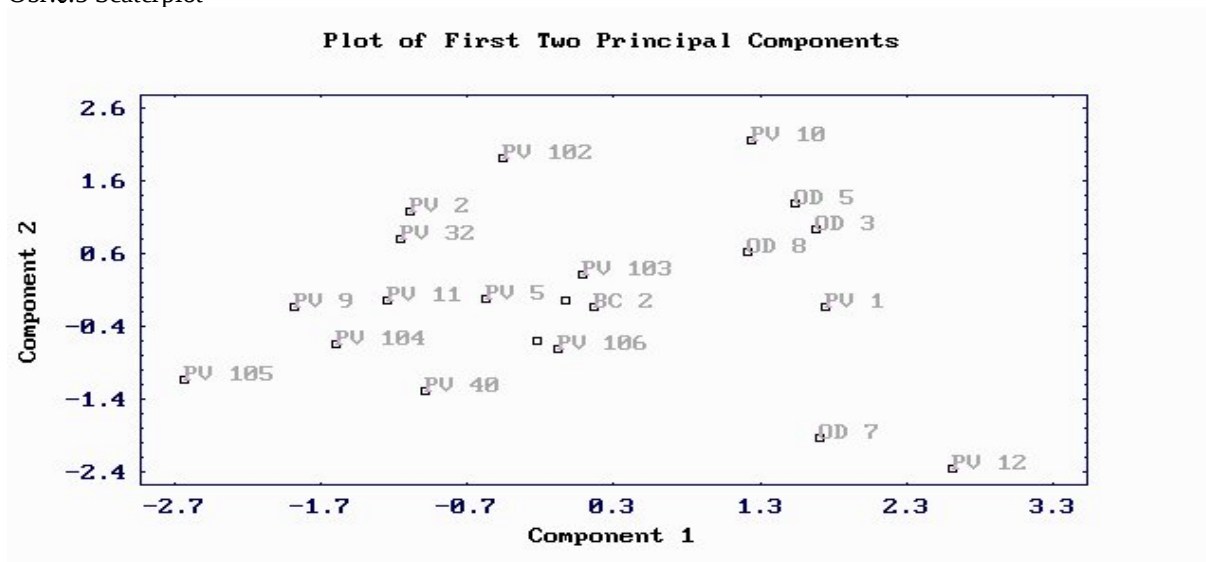
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	39.81035	39.81035
2	27.05304	66.86339
3	17.85064	84.71404
4	13.46849	98.18253
5	1.81747	100.00000

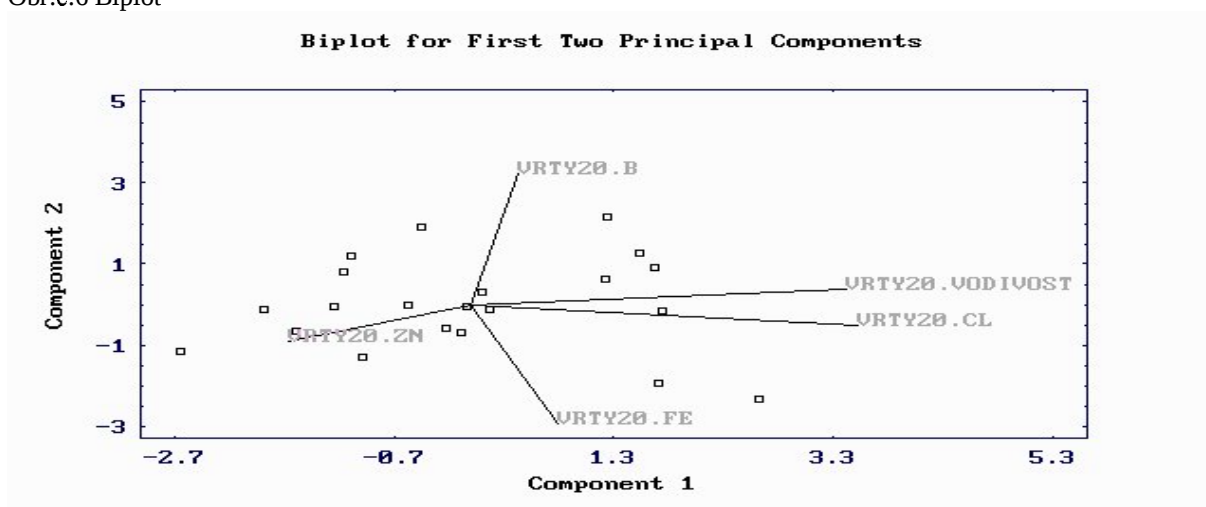
Obr.č.4 Plot components weights (reduk)



Obr.č.5 Scatterplot



Obr.č.6 Biplot



Metodou PCA byl snížen počet proměnných na 5 základních postačujících k vysvětlení zdrojové matice.

- Faktorová analýza

Metody s latentními proměnnými a klasifikační metody

Faktorová analýza, podobně jako analýza hlavních komponent patří mezi metody redukce počtu proměnných. Na rozdíl od komponentní analýzy lze při faktorové analýze vysvětlit závislosti proměnných. Vstupní data byla standardizována, podobně jako u analýzy hlavních komponent. Jako typ rotace byla zvolena metoda Equimax, která spojuje kritéria metody Varimax a Quartimax, což znamená, že při rotaci je maximalizován rozptyl čtvercových vah a zároveň je maximalizován součet čtvrtých mocnin faktorových zátěží. V tab.č.3 jsou ve druhém sloupci uvedeny odhady komunalit(čtverce indexu korelace), příslušná charakteristická čísla, dále procenta variability v pátém sloupci a v posledním sloupci kumulativně. Komunalita představuje váhu s jakou jednotlivé manifestní proměnné přispívají ke konstrukci příslušných latentních proměnných (nejvíc přispívají Vodivost a Cl, nejméně Fe a Zn). V tab.č.4 je uvedena matice faktorových vah, její jednotlivé prvky představují kovariance mezi jednotlivými proměnnými(v řádcích) a společnými faktory (ve sloupcích).

První faktor je korelován proměnnými vodivost a Cl.

Druhý faktor je korelován proměnnými B a Fe.

Třetí faktor hlavně proměnnou Zn.

V tab.č.5 jsou uvedeny odhady komunalit odpovídající faktorové matici v tab.č.4. Na celkové společné vysvětlené variabilitě se nejvíce podílí Zn dále vodivost. V tab.č.6 a 7 je uveden výpis matice faktorových vah po rotaci. Ve srovnání s původní maticí je zřejmé, že se sloupce zjednodušily.

Tab.č.3 Faktorová analýza

Variable	Communality	Factor	Eigenvalue	Percent Var	Cum Percent

--					
VRTY20.VODIVOST	0.81041	1	1.99052	39.8	39.8
VRTY20.CL	0.82040	2	1.35265	27.1	66.9
VRTY20.B	0.37641	3	.89253	17.9	84.7
VRTY20.FE	0.11689	4	.67342	13.5	98.2
VRTY20.ZN	0.21437	5	.09087	1.8	100.0

Tab.č.4 Matice faktorových vah

Factor Matrix

Variable/Factor		1	2	3
VRTY20.VODIVOST	0.91601	0.09536	0.33245	
VRTY20.CL	0.94596	-0.13454	0.11462	
VRTY20.B	0.11368	0.83280	0.06058	
VRTY20.FE	0.21080	-0.75914	-0.17816	
VRTY20.ZN	-0.44637	-0.23583	0.85642	

Tab.č. 5 Odhady komunalit faktorové matice

Variable	Est Communality

VRTY20.VODIVOST	0.95869
VRTY20.CL	0.92607
VRTY20.B	0.71015
VRTY20.FE	0.65247
VRTY20.ZN	0.98833

Tab.č.6 Matice faktorových vah po rotaci

EQUIMAX ROTATED FACTOR MATRIX

Variable/Factor	1	2	3
VRTY20.VODIVOST	0.97270	-0.10621	-0.03565
VRTY20.CL	0.92724	0.16254	-0.19969
VRTY20.B	0.11422	-0.82128	-0.15034
VRTY20.FE	0.14630	0.78977	-0.08562
VRTY20.ZN	-0.11002	0.04175	0.98716

Tab.č.7 Transformační matice

FINAL TRANSITION MATRIX

Factor	1	2	3
1	0.93497	0.05574	-0.35033
2	-0.01625	-0.97981	-0.19927
3	0.35436	-0.19201	0.91518

Faktorově čisté zvláště po rotaci vycházejí proměnné Vodivost a Cl (první faktor). Proměnné B (2. faktor), Zn (3. faktor), vycházejí faktorově nejčistší, což znamená, že jsou nejméně závislé a podobné proměnným Vodivost a Cl a sobě navzájem.

- Shluková analýza

Shlukovou analýzou byla zkoumána podobnost vícerozměrného objektu a provedeno roztrídění do skupin.

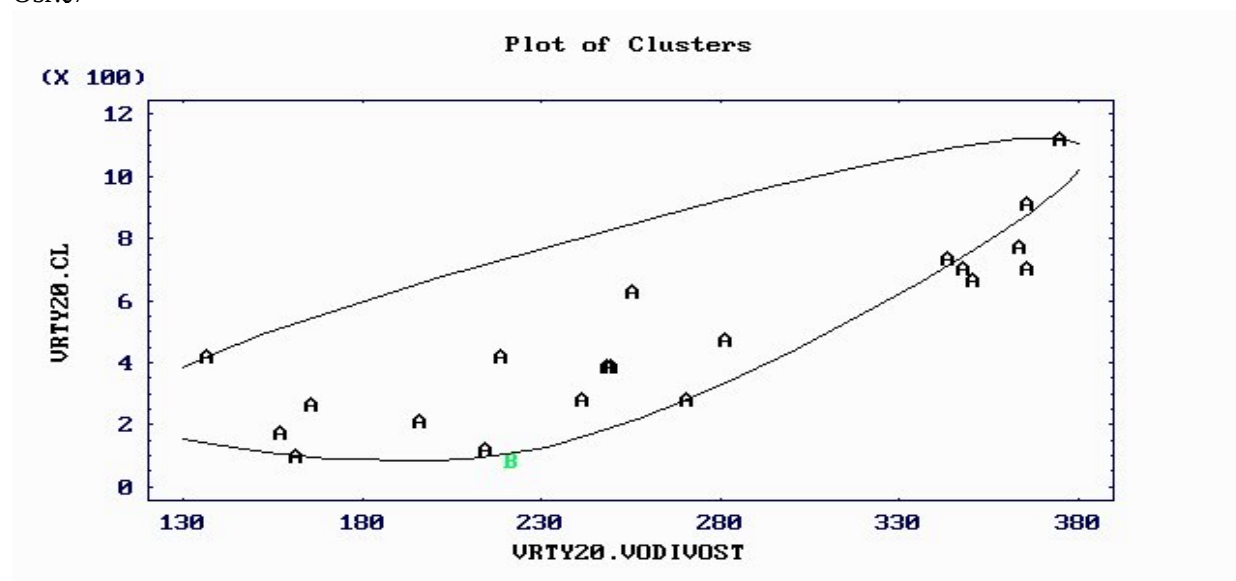
Program Statgraphics

Soubor byl zpracován eukleidovskou metrikou metodou nejbližšího souseda (Nearest), kde kritériem pro spojování shluků je minimum z možných mezishlukových vzdáleností objektů. Vstupní proměnné jsou vyjádřeny v různých měrných jednotkách, proto byla provedena standardizace. Jak vyplývá z Tabulky č.8 a grafů na obr. č.7-8 bylo do jednoho shluku zařazeno 20 vzorků a do druhého 1 vzorek.

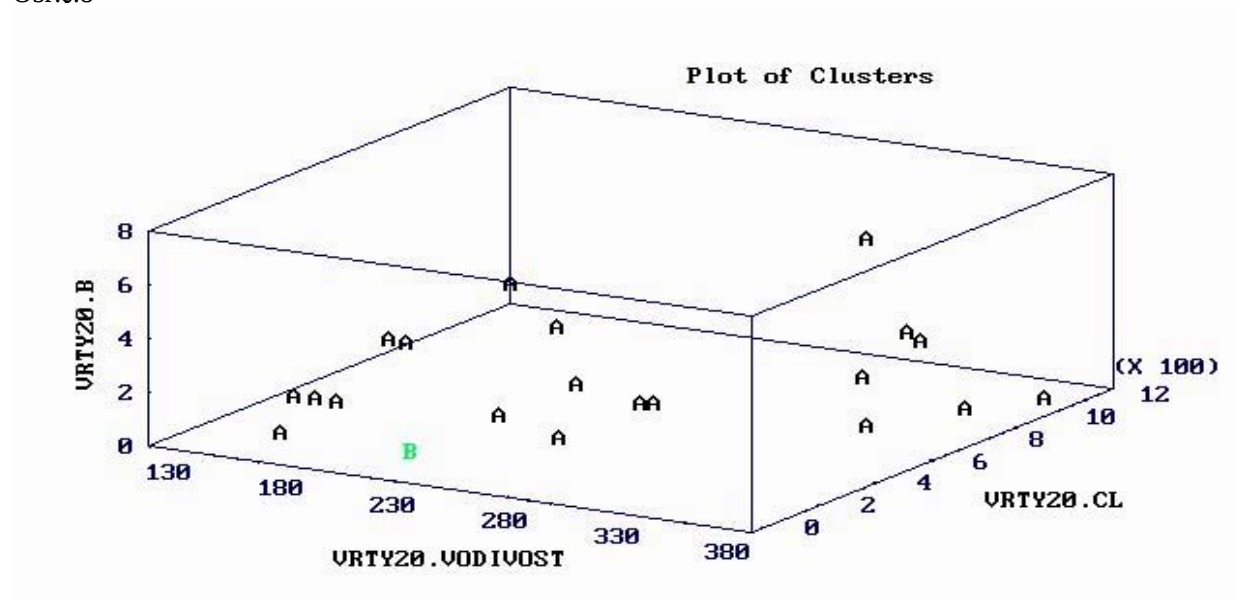
Tab č.8 Results of Clustering by Single Linkage or Nearest Neighbor Method

Cluster	Frequency	Percentage
1	20	95.2381
2	1	4.7619

Obr.č7



Obr.č.8



Program Scan

Tab.č.9 Hierarchical Cluster Analysis

Average Linkage
Euclidean Distance
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	20	98.10	20.111	4 16	4	2
2	19	96.26	39.498	1 3	1	2
3	18	95.89	43.382	7 9	7	2
4	17	94.78	55.147	13 15	13	2
5	16	94.72	55.766	4 6	4	3
6	15	94.08	62.562	7 21	7	3
7	14	92.64	77.780	13 17	13	3
8	13	92.33	81.030	1 14	1	3

Metody s latentními proměnnými a klasifikační metody

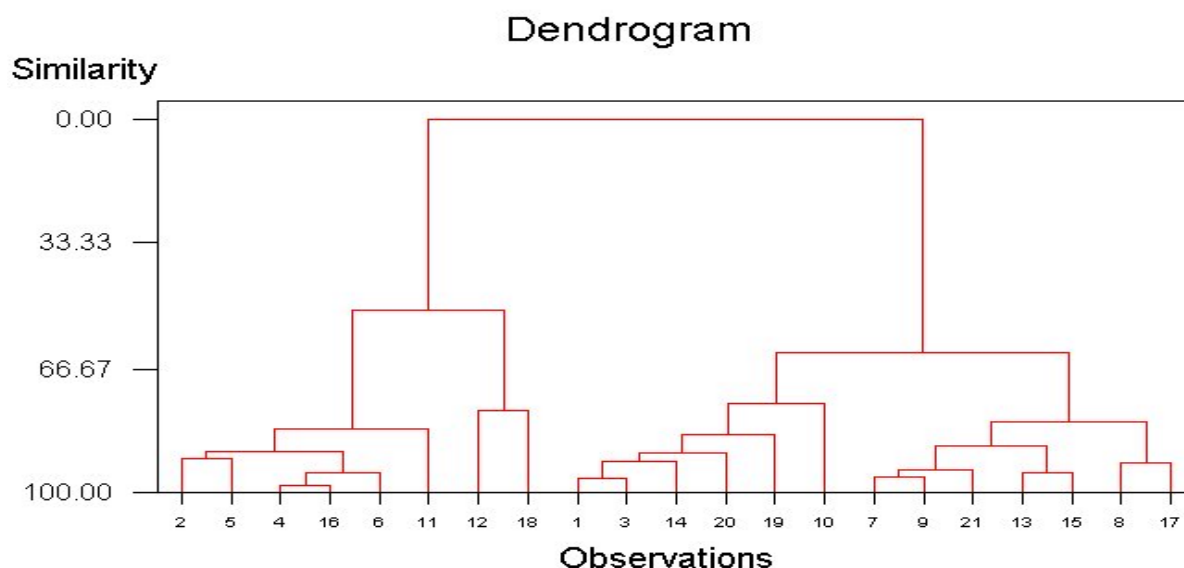
9	12	91.87	85.911	2	4	2	4
10	11	90.51	100.293	1	20	1	4
11	10	90.36	101.798	8	13	8	4
12	9	90.08	104.814	2	5	2	5
13	8	87.70	129.915	1	19	1	5
14	7	86.63	141.269	2	11	2	6
15	6	86.54	142.171	7	8	7	7
16	5	80.24	208.723	1	10	1	6
17	4	79.28	218.863	2	12	2	7
18	3	76.51	248.155	1	7	1	13
19	2	61.71	404.516	2	18	2	8
20	1	49.93	528.992	1	2	1	21

Final Partition

Number of clusters: 1

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	21	1821957.125	257.718	667.509

Obr.č.9 Dendrogram



Dendrogram viz Obr č.9 a tabulka č.9 znázorňují strukturu tvorby výsledného shluku. Nejvíce podobné jsou objekty, 4 a 16 (z 98,1 %). Při 70%-ní podobnosti tedy v podstatě existují 4 shluky.

Závěr CA: Objekty 2, 5, 4, 16, 6 a 11 mají více jak 75 % společných vlastností. Objekty 12 a 18 mají více jak 73 % společných vlastností. Objekty 1, 3, 20, 14, 19 a 10 mají více jak 69 % společných vlastností. Objekty 7, 9, 21, 13, 15, 8 a 17 mají společných více jak 74 % vlastností.

- Výpočet korelační matice

Výsledkem korelační analýzy je korelační matice uvedená v tab.č. 10a,b pro každou dvojici proměnných je uveden korelační koeficient, tj. číslo popisující míru souvislosti mezi dvěma proměnnými. Korelační matice byla vypočtena pro 21 vzorků. V prvním řádku a prvním sloupci jsou uvedeny názvy proměnných. V prvním řádku každé proměnné jsou uvedeny korelační koeficienty. Na diagonále jsou jednotky, protože souvislost každé proměnné se sebou samou je samozřejmě absolutní. V druhém řádku každé proměnné je uvedena hladina významnosti, tj. pravděpodobnost, že při skutečné nulové korelaci by empiricky zjištěný korelační koeficient mohl dosáhnout nynější hodnoty. Je-li toto číslo malé (menší než 0,05), pak lze usoudit, že mezi příslušnými proměnnými existuje nějaká souvislost. Její síla je posuzována podle velikosti korelačního koeficientu. Je-li velikost korelačního koeficientu menší než 0,5 nelze mezi proměnnými hovořit o závislosti.

Tab.č.10a Korelační matice

--						
	PH	VODIVOST	F	CL	CHSK_CR	B
PH	1.0000	.1902	.0136	-.0939	.0641	.3118
	.0000	.4090	.9532	.6857	.7826	.1688
VODIVOST	.1902	1.0000	.0875	.8452	.2487	.2038
	.4090	.0000	.7061	.0000	.2771	.3757
F	.0136	.0875	1.0000	.2260	-.1213	.1454
	.9532	.7061	.0000	.3246	.6003	.5293
CL	-.0939	.8452	.2260	1.0000	-.0277	-.0764
	.6857	.0000	.3246	.0000	.9050	.7420
CHSK_CR	.0641	.2487	-.1213	-.0277	1.0000	.0789
	.7826	.2771	.6003	.9050	.0000	.7339
B	.3118	.2038	.1454	-.0764	.0789	1.0000
	.1688	.3757	.5293	.7420	.7339	.0000
BA	-.5339	.0847	.3751	.4464	-.0975	-.4271
	.0127	.7151	.0938	.0425	.6741	.0535
CU	-.0468	-.1982	-.1127	-.3327	-.0730	-.1229
	.8404	.3890	.6267	.1406	.7533	.5955
FE	-.5495	.0770	.1770	.1780	.4122	-.3043
	.0099	.7399	.4426	.4400	.0634	.1799
MN	-.3296	.1975	.0929	.2939	.3887	-.3897
	.1445	.3908	.6886	.1960	.0816	.0807
ZN	-.0547	-.1532	-.1509	-.3000	-.0679	-.1398
	.8137	.5073	.5138	.1864	.7698	.5457
SN	.3689	.2350	-.0078	-.0945	.6217	.0719
	.0998	.3051	.9732	.6836	.0026	.7567

Tab.č.10b Korelační matice

--						
	BA	CU	FE	MN	ZN	SN
PH	-.5339	-.0468	-.5495	-.3296	-.0547	.3689
	.0127	.8404	.0099	.1445	.8137	.0998
VODIVOST	.0847	-.1982	.0770	.1975	-.1532	.2350
	.7151	.3890	.7399	.3908	.5073	.3051
F	.3751	-.1127	.1770	.0929	-.1509	-.0078
	.0938	.6267	.4426	.6886	.5138	.9732
CL	.4464	-.3327	.1780	.2939	-.3000	-.0945
	.0425	.1406	.4400	.1960	.1864	.6836
CHSK_CR	-.0975	-.0730	.4122	.3887	-.0679	.6217
	.6741	.7533	.0634	.0816	.7698	.0026
B	-.4271	-.1229	-.3043	-.3897	-.1398	.0719

Metody s latentními proměnnými a klasifikační metody

	.0535	.5955	.1799	.0807	.5457	.7567
BA	1.0000	-.1736	.7419	.3657	-.0871	-.3263
	.0000	.4516	.0001	.1030	.7074	.1489
CU	-.1736	1.0000	-.0522	-.1143	.9300	.0315
	.4516	.0000	.8223	.6217	.0000	.8922
FE	.7419	-.0522	1.0000	.5843	-.0099	.0233
	.0001	.8223	.0000	.0054	.9662	.9202
MN	.3657	-.1143	.5843	1.0000	-.0137	.3388
	.1030	.6217	.0054	.0000	.9530	.1330
ZN	-.0871	.9300	-.0099	-.0137	1.0000	.1606
	.7074	.0000	.9662	.9530	.0000	.4869
SN	-.3263	.0315	.0233	.3388	.1606	1.0000
	.1489	.8922	.9202	.1330	.4869	.0000

Z korelační matice lze vyčíst následující závislosti. Poměrně vysoká závislost je patrná mezi Cu a Zn , Vodivostí a chloridy. Dále byla zjištěna korelace i mezi CHSK_Cr a Sn, Fe a Ph, Ba a Ph. Ostatní závislosti jsou statisticky nevýznamné.

- Výpočet kovarianční matice

Výsledkem kovarianční analýzy je kovarianční matice uvedená v tab.č.11a,b,c. V hodnotě kovariance se odráží sice variabilita jednotlivých proměnných, ale nelze bezprostředně usuzovat na míru souvislosti. Používá se především jako vstup pro další výpočty.

Tab.č.11a Kovarianční matice

	PH	VODIVOST	F	CL
PH	0.37616	9.17141	0.01374	-16.5584
VODIVOST	9.17141	6183.13	11.2942	19118.1
F	0.01374	11.2942	2.69432	106.717
CL	-16.5584	19118.1	106.717	82748.3
CHSK_CR	1.30429	648.966	-6.61086	-264.717
B	0.41451	34.7289	0.51747	-47.6458
BA	-0.05530	1.12446	0.10397	21.6837
CU	-0.00102	-0.55238	-0.00655	-3.39104
FE	-10.7453	193.167	9.26564	1632.92
MN	-1.28641	98.8180	0.97075	537.963
ZN	-0.00263	-0.94394	-0.01941	-6.76209
SN	0.04825	3.94084	-0.00273	-5.79881

Tab.č.11b Kovarianční matice

	CHSK_CR	B	BA	CU
PH	1.30429	0.41451	-0.05530	-0.00102
VODIVOST	648.966	34.7289	1.12446	-0.55238
F	-6.61086	0.51747	0.10397	-0.00655

Metody s latentními proměnnými a klasifikační metody

CL	-264.717	-47.6458	21.6837	-3.39104
CHSK_CR	1101.56	5.67560	-0.54652	-0.08581
B	5.67560	4.69809	-0.15630	-0.00944
BA	-0.54652	-0.15630	0.02851	-0.00104
CU	-0.08581	-0.00944	-0.00104	0.00126
FE	436.160	-21.0302	3.99419	-0.05896
MN	82.0828	-5.37488	0.39297	-0.02577
ZN	-0.17667	-0.02374	-0.00115	0.00258
SN	4.40021	0.03325	-0.01175	0.00024

--

Tab.č.11c Kovarianční matice

--

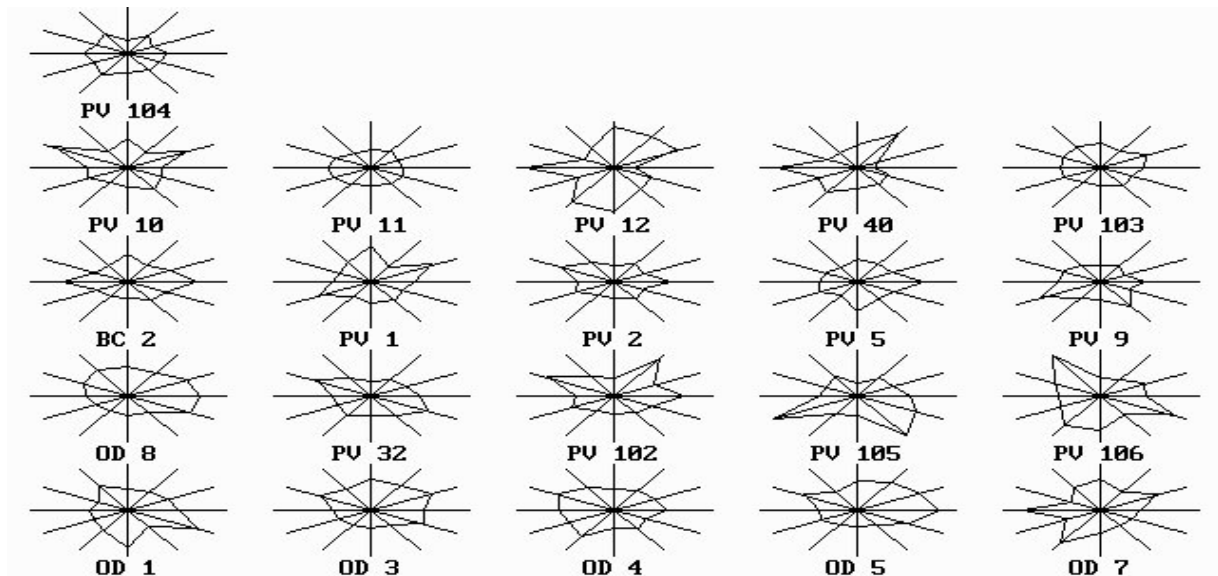
	FE	MN	ZN	SN
PH	-10.7453	-1.28641	-0.00263	0.04825
VODIVOST	193.167	98.8180	-0.94394	3.94084
F	9.26564	0.97075	-0.01941	-0.00273
CL	1632.92	537.963	-6.76209	-5.79881
CHSK_CR	436.160	82.0828	-0.17667	4.40021
B	-21.0302	-5.37488	-0.02374	0.03325
BA	3.99419	0.39297	-0.00115	-0.01175
CU	-0.05896	-0.02577	0.00258	0.00024
FE	1016.55	118.543	-0.02462	0.15833
MN	118.543	40.4874	-0.00683	0.45971
ZN	-0.02462	-0.00683	0.00614	0.00268
SN	0.15833	0.45971	0.00268	0.04548

--

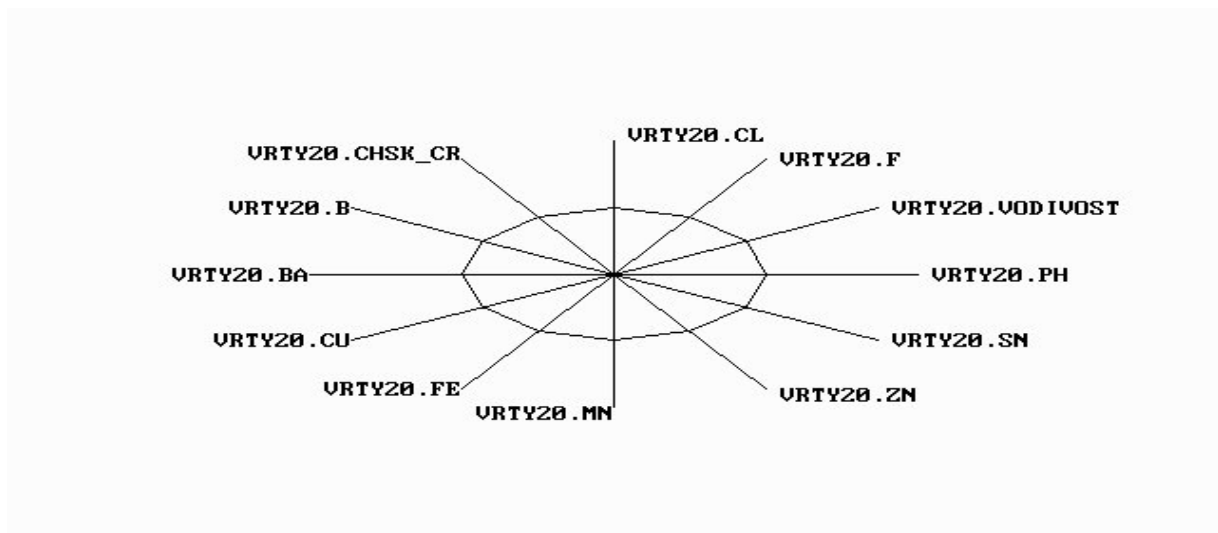
- Grafické zkoumání podobnosti objektů

Sun Ray Plot – graf slunečních paprsků

Graf slouží k vizuálnímu srovnání zadaných vícerozměrných objektů. Procedura Statgraphicsu Sun Ray Plot nakreslí pro každý vzorek hvězdici, která se skládá z paprsků spojených ve společném bodě a úseček mezi paprsky, které tvoří polygon. Počet paprsků odpovídá počtu proměnných. Střed polygonu představuje průměr odpovídající proměnné a délka paprsku je 2n násobek směrodatné odchylky dané proměnné. Na obr.č. 10 je 21 polygonů testovaných vzorků, klíč je znázorněn na obr.č. 11.



Obr.č.10 Graf slunečních paprsků



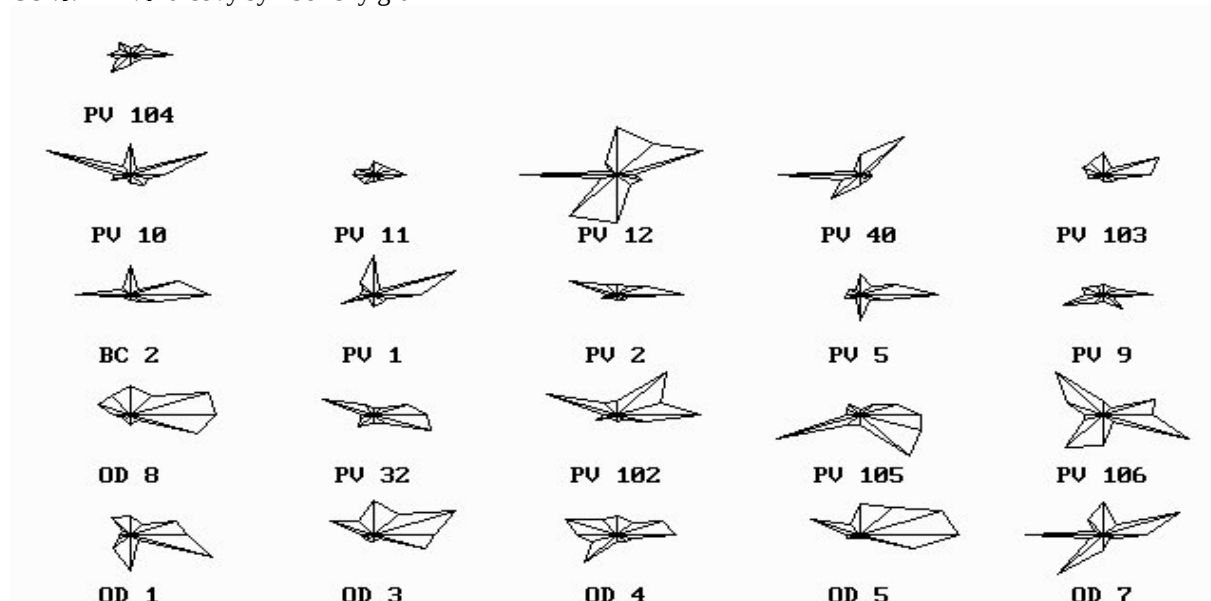
Obr.č.11 Klíč ke grafu slunečních paprsků

Dle vizuálního posouzení jednotlivých polygonů je zřejmé, že z hlediska jejich velikosti lze zaznamenat jejich podobnost. Hledáním podobnosti na základě tvaru lze usoudit na podobnost vrtů OD5 a OD8.

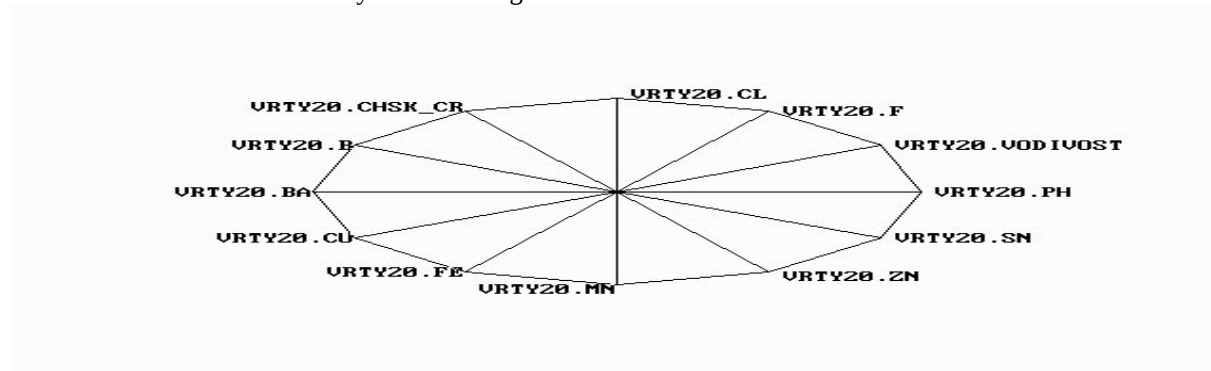
Star Symbol Plot – hvězdicový symbolický graf

Graf podobně jako předchozí slouží rovněž k vizuálnímu zobrazení podobnosti vícerozměrných objektů. Paprsky se spojují v jednom bodě. Nejkratší paprsek signalizuje, že příslušná proměnná má nejmenší hodnotu a nejdelší paprsek informuje o nejvyšší hodnotě příslušné proměnné. Na obr.č. 12 je 21 polygonů testovaných vzorků, klíč je znázorněn na obr.č.13.

Obr.č.12 Hvězdicový symbolický graf



Obr.č.13 Klíč k hvězdicovému symbolickému grafu



Podobně jako u předchozího grafu lze konstatovat, že polygony OD5 a OD8 jeví určitou podobnost. Zhodnocení souboru na základě tvaru polygonu je ještě více zřetelnější než u grafu slunečních paprsků.

Závěr:

Metoda hlavních komponent

Metodou PCA byl snížen počet proměnných na 5 základních postačujících k vysvětlení zdrojové matice(B, Zn, Fe, Cl, Vodivost).

Faktorová analýza

Faktorově čisté zvláště po rotaci vycházejí proměnné Vodivost a Cl (první faktor). Proměnné B (2. faktor), Zn (3. faktor), vycházejí faktorově nejčistší, což znamená, že jsou nejméně závislé a podobné proměnným Vodivost a Cl a sobě navzájem.

Shluková analýza

Objekty 2, 5, 4, 16, 6 a 11 mají více jak 75 % společných vlastností. Objekty 12 a 18 mají více jak 73 % společných vlastností. Objekty 1, 3, 20, 14, 19 a 10 mají více jak 69 % společných vlastností. Objekty 7, 9, 21, 13, 15, 8 a 17 mají společných více jak 74 % vlastností.

Z korelační matice lze vyčíst následující závislosti. Poměrně vysoká závislost je patrná mezi Cu a Zn , Vodivostí a chloridy. Dále byla zjištěna korelace i mezi CHSK_Cr a Sn, Fe a Ph, Ba a Ph. Ostatní závislosti jsou statisticky nevýznamné.

Grafické zkoumání podobnosti objektů

Sun Ray Plot – graf slunečních paprsků i Star Symbol Plot – hvězdicový symbolický graf

Dle vizuálního posouzení jednotlivých polygonů je zřejmé, že z hlediska jejich velikosti lze zaznamenat jejich podobnost. Hledáním podobnosti na základě tvaru lze usoudit na podobnost vrtů OD5 a OD8.

Je možné redukovat počet odběrových míst (vrtů) z 21 na 20. (Vrtů OD5 a OD8 jsou podobné). Jednotlivé vrtů lze charakterizovat 5 proměnnými B, Zn, Fe, Vodivost a Cl.

Příklad č.2**Zadání :**

Na 61 vzorcích povrchové vody , u kterých bylo provedeno 12 stanovení parametrů kvality, proveďte vícerozměrnou statistickou analýzu.

Rozhodněte, zda je možné charakterizovat jednotlivé odběry menším počtem stanovovaných ukazatelů.

Data :

Rada	VZ	PH	VODI	NL	RL	TVRD	CL	CHSKCR	SIRANY	BA	FE	MN	SR
1	10	7,470	296,000	10,000	1716,000	23,100	799,000	21,000	98,000	0,2860	0,5340	0,0940	1,8830
2	11	7,580	67,000	50,000	380,500	13,900	45,200	17,800	101,000	0,0770	0,6590	0,2300	0,2980
3	12	7,660	45,100	17,000	190,000	9,500	31,300	16,400	68,000	0,0750	0,9910	0,1090	0,2510
4	20	7,950	206,000	32,000	1473,000	21,300	418,000	21,000	161,000	0,2770	0,6350	0,1850	1,3180
5	23	8,040	129,500	26,000	650,000	18,200	202,000	18,800	181,000	0,1280	0,5970	0,0330	0,6000
6	25	7,600	56,700	15,000	407,000	11,900	45,200	16,300	119,000	0,1050	1,0100	0,1320	0,2260
7	356	7,280	161,300	37,000	815,500	18,100	310,000	20,000	127,000	0,1660	0,5290	0,2420	0,9330
8	359	6,980	283,000	24,000	1615,000	21,600	757,000	23,200	92,000	0,2010	0,3960	0,2230	1,6400
9	360	6,820	49,000	33,000	336,500	9,900	34,000	28,100	83,000	0,0590	0,8410	0,1680	0,1800
10	361	7,110	55,300	12,000	385,000	11,900	44,700	17,190	97,000	0,0670	0,4980	0,1680	0,1680
11	363	7,160	44,100	49,000	323,000	9,500	31,000	29,580	67,000	0,0500	0,7440	0,2470	0,1960
12	364	7,470	129,400	16,000	809,500	18,200	203,000	16,600	177,000	0,0970	0,1740	0,0630	0,5500
13	742	7,690	171,100	24,000	901,000	19,000	344,000	22,000	162,000	0,2240	0,3230	0,1200	1,0960
14	746	7,280	62,000	5,000	170,500	13,700	34,400	16,000	106,000	0,0780	0,4300	0,0370	0,2960
15	747	7,460	56,300	8,000	369,000	12,000	41,300	16,000	115,000	0,0980	0,1480	0,0490	0,2190
16	749	7,460	33,600	4,000	318,000	7,000	20,700	16,000	58,000	0,0650	0,2360	0,0480	0,1880
17	750	7,900	125,100	19,000	772,500	17,600	207,000	11,000	181,000	0,1210	0,2190	0,0900	0,6070
18	1155	7,770	164,700	14,000	862,000	18,900	313,000	38,650	179,000	0,2070	0,6820	0,2480	1,0940
19	1158	6,890	99,000	33,000	526,500	9,800	215,000	20,310	59,000	0,2600	0,7300	0,1920	0,9220
20	1159	6,790	60,700	10,000	686,000	13,200	34,800	24,440	104,000	0,0660	0,6390	0,1110	0,3040
21	1160	6,780	60,000	5,000	462,500	12,000	45,200	11,950	141,000	0,1040	7,7500	1,3400	0,2330
22	1162	6,850	27,800	13,000	273,000	5,600	13,900	21,030	52,000	0,0480	0,3090	0,0260	0,1540
23	1163	7,710	113,600	25,000	768,500	17,900	198,000	25,530	184,000	0,1130	0,3370	0,0800	0,6230
24	1537	7,760	266,000	24,000	1619,000	24,200	597,000	29,000	162,000	0,3590	1,2300	0,5150	2,0450
25	1540	7,570	257,000	20,000	1690,000	20,300	702,000	21,000	92,000	0,2220	0,7460	0,0310	1,7160
26	1541	7,650	65,500	19,000	475,000	14,400	45,600	26,000	100,000	0,0850	1,2600	0,9580	0,3170
27	1542	7,750	58,100	17,000	422,500	12,700	49,100	22,000	107,000	0,0920	0,5100	0,0520	0,2120
28	1544	7,540	29,000	5,000	197,500	7,000	21,100	17,000	48,000	0,0630	0,5830	0,0570	0,1880
29	1545	7,610	128,600	7,000	803,500	18,500	203,500	18,000	188,000	0,1410	0,3870	0,2690	0,6370
30	2049	7,200	191,200	43,000	1227,000	21,800	451,000	24,000	87,000	0,1960	1,4300	0,0640	1,0500
31	2050	7,440	68,300	33,000	425,500	13,000	48,600	26,000	82,000	0,1120	1,1250	0,5830	0,3190
32	2051	7,450	58,600	14,000	404,000	11,700	45,100	11,000	116,000	0,1250	0,6700	0,0420	0,2400
33	2053	7,240	37,400	20,000	291,500	7,500	27,800	17,000	54,000	0,0780	0,5010	0,1170	0,2160
34	2062	8,180	269,000	10,000	1547,000	23,900	695,000	28,000	171,000	0,3160	0,6780	0,1460	1,9100
35	2065	7,970	130,800	22,000	777,500	18,400	205,000	20,000	195,000	0,1460	0,2500	0,0480	0,6470
36	2303	7,530	314,000	66,500	1862,000	24,800	869,000	19,000	137,000	0,2330	0,8980	0,1690	1,9280
37	2304	7,680	70,400	18,000	442,500	13,200	45,200	25,000	94,000	0,0610	0,8730	0,5500	0,3070
38	2305	7,670	59,300	56,500	420,500	12,600	45,200	12,000	119,000	0,1010	1,8900	0,1540	0,2250
39	2307	7,570	46,700	58,000	319,000	10,400	31,300	21,000	73,000	0,0470	0,4270	0,2650	0,2440
40	2319	7,920	330,000	48,000	1341,000	27,100	834,000	30,000	178,000	0,4050	0,5710	0,1710	2,0620
41	2322	8,700	128,000	61,500	690,000	15,100	226,000	13,000	188,000	0,1370	0,3700	0,1090	0,7160
42	2738	7,970	131,500	59,000	710,000	17,400	209,000	39,000	166,000	0,2200	1,1100	0,1580	0,8740
43	2741	7,040	55,100	42,000	297,000	13,700	33,000	29,000	85,000	0,0870	0,9760	0,5020	0,2980
44	2742	7,360	56,900	84,000	307,000	11,800	50,300	19,000	101,000	0,1150	2,6400	0,2010	0,2240
45	2744	7,430	42,700	24,000	230,000	9,100	29,500	17,000	63,000	0,0860	0,3800	0,1060	0,2340

Metody s latentními proměnnými a klasifikační metody

46	3059	7,440	63,300	24,000	414,500	12,700	41,700	29,000	99,000	0,0750	0,6440	0,2960	0,2730
47	3060	7,560	58,300	24,000	465,500	11,800	48,700	18,000	123,000	0,1080	1,1800	0,1330	0,2000
48	3062	7,390	40,100	15,000	199,000	7,700	31,300	23,000	60,000	0,0490	0,1560	0,0170	0,1840
49	3063	8,160	127,700	12,000	808,500	17,600	205,000	22,000	191,000	0,1160	0,2320	0,0400	0,5660
50	3074	7,990	201,000	89,000	1232,000	22,300	417,000	32,000	175,000	0,2530	0,7110	0,1060	1,2240
51	3421	7,640	437,000	55,000	2533,000	29,500	1263,000	25,000	117,000	0,3370	0,7490	0,1410	2,7680
52	3422	7,630	67,200	55,000	410,000	13,100	45,630	33,000	89,000	0,0760	0,5030	0,1620	0,3140
53	3423	7,610	58,500	9,000	434,000	12,300	45,630	18,000	109,000	0,0960	0,4810	0,0370	0,2310
54	3424	7,590	38,700	27,000	285,000	7,560	24,570	29,000	58,000	0,0660	0,3880	0,0150	0,2130
55	3437	7,790	131,500	15,000	753,000	18,400	207,000	35,000	164,000	0,1310	0,3720	0,0470	0,6450
56	3991	7,950	223,000	39,000	1369,000	25,400	532,000	21,000	187,000	0,2880	8,9400	0,2520	1,5780
57	3994	7,650	327,000	67,000	1980,000	19,800	922,000	12,000	108,000	0,3210	0,5460	0,0600	2,5000
58	3995	7,370	62,100	43,000	328,000	12,800	33,680	13,000	103,000	0,0590	0,0320	0,0310	0,2780
59	3996	7,600	57,700	43,000	253,500	12,700	42,540	19,000	99,000	0,0950	0,5450	0,0850	0,2170
60	3997	8,100	36,900	18,000	645,000	9,660	21,270	17,000	38,000	0,0640	1,3000	0,1300	0,1860
61	3998	7,850	129,400	15,000	799,500	17,600	198,500	19,000	189,000	0,1370	0,4290	0,0310	0,6200

Program :

Statgraphics, Microsoft Word, Scan

Řešení :

- Metoda hlavních komponent

Cílem metody PCA je redukce počtu proměnných a zároveň získání co nejmenšího počtu hlavních komponent vysvětlujících maximální variabilitu původní matice. Metodou hlavních komponent jsme zpracovali výše uvedená data. Data byla standardizována, protože nejsou ve stejných jednotkách. V tabulce č. 12 jsou uvedeny výsledky. Ve druhém sloupci jsou uvedena procenta variability vysvětlená jednotlivými hlavními komponentami, tj. podíly rozptylů jednotlivých hlavních komponent na celkové variabilitě, v pořadí od nejvyššího k nejnižšímu. Ve třetím sloupci jsou tytéž podíly uvedeny kumulovaně a lze z nich určit optimální počet komponent. Většinou se požaduje, aby součet prvních nejvyšších komponent měl 85 až 92 %. Z tabulky č.12 vyplývá, že tento předpoklad splňuje 5 hlavní komponenta.

Na obr. č. 14 – 16 jsou znázorněny grafy metody hlavních komponent. Na obr.č.14 je znázorněn graf komponentních vah pro první dvě hlavní komponenty u jednotlivých proměnných vstupujících do analýzy. Na obr.č.15 je znázorněn bodový diagram, který zobrazuje hodnoty prvních dvou hlavních komponent u jednotlivých objektů. Tohoto grafu se využívá ke grafickému ověření předpokladu normality pro další analýzy. Na obr.č.16 je znázorněn dvojný graf, který kombinuje oba předchozí grafy. Body představují komponentní skóre jednotlivých objektů. Přímký spojující se v bodě (0,0) představují původní proměnné, délka každé přímký je úměrná svému příspěvku k prvním dvěma hlavním komponentám.

Z číselného výstupu (Tab.č12) je zřejmé, že 91,64% splňuje součet rozptylů 5 hlavních komponent. Z grafického výstupu obr.č.14-16 vyplývá, že můžeme redukovat počet proměnných na 5 (Ph, Vodivost, Tvrdost, Ba, Mn). Výsledky jsme uvedli do Tabulky č.13, ze které vyplývá, že 85% splňuje součet rozptylů 3 hlavních komponent. Na obr. č.17 – 19 jsou znázorněny grafy metody hlavních komponent pro zredukovaná data.

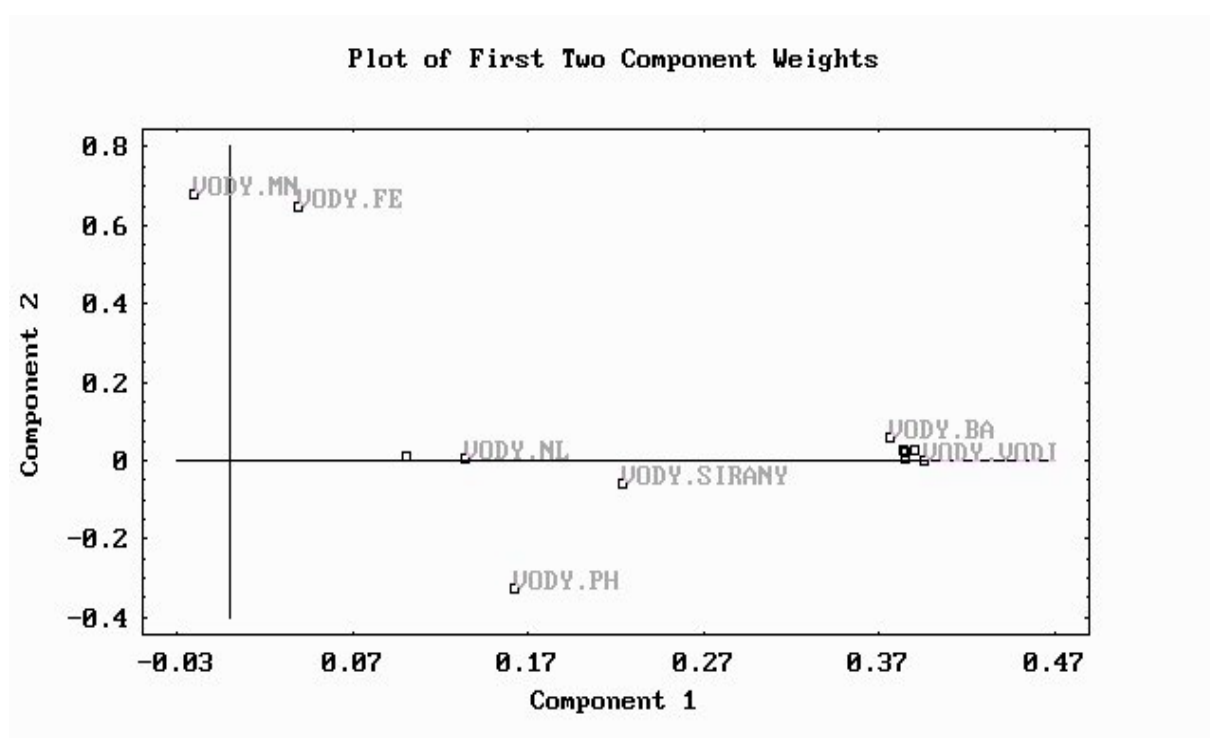
Tab. č. : 12 Analýza hlavních komponent

Principal Components Analysis

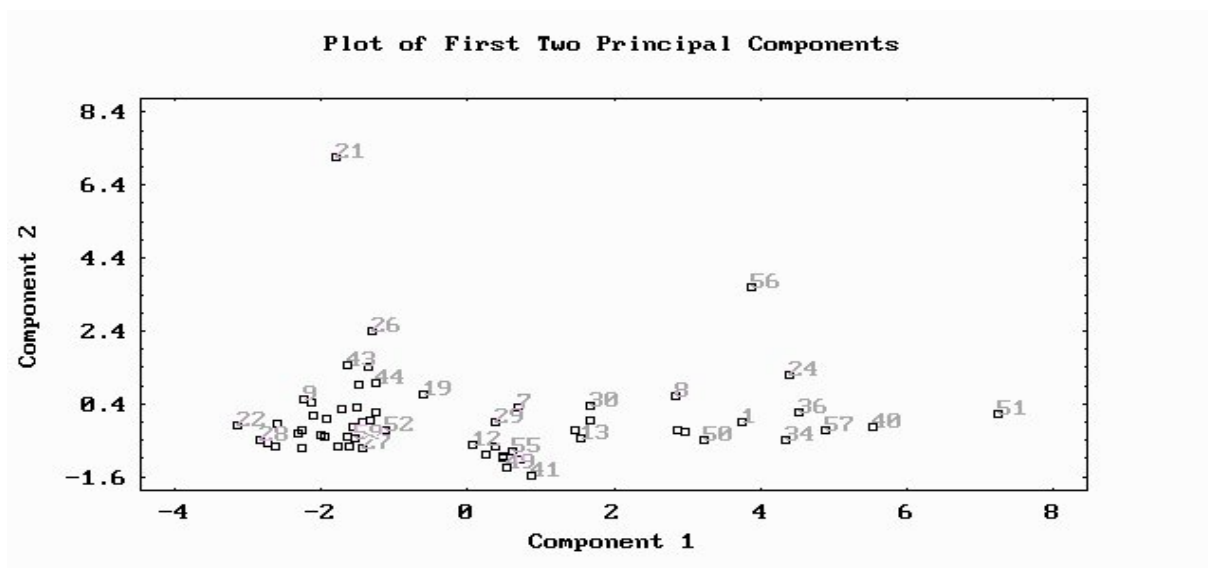
Metody s latentními proměnnými a klasifikační metody

Component Number	Percent of Variance	Cumulative Percentage
1	51.22548	51.22548
2	13.51319	64.73868
3	10.65533	75.39401
4	8.86823	84.26223
5	7.37401	91.63624
6	3.24648	94.88272
7	3.06613	97.94885
8	1.31287	99.26172
9	.41236	99.67408
10	.25179	99.92587
11	.06573	99.99160
12	.00840	100.00000

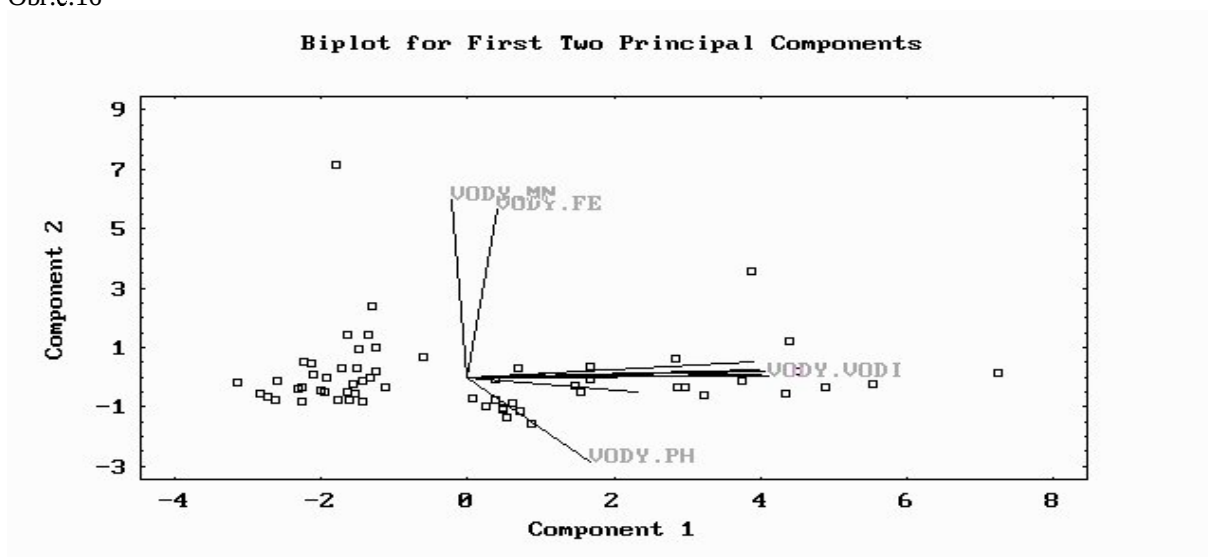
Obr.č.14 Plot components weights



Obr.č.15 Scatterplot



Obr.č.16

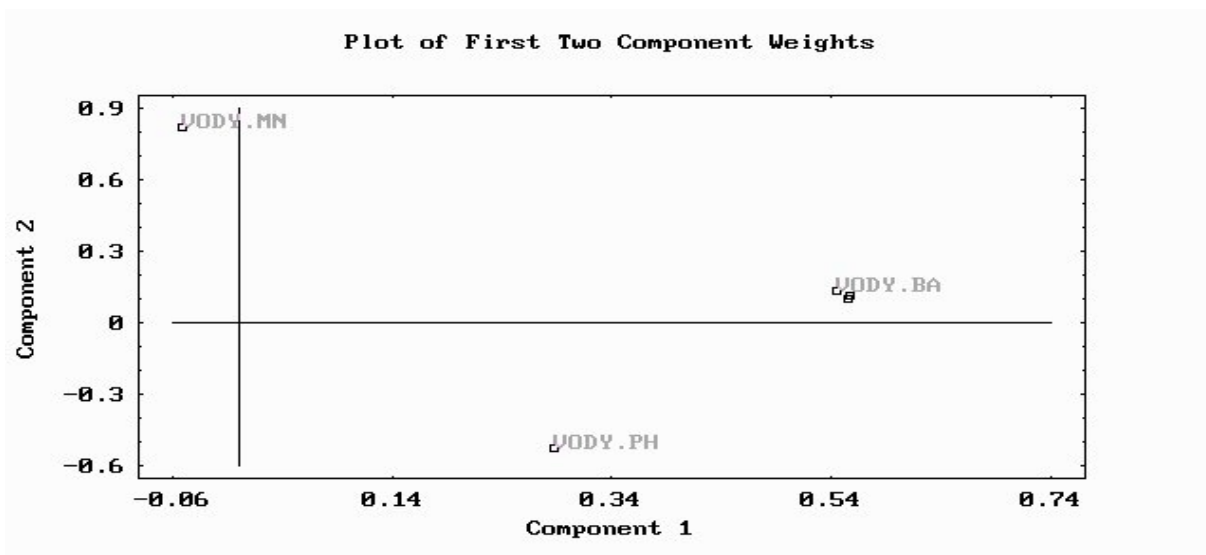


Tab. č. : 13 Analýza hlavních komponent

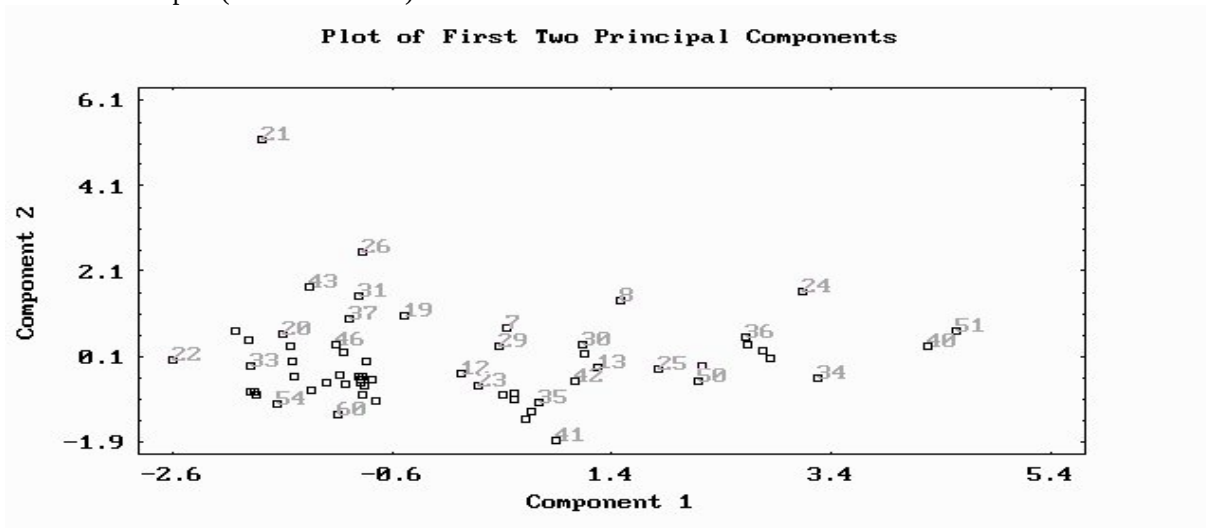
Principal Components Analysis

Component Number	Percent of Variance	Cumulative Percentage
1	59.00790	59.00790
2	22.81961	81.82751
3	14.00464	95.83215
4	3.09486	98.92701
5	1.07299	100.00000

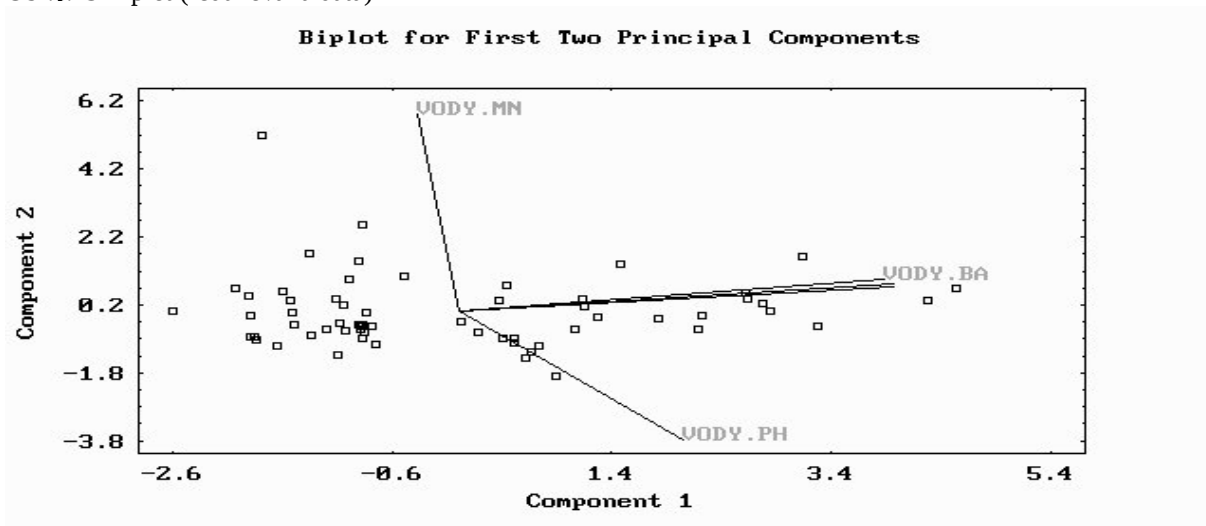
Obr.č.17 Plot components weights (redukována data)



Obr.č.18 Scatterplot (redukovaná data)



Obr.č.19 Biplot (redukovaná data)



Metodou PCA byl snížen počet proměnných na 5 základních postačujících k vysvětlení zdrojové matice.

- Faktorová analýza

Metody s latentními proměnnými a klasifikační metody

Faktorová analýza, podobně jako analýza hlavních komponent patří mezi metody redukce počtu proměnných. Na rozdíl od komponentní analýzy lze při faktorové analýze vysvětlit závislosti proměnných. Vstupní data byla standardizována, podobně jako u analýzy hlavních komponent. Jako typ rotace byla zvolena metoda Equimax, která spojuje kritéria metody Varimax a Quartimax, což znamená, že při rotaci je maximalizován rozptyl čtvercových vah a zároveň je maximalizován součet čtvrtých mocnin faktorových zátěží. V tab.č.14 jsou ve druhém sloupci uvedeny odhady komunalit(čtverce indexu korelace), příslušná charakteristická čísla, dále procenta variability v pátém sloupci a v posledním sloupci kumulativně. Komunalita představuje váhu s jakou jednotlivé manifestní proměnné přispívají ke konstrukci příslušných latentních proměnných (nejvíc přispívají Vodivost a Tvrdost, nejméně Ph a Mn). V tab.č.15 je uvedena matice faktorových vah, její jednotlivé prvky představují kovariance mezi jednotlivými proměnnými(v řádcích) a společnými faktory (ve sloupcích). První faktor je korelován proměnnými vodivost, Tvrdost a Ba. Druhý faktor je korelován proměnnou Mn. Třetí faktor hlavně proměnnou Ph. V tab.č.16 jsou uvedeny odhady komunalit odpovídající faktorové matici v tab.č.15. Na celkové společné vysvětlené variabilitě se nejvíce podílí Mn dále Ph. V tab.č.17 a 18 je uveden výpis matice faktorových vah po rotaci. Ve srovnání s původní maticí je zřejmé, že se sloupce zjednodušily.

Tab. č. : 14 Faktorová analýza

Variable	Communality	Factor	Eigenvalue	Percent Var	Cum Percent

--					
VODY.PH	0.29585	1	2.95039	59.0	59.0
VODY.VODI	0.91425	2	1.14098	22.8	81.8
VODY.TVRD	0.87163	3	.70023	14.0	95.8
VODY.BA	0.82494	4	.15474	3.1	98.9
VODY.MN	0.13860	5	.05365	1.1	100.0

Tab. č. : 15 Matice faktorových vah

Factor Matrix

Variable/Factor	1	2	3

VODY.PH	0.49431	-0.56690	0.65790
VODY.VODI	0.95613	0.12429	-0.18723
VODY.TVRD	0.95386	0.10924	-0.00243
VODY.BA	0.93489	0.14526	-0.10893
VODY.MN	-0.08953	0.87813	0.46955

Tab. č. : 16 Odhady komunalit faktorové matice

Variable	Est Communality

VODY.PH	0.99856
VODY.VODI	0.96468
VODY.TVRD	0.92178
VODY.BA	0.90698
VODY.MN	0.99961

Tab. č. : 17 Matice faktorových vah po rotaci

EQUIMAX ROTATED FACTOR MATRIX

Metody s latentními proměnnými a klasifikační metody

Variable/Factor	1	2	3
VODY.PH	0.14757	0.98037	-0.12508
VODY.VODI	0.97188	0.13149	-0.05324
VODY.TVRD	0.91663	0.28304	0.03819
VODY.BA	0.93618	0.17454	0.00916
VODY.MN	0.00330	-0.11156	0.99355

Tab. č. : 18 Transformační matice

FINAL TRANSITION MATRIX

Factor	1	2	3
1	0.93257	0.35706	-0.05311
2	0.24204	-0.50932	0.82584
3	-0.26782	0.78301	0.56140

Faktorově čisté zvláště po rotaci vycházejí proměnné Vodivost, Ba a Tvrdost (první faktor). Proměnné Ph (2. faktor), Mn (3. faktor), vycházejí faktorově nejčistší, což znamená, že jsou nejméně závislé a podobné proměnným Vodivost, Ba a Tvrdost a sobě navzájem.

- Shluková analýza

Shlukovou analýzou byla zkoumána podobnost vícerozměrného objektu a provedeno roztrídění do skupin.

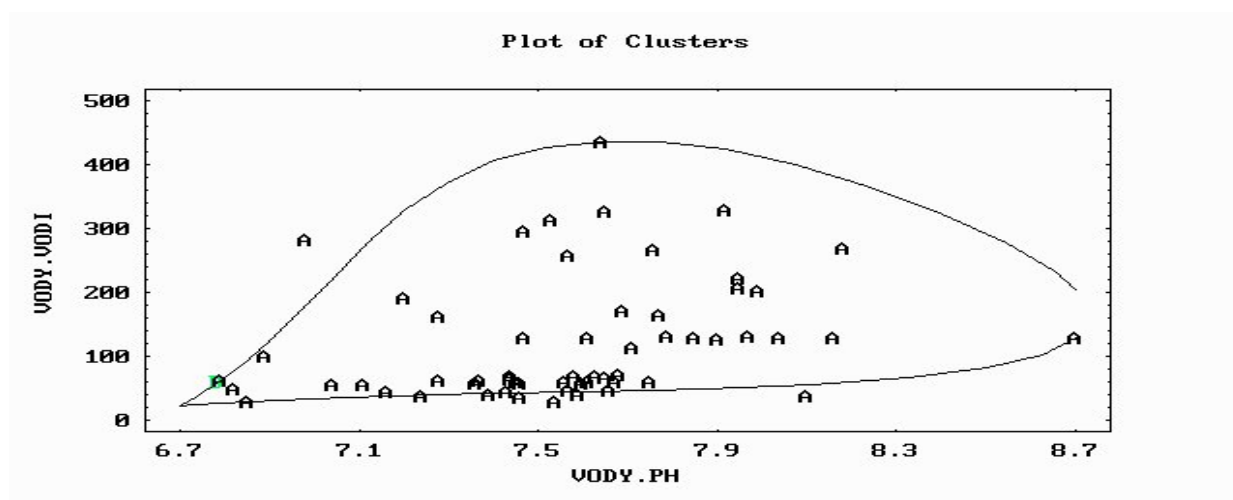
Program Statgrahics

Soubor byl zpracován eukleidovskou metrikou metodou nejbližšího souseda (Nearest), kde kritériem pro spojování shluků je minimum z možných mezishlukových vzdáleností objektů. Vstupní proměnné jsou vyjádřeny v různých měrných jednotkách, proto byla provedena standardizace. Jak vyplývá z Tabulky č.19 a grafů na obr. č.20-21 bylo do jednoho shluku zařazeno 60 vzorků a do druhého 1 vzorek.

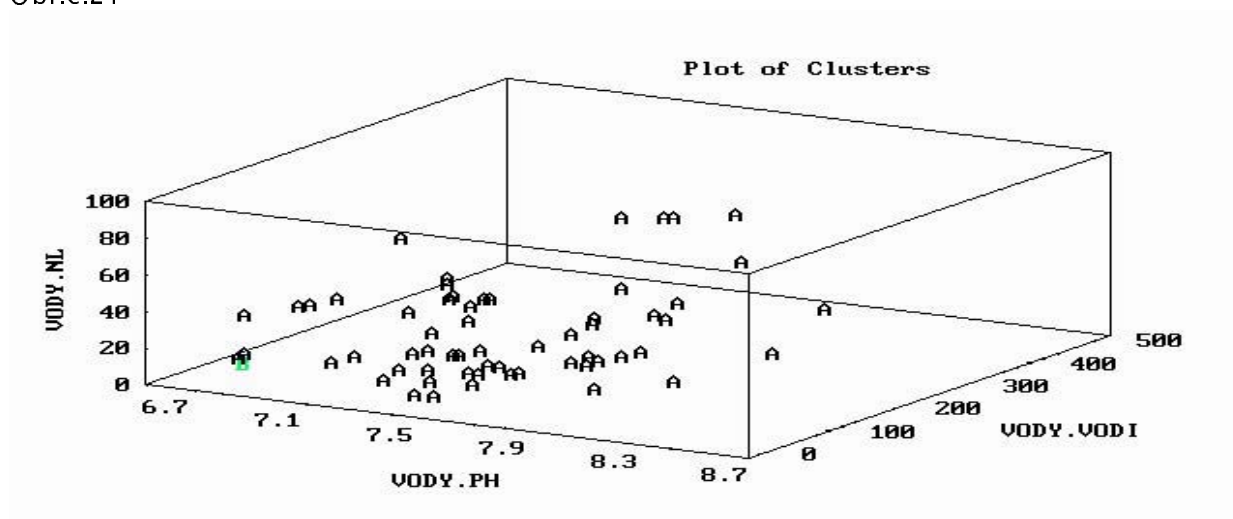
Tab č.19 Results of Clustering by Single Linkage or Nearest Neighbor Method

Cluster	Frequency	Percentage
1	60	98.3607
2	1	1.6393

Obr.č.20



Obr.č.21



Program Scan

Tab.č.20 Hierarchical Cluster Analysis

Complete Linkage
Euclidean Distance
Amalgamation Steps

Step	Number of Similarity	Distance	Clusters	New	Number of obs.
clusters	level	level	joined	cluster	in new cluster
1	60	99.73	7.134	6 32	6 2
2	59	99.67	8.903	29 49	29 2
3	58	99.55	12.177	29 61	29 3
4	57	99.46	14.647	11 39	11 2
5	56	99.45	14.886	3 48	3 2
6	55	99.44	15.121	27 53	27 2
7	54	99.40	16.233	33 54	33 2
8	53	99.39	16.480	12 29	12 4
9	52	99.31	18.659	17 35	17 2

Metody s latentními proměnnými a klasifikační metody

10	51	99.11	23.938	31	46	31	2
11	50	99.10	24.223	17	23	17	3
12	49	99.08	24.703	10	15	10	2
13	48	99.02	26.470	22	33	22	3
14	47	98.98	27.316	27	37	27	3
15	46	98.96	27.878	26	47	26	2
16	45	98.84	31.150	3	28	3	3
17	44	98.84	31.327	9	58	9	2
18	43	98.75	33.588	31	52	31	3
19	42	98.72	34.527	11	43	11	3
20	41	98.55	39.054	6	10	6	4
21	40	98.41	42.893	17	55	17	4
22	39	98.39	43.289	41	42	41	2
23	38	98.38	43.646	3	45	3	4
24	37	98.35	44.436	9	11	9	5
25	36	98.31	45.414	2	38	2	2
26	35	98.24	47.429	16	22	16	4
27	34	98.21	48.061	21	26	21	3
28	33	98.05	52.504	2	31	2	5
29	32	97.93	55.737	21	27	21	6
30	31	97.90	56.474	13	18	13	2
31	30	97.65	63.297	12	17	12	8
32	29	97.59	64.910	9	44	9	6
33	28	97.32	72.151	2	6	2	9
34	27	97.27	73.322	5	41	5	3
35	26	97.08	78.543	3	14	3	5
36	25	96.91	83.190	20	60	20	2
37	24	96.78	86.516	9	59	9	7
38	23	96.41	96.690	8	25	8	2
39	22	96.29	99.807	7	13	7	3
40	21	96.13	104.010	9	16	9	11
41	20	96.06	105.842	30	50	30	2
42	19	95.96	108.596	2	21	2	15
43	18	95.86	111.236	1	8	1	3
44	17	95.44	122.783	24	34	24	2
45	16	95.04	133.483	36	57	36	2
46	15	94.13	157.835	4	56	4	2
47	14	94.06	159.883	5	12	5	11
48	13	93.65	170.878	3	9	3	16
49	12	91.24	235.523	1	24	1	5
50	11	90.85	246.144	2	19	2	16
51	10	90.35	259.676	4	30	4	4
52	9	89.14	292.014	5	7	5	14
53	8	88.20	317.410	2	20	2	18
54	7	85.46	390.975	1	40	1	6

Metody s latentními proměnnými a klasifikační metody

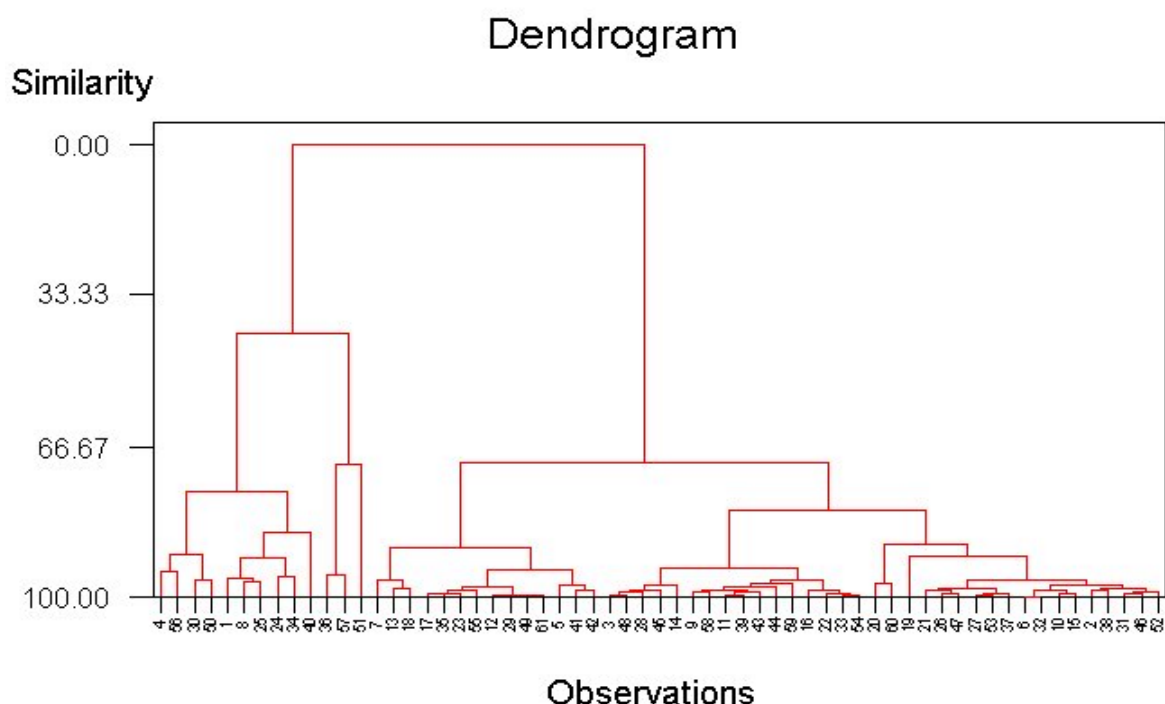
55	6	80.83	515.599	2	3	2	34
56	5	76.44	633.638	1	4	1	10
57	4	70.70	788.161	36	51	36	3
58	3	70.14	803.086	2	5	2	48
59	2	41.58	1571.190	1	36	1	13
60	1	0.00	2689.693	1	2	1	61

Final Partition

Number of clusters: 1

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	61	22738662.000	481.799	2101.684

Obr.č.22 Dendrogram



Dendrogram viz Obr. č.22 a tabulka č.20 znázorňují strukturu tvorby výsledného shluku pro proměnné Ph, Vodivost, Tvrdost, Ba a Mn. Nejvíce podobné jsou objekty, 6 a 32 (z 99,7 %). Při 70%-ní podobnosti tedy v podstatě existují 3 shluky.

- Výpočet korelační matice

Výsledkem korelační analýzy je korelační matice uvedená v tab.č. 21a,b,c pro každou dvojici proměnných je uveden korelační koeficient, tj. číslo popisující míru souvislosti mezi dvěma proměnnými. Korelační matice byla vypočtena pro 61 vzorků. V prvním řádku a prvním sloupci jsou uvedeny názvy proměnných. V prvním řádku každé proměnné jsou uvedeny korelační koeficienty. Na diagonále jsou jednotky, protože souvislost každé proměnné se sebou samou je samozřejmě absolutní. V druhém řádku každé proměnné je uvedena hladina významnosti, tj. pravděpodobnost, že při skutečné nulové korelaci by empiricky zjištěný korelační koeficient mohl dosáhnout nynější hodnoty. Je-li toto číslo malé (menší než 0,05), pak lze usoudit, že mezi příslušnými proměnnými existuje nějaká souvislost. Její síla je posuzována podle velikosti korelačního koeficientu. Je-li velikost korelačního koeficientu menší než 0,5 nelze mezi proměnnými hovořit o závislosti.

Metody s latentními proměnnými a klasifikační metody

Tab.č.21a Korelační matice

Sample Correlations							
	PH	VODI	NL	RL	TVRD	CL	
PH	1.0000	.2822	.1705	.2658	.3978	.2234	
	.0000	.0276	.1890	.0384	.0015	.0835	
VODI	.2822	1.0000	.2806	.9725	.9172	.9902	
	.0276	.0000	.0285	.0000	.0000	.0000	
NL	.1705	.2806	1.0000	.2428	.2572	.2781	
	.1890	.0285	.0000	.0593	.0454	.0300	
RL	.2658	.9725	.2428	1.0000	.8863	.9688	
	.0384	.0000	.0593	.0000	.0000	.0000	

Tab.č.21b Korelační matice

Sample Correlations							
	PH	VODI	NL	RL	TVRD	CL	
TVRD	.3978	.9172	.2572	.8863	1.0000	.8626	
	.0015	.0000	.0454	.0000	.0000	.0000	
CL	.2234	.9902	.2781	.9688	.8626	1.0000	
	.0835	.0000	.0300	.0000	.0000	.0000	
CHSKCR	.0684	.1927	.1747	.1544	.2618	.1549	
	.6002	.1368	.1782	.2347	.0415	.2334	
SIRANY	.5813	.4403	.0646	.3986	.6748	.3334	
	.0000	.0004	.6207	.0015	.0000	.0087	

Tab.č.21c Korelační matice

Sample Correlations							
	PH	VODI	NL	RL	TVRD	CL	
BA	.3145	.9038	.2811	.8599	.8421	.8819	
	.0136	.0000	.0282	.0000	.0000	.0000	
FE	-.0849	.0465	.0705	.0776	.1286	.0465	
	.5151	.7222	.5892	.5522	.3234	.7221	
MN	-.2325	-.0615	-.0322	-.0504	.0052	-.0824	
	.0714	.6380	.8055	.6997	.9681	.5279	
SR	.2631	.9837	.2852	.9612	.8687	.9824	

.0405 .0000 .0259 .0000 .0000 .0000

Z korelační matice lze vyčíst následující závislosti. Poměrně vysoká závislost je patrná mezi Vodivostí a Tvrdostí, Vodivostí a RL, Vodivostí a Cl, RL a Tvrdost, RL a Cl, Tvrdost a Cl, Ba a Vodivost, Ba a RL, Ba a Tvrdost, Ba a Cl, SR a Vodivost, SR a RL, SR a Tvrdost, SR a Cl. Ostatní závislosti jsou statisticky nevýznamné.

- Výpočet kovarianční matice

Výsledkem kovarianční analýzy je kovarianční matice uvedená v tab.č.22a,b,c. V hodnotě kovariance se odráží sice variabilita jednotlivých proměnných, ale nelze bezprostředně usuzovat na míru souvislosti. Používá se především jako vstup pro další výpočty.

Tab.č.22a Kovarianční matice

Covariances					

--					
	PH	VODI	NL	RL	
PH	0.13778	9.89048	1.28317	52.6456	
VODI	9.89048	8914.53	537.246	48993.7	
NL	1.28317	537.246	411.263	2627.63	
RL	52.6456	48993.7	2627.63	284689.	
TVRD	0.82007	480.935	28.9669	2626.31	

Tab.č.22b Kovarianční matice

Covariances					

--					
	PH	VODI	NL	RL	
CL	23.8600	26901.3	1622.68	148730.	
CHSKCR	0.16625	119.042	23.1777	539.147	
SIRANY	9.87845	1903.48	60.0097	9738.45	
BA	0.01053	7.69907	0.51433	41.3949	
FE	-0.04570	6.35710	2.07262	60.0195	

Tab.č.22c Kovarianční matice

Covariances					

--					
	PH	VODI	NL	RL	
MN	-0.01925	-1.29440	-0.14562	-5.99778	
SR	0.06513	61.9380	3.85722	342.017	

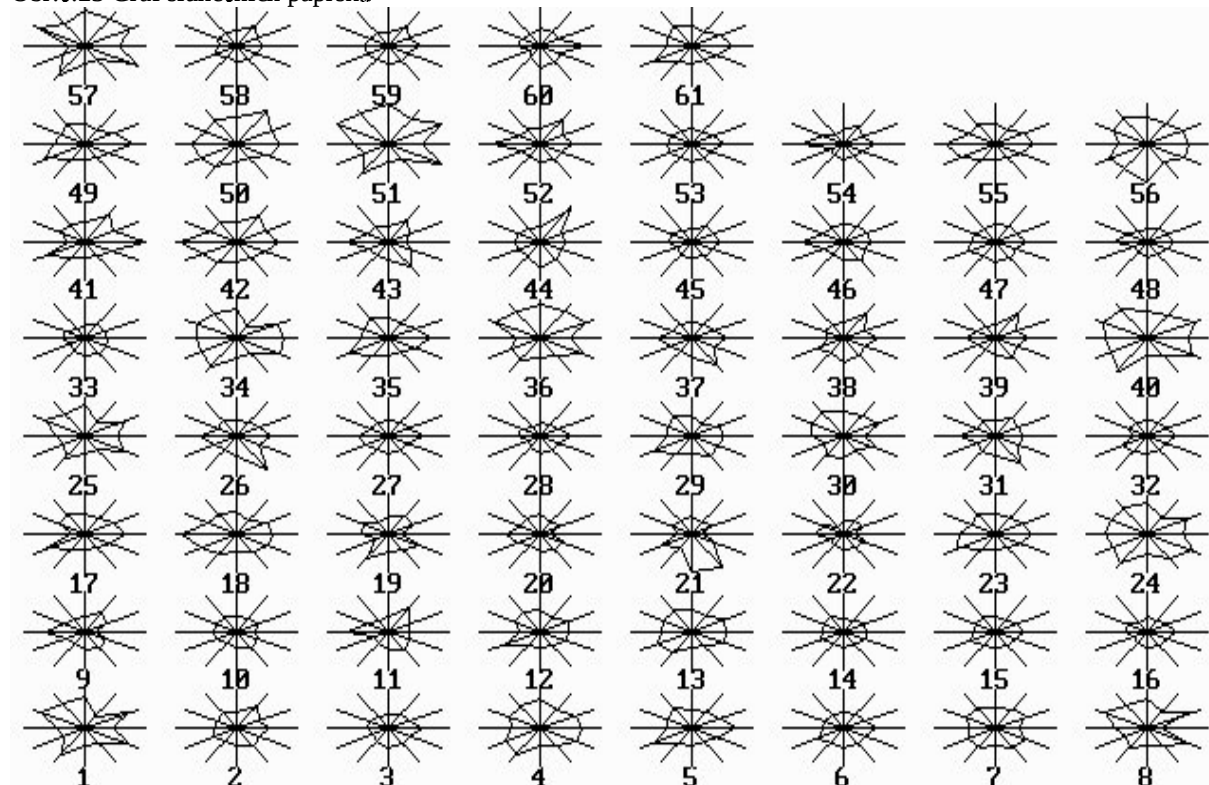
- Grafické zkoumání podobnosti objektů

Sun Ray Plot – graf slunečních paprsků

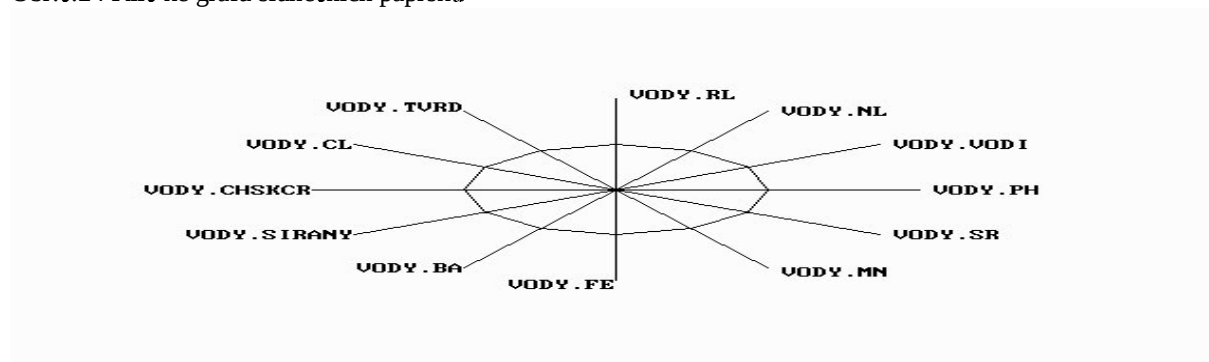
Graf slouží k vizuálnímu srovnání zadaných vícerozměrných objektů. Procedura Statgraphicsu Sun Ray Plot nakreslí pro každý vzorek hvězdici, která se skládá z paprsků spojených ve společném bodě a úseček mezi paprsky, které tvoří polygon. Počet paprsků odpovídá počtu proměnných. Střed polygonu představuje průměr

odpovídající proměnné a délka paprsku je 2n násobek směrodatné odchylky dané proměnné. Na obr.č. 23 je 61 polygonů testovaných vzorků, klíč je znázorněn na obr.č. 24. Sun Ray Plot – graf slunečních paprsků

Obr.č.23 Graf slunečních paprsků



Obr.č.24 Klíč ke grafu slunečních paprsků

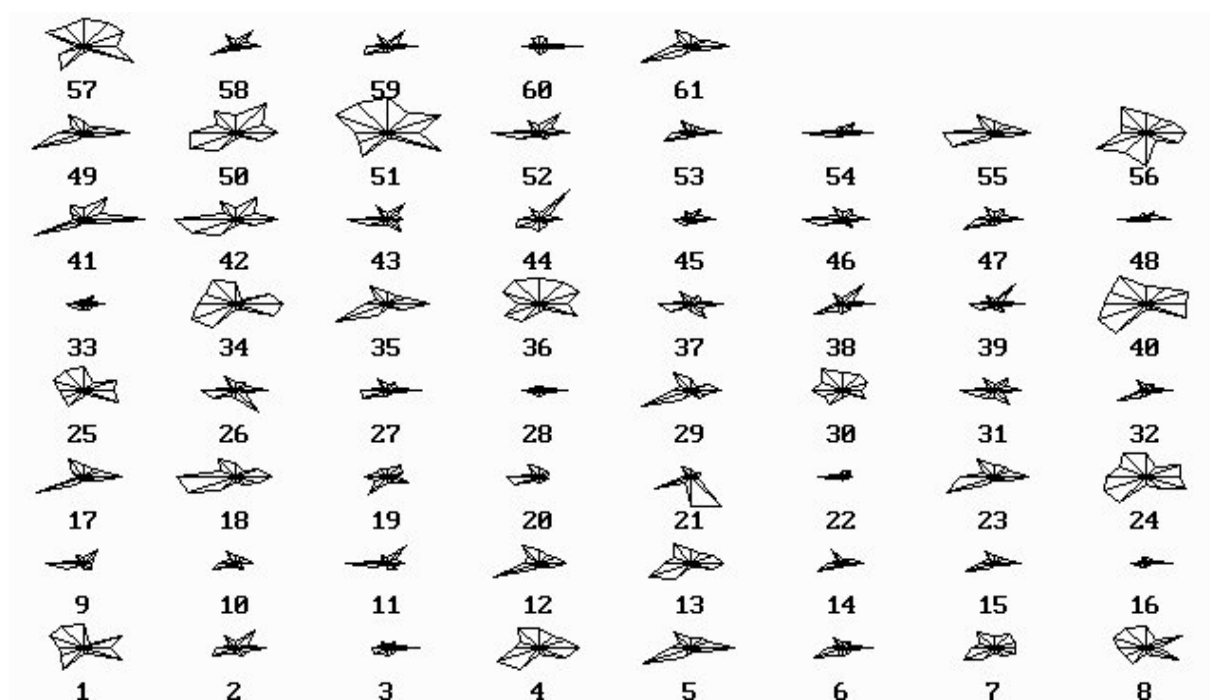


Dle vizuálního posouzení jednotlivých polygonů je zřejmé, že z hlediska jejich velikosti lze zaznamenat jejich podobnost. Hledáním podobnosti na základě tvaru lze usoudit na podobnost 36 (vzorek č. 2303) a 57(vzorek č. 3994), 24(vzorek č.1537) a 40(vzorek č.2319).

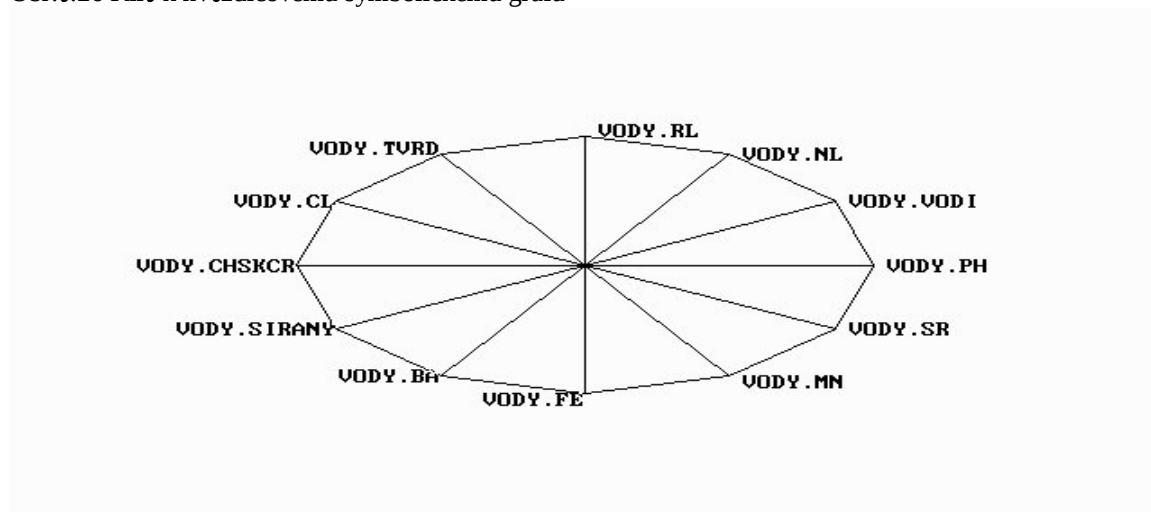
Star Symbol Plot – hvězdicový symbolický graf

Graf podobně jako předchozí slouží rovněž k vizuálnímu zobrazení podobnosti vícerozměrných objektů. Paprsky se spojují v jednom bodě. Nejkratší paprsek signalizuje, že příslušná proměnná má nejmenší hodnotu a nejdelší paprsek informuje o nejvyšší hodnotě příslušné proměnné. Na obr.č. 25 je 61 polygonů testovaných vzorků, klíč je znázorněn na obr.č.26. Podobně jako u předchozího grafu lze konstatovat, že polygony 36 (vzorek č. 2303) a 57(vzorek č. 3994), 24(vzorek č.1537) a 40(vzorek č.2319) jeví určitou podobnost. Zhodnocení souboru na základě tvaru polygonu je ještě více zřetelnější než u grafu slunečních paprsků.

Obr.č.25 Hvězdicový symbolický graf



Obr.č.26 Klíč k hvězdicovému symbolickému grafu



Závěr:

Metoda hlavních komponent

Metodou PCA byl snížen počet proměnných na 5 základních postačujících k vysvětlení zdrojové matice (Ph, Tvrlost, Ba, Vodivost a Mn).

Faktorová analýza

Faktorově čisté zvláště po rotaci vycházejí proměnné Vodivost, Ba a Tvrlost (první faktor). Proměnné Ph (2. faktor), Mn (3. faktor), vycházejí faktorově nejčistší, což znamená, že jsou nejméně závislé a podobné proměnným Vodivost, Ba a Tvrlost a sobě navzájem.

Shluková analýza

Dendrogram viz Obr. č.22 a tabulka č.20 znázorňují strukturu tvorby výsledného shluku pro proměnné Ph, Vodivost, Tvrlost, Ba a Mn. Nejvíce podobné jsou objekty, 6 a 32 (z 99,7 %). Při 70%-ní podobnosti tedy v podstatě existují 3 shluky.

Z korelační matice lze vyčíst následující závislosti. Poměrně vysoká závislost je patrná mezi Vodivostí a Tvrlostí, Vodivostí a RL, Vodivostí a Cl, RL a Tvrlost, RL a Cl, Tvrlost a Cl, Ba a Vodivost, Ba a RL, Ba a Tvrlost, Ba a Cl, SR a Vodivost, SR a RL, SR a Tvrlost, SR a Cl. Ostatní závislosti jsou statisticky nevýznamné.

Grafické zkoumání podobnosti objektů

Sun Ray Plot – graf slunečních paprsků i Star Symbol Plot – hvězdicový symbolický graf

Metody s latentními proměnnými a klasifikační metody

Dle vizuálního posouzení jednotlivých polygonů je zřejmé, že z hlediska jejich velikosti lze zaznamenat jejich podobnost. Hledáním podobnosti na základě tvaru lze usoudit na podobnost 36 (vzorek č. 2303) a 57(vzorek č. 3994), 24(vzorek č.1537) a 40(vzorek č.2319).

Jednotlivé odběry lze charakterizovat 5 proměnnými Ph, Tvrdost, Ba, Vodivost a Mn.