

Zpracování a vyhodnocování analytických dat

naměřená data



Zpracování a
statistická
analýza dat

analytické výsledky

Naměřená data

- **jedna hodnota** 5,00 mg (bod - 1D)
 - navážka, odměřený objem
- **řada dat** 15,8; 27,2; 38,3 ... (vektor)
 - Uspořádaná N-tice hodnot, poloha bodu v N-dimenzionálním prostoru - polohový vektor bodu
 - sada navážek, sada hodnot koncentrace roztoků standardu
 - sada hodnot absorbancí pro sadu vzorků

Naměřená data

- **funkční závislosti** – „matice“, „tabulky“, „grafy“
 - **dvojice hodnot** - x, y
 - spektrum, jednoduchý chromatogram, voltamogram,
 - závislost na čase (kinetika)
 - **trojice hodnot** - x, y, z
 - sada spekter v závislosti na čase
 - 2D spektra
 - **více dimenzionální** –
 - MS detekce v chromatografii
 - spektrální mapy povrchů
 - uspořádané sady spekter

Naměřená data

- **odrážejí jak vlastnosti analytu (analytů), tak řadu dalších vlivů**
 - **Systematické vlivy**
 - **Nahodilé (pseudonahodilé) vlivy**
- **Chyby měření**
 - **Hrubé** – vyloučení příslušných výsledků měření
 - **Systematické** – vliv přístroje, obsluhy, zvolené metody
 - **Náhodné** – fluktuace ve vzorku, fluktuace vnějších podmínek a další náhodné (často neidentifikované) vlivy

Chyby měření

- **Hrubé**
 - **Lidský faktor**
 - **Chybné nastavení či porucha přístroje**
 - **TŘEBA VYLOUČIT PŘÍSLUŠNÁ MĚŘENÍ**
- **Systematické**
 - **Dané zvolenou metodou analýzy, odečtu dat, volbou referentního signálu**
 - **Dané nastavením, konstrukcí přístroje**
 - Složka chyby, která se u řady výsledků zkoušky na tutéž charakteristiku zůstává konstantní nebo se mění předvídatelným způsobem.
 - **IDENTIFIKACE – analýza (certifikovaných) referenčních materiálů**

Chyby měření

- **Náhodné**

- **Způsobují rozptyl opakovaně naměřených dat**

- Složka chyby, která se u řady výsledků zkoušky na tutéž charakteristiku mění nepředvídatelným způsobem.

- **Je možný pouze odhad hodnoty náhodné chyby**

- **Jejich efekty na datový soubor lze mnohdy popsat tzv. normálním (Gaussovým) rozdělením s charakteristickou střední hodnotou a směrodatnou odchylkou (odhad střední hodnoty – aritmetický průměr a odhad směrodatné odchylky)**

Hodnoty

- **Skutečná hodnota – true value**
 - **Hodnota charakterizující dokonale definovanou veličinu za podmínek, při nichž je veličina uvažována.**
 - **Je to hodnota, která by mohla být získána dokonalým měřením. Skutečné hodnoty jsou v principu nestanovitelné.**
- **Konvenční skutečná hodnota – conventional true value**
 - **Hodnota veličiny, kterou můžeme pro daný účel nahradit skutečnou hodnotu veličiny.**

Hodnoty

- **Konvenční skutečná hodnota – conventional true value**
 - Je to hodnota přiřazená dané veličině a přijatá někdy na základě konvence s tím, že její nejistota je vhodná pro dané použití.
 - Pro daný případ může být hodnota přisouzená dané veličině a realizovaná referenčním etalonem nebo CRM považována za konvenční skutečnou hodnotu.
 - Jako příklad uveďme tabelovanou hodnotu Faradayovy konstanty: $F = 96\,485,3415(39) \text{ C mol}^{-1}$
 - Pro stanovení konvenční skutečné hodnoty se často používá řady výsledků měření dané veličin.

Hodnoty

- **Dohodnutá referenční hodnota – Accepted reference value**
 - **Hodnota, která se používá jako odsouhlasená referenční hodnota pro srovnání a která se získá jako**
 - a) teoretická nebo zavedená (ustanovená) hodnota, založená na vědeckých principech (viz skutečná hodnota)
 - b) přidělená (určená) nebo certifikovaná hodnota založená na experimentální práci národní nebo mezinárodní organizace
 - c) dohodnutá nebo certifikovaná hodnota založená na experimentální spolupráci pod dohledem (vedením) vědecké nebo inženýrské skupiny (viz konvenční skutečná hodnota)
 - d) střední hodnota specifikovaného souboru výsledků měření, nejsou-li hodnoty podle a), b) a c) dosažitelné

Hodnoty

- **Zjištěná hodnota – Observed value**
 - **Hodnota určité vlastnosti získaná jako výsledek jediného pozorování**
 - **Výsledek zkoušky (testu) – test result**
 - **Hodnota** určité vlastnosti získaná použitím určité **zkušební metody**. Zkušební metoda má specifikovat, že má být provedeno jedno nebo určitý počet pozorování, a jako výsledek zkoušky (testu) se má uvést jejich průměr nebo jiná vhodná funkce (např. jejich medián). Může rovněž vyžadovat použití předepsaných korekcí na normální podmínky, např. korekci objemů plynů na normální teplotu a tlak. Výsledkem zkoušky tedy může být hodnota vypočtená z několika pozorovaných údajů.
V jednoduchém případě je výsledkem zkoušky sama pozorovaná hodnota.

Hodnoty

- **Zjištěná hodnota – Observed value**
 - **Hodnota určité vlastnosti získaná jako výsledek jediného pozorování**

– Výsledek měření

- result of a measurement

- Hodnota přisuzovaná měřené veličině získaná měřením.
- Je-li použit pojem „**výsledek měření**“, mělo by být zřejmé, zda se jedná o:
 - indikaci
 - nekorigovaný výsledek
 - korigovaný výsledeka zda se jedná o průměr několika hodnot.

Úplný údaj výsledku měření obsahuje informaci o nejistotě měření.

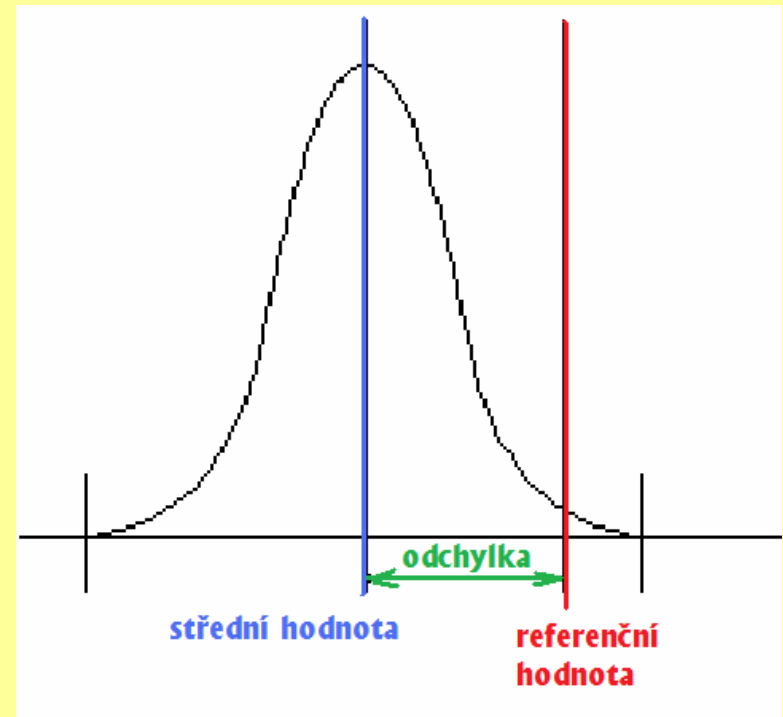
Hodnoty

- **Odchylka – Bias**

- **Rozdíl mezi střední hodnotou výsledku zkoušky a přijatou referenční hodnotou.**

- The difference between a population mean of measurements or test results and an accepted reference value.

- Souvisí s celkovou systematickou chybou.



Výsledky

- **Pravdivost – Trueness**

- **Těsnost souhlasu mezi průměrnou hodnotou získanou z velkého počtu výsledků měření a dohodnutou referenční hodnotou (skutečnou hodnotou, konvenční skutečnou hodnotou).**
- **Pravdivý výsledek je zatížen nulovou systematickou chybou. Mírou pravdivosti je odchylka.**
- **Průměrnou hodnotou je v definici myšlena střední hodnota základního (statistického) souboru.**

Výsledky

- **Přesnost – Precision**

- **Těsnost souhlasu mezi nezávislými výsledky zkoušky získanými za předem specifikovaných podmínek.**
- Přesnost vyjadřuje míru rozptýlení pozorování okolo střední hodnoty. Nemá vztah ke skutečné hodnotě. Souvisí s opakovatelností.

- **Správnost – Accuracy**

- **Těsnost souhlasu mezi výsledkem měření a skutečnou hodnotou (konvenční skutečnou hodnotou) měřené veličiny.**
- Správnost kombinuje přesnost a pravdivost, tj. vlivy náhodných a systematických faktorů.

Výsledky

- **Opakovatelnost – Repeatability**
 - **Opakovatelnost vyjadřuje těsnost souhlasu mezi výsledky nezávislých měření stejného analytu provedených stejnou metodou, stejným experimentátorem, na stejném přístroji, na stejném místě, za stejných podmínek v krátkém časovém intervalu.**
 - **Opakovatelnost je vlastností metody, ne výsledku.**
 - **Opakovatelnost souvisí s rozptylem hodnot opakovaných měření.**

Výsledky

- **Reprodukovatelnost – Reproducibility**
 - **Reprodukovatelnost vyjadřuje těsnost souhlasu mezi výsledky měření stejného analytu ve vzorcích stejného materiálu, kdy jsou jednotlivá měření prováděna za různých podmínek (experimentátor, přístroj, místo, podmínky, čas, avšak stejná metoda).**
 - **Mezilaboratorní testy – porovnání shody výsledků**
 - **Různí pracovníci – stejná metoda**
 - **Větší rozptyl hodnot než při opakovatelnosti.**

Výsledky

- **Nejistota měření – Uncertainty of measurement**
 - **Parametr přidružený k výsledku měření, který charakterizuje rozptyl hodnot, které by mohly být důvodně přisuzovány k měřené veličině.**
 - Nejistota vymezuje hranice, v nichž je výsledek považován za správný, tj. **přesný a pravdivý**. Nejistota obecně zahrnuje mnoho složek. Některé z nich mohou být získány ze statistického rozdělení výsledků série měření, charakterizovaných výběrovou směrodatnou odchylkou. Ostatní složky, které mohou být rovněž charakterizovány výběrovými směrodatnými odchylkami, se vyhodnocují z předpokládaných rozdělení na základě zkušeností nebo na základě jiných informací.

Výsledky

- **Nejistota měření – Uncertainty of measurement**

- **Celková nejistota** se vyčíslí po odhalení všech faktorů, které k ní přispívají. Jejich příspěvky se vyjadřují jako směrodatné odchylky vybraných pozorování (pro **náhodné** složky) nebo z jiných zdrojů informací (pro **systematické** složky).
- Kombinovaná standardní nejistota se vypočte jako kombinace rozptylů (propagace rozptylů) všech složek, vyjádřená jako směrodatná odchylka.
- Kombinovaná standardní nejistota násobená faktorem 2 dává (přibližně) 95% hladinu spolehlivosti.

Výsledky

- **Nejistota měření – Uncertainty of measurement**
 - Souhrnná charakterizace kvality výsledku jediným parametrem.
 - Kvantitativní odhad její velikosti podle jednotné metodiky platné pro všechny obory měření.
 - 1993 – Guide to the Expression of Uncertainty in Measurement - International Organization for Standardization
 - **Dva postupy pro odhad nejistoty**
 - **Postup zdola nahoru – složka po složce**
 - analýza procesu měření krok po kroku
 - **Postup shora dolů** – z validační studie, z dlouhodobého sledování v laboratoři, z mezilaboratorních testů, z analýzy CRM

Pravděpodobnost

- **Pravděpodobnost (náhodného jevu)**
 - číselný údaj, který je mírou očekávatelnosti výskytu jevu
 - interval $0 - 1$, $0 - 100\%$
 - náhodný jev – opakovaná činnost prováděná za „stejných“ podmínek, jejíž výsledek je nejistý a závisí na náhodě
- **Rozdělení pravděpodobnosti**
 - funkce, přiřazující pravděpodobnosti událostem, tvrzením, jevům...
 - pravidlo, přiřazující každému jevu určitou pravděpodobnost
 - pro různé sady událostí (jevů) – různé rozdělení
 - zobrazení, které každému elementárnímu jevu přiřadí číslo, charakterizující jeho pravděpodobnost

 - každé hodnotě diskrétní veličiny přiřadíme pravděpodobnost
 - intervalu hodnot spojité veličiny přiřadíme pravděpodobnost

Pravděpodobnost

- **Rozdělení pravděpodobnosti**

- **diskrétní veličiny**

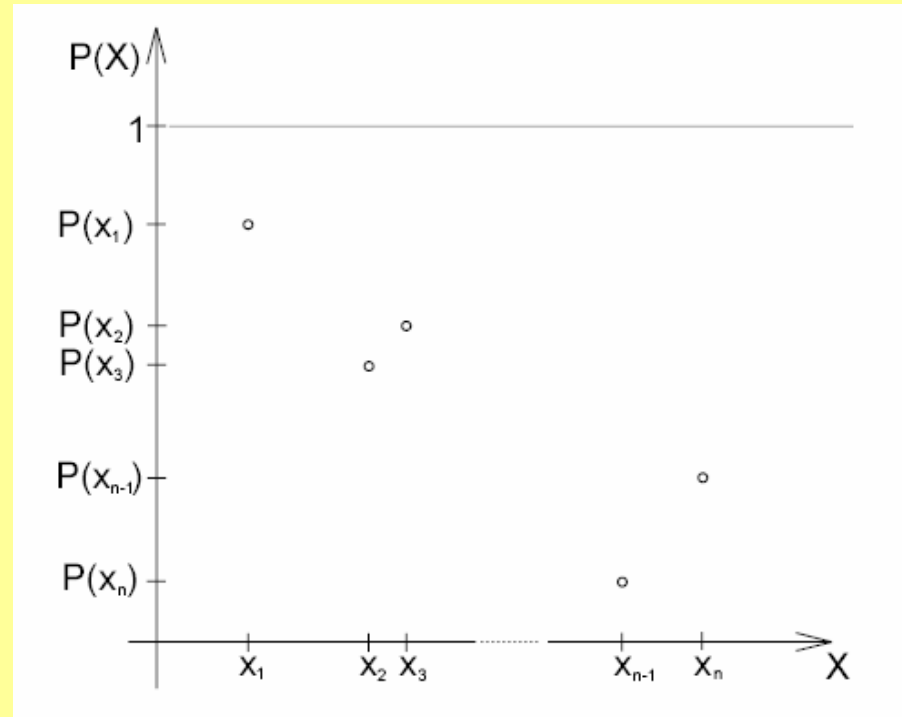
- **pravděpodobnostní funkce $P(x)$**

- s hodnotami pro x_i , kde $i = 1, .. n$**

$$\sum_{i=1}^n P(x_i) = 1$$

pravděpodobnost hodnoty v intervalu

$$P(x_k \leq x \leq x_m) = \sum_{x=x_k}^{x_m} P(x)$$



Pravděpodobnost

- **Rozdělení pravděpodobnosti**
 - **diskrétní veličiny**
 - **distribuční funkce**

$$F(x) = P(t < x) = \sum_{t < x} P(t)$$

pravděpodobnost hodnoty v intervalu

$$P(x_k \leq x < x_m) = F(x_m) - F(x_k)$$

Pravděpodobnost

- **Rozdělení pravděpodobnosti**

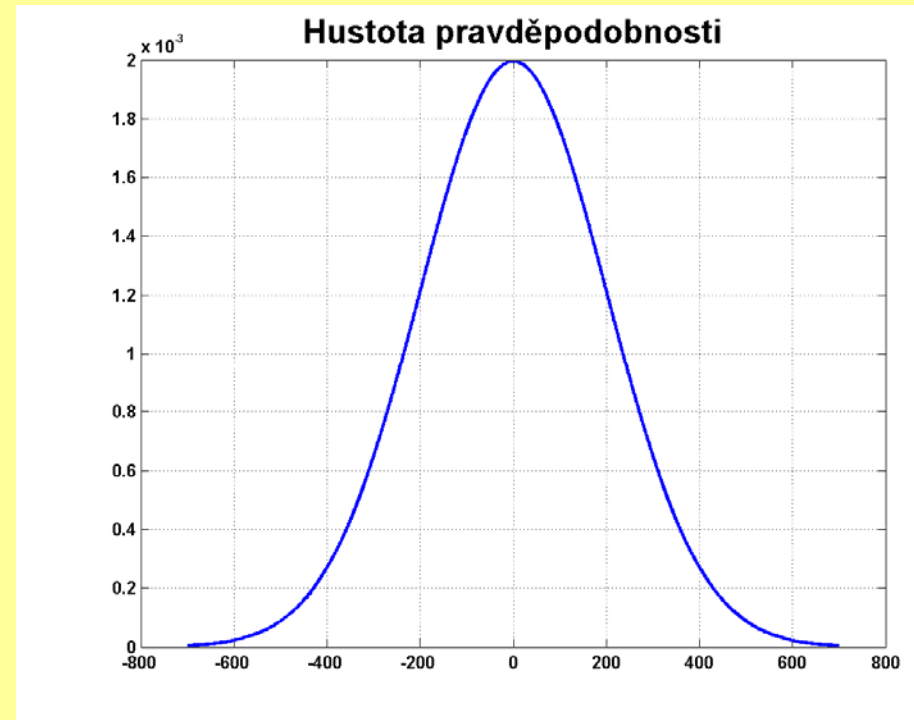
- **spojité veličiny**

- **funkce - hustota (rozdělení) pravděpodobnosti**

$$\int_{-\infty}^{\infty} \rho(x) = 1$$

pravděpodobnost hodnoty v intervalu

$$P(x_k \leq x \leq x_m) = \int_{x_k}^{x_m} \rho(x)$$



Pravděpodobnost

- **Rozdělení pravděpodobnosti**
 - **spojité veličiny**
 - **distribuční funkce**

$$F(x) = P(t < x) = \int_{-\infty}^x \rho(t)$$

pravděpodobnost hodnoty v intervalu

$$P(x_k \leq x \leq x_m) = F(x_m) - F(x_k)$$

Pravděpodobnost

• Rozdělení pravděpodobnosti

- každé hodnotě diskrétní veličiny přiřadíme pravděpodobnost
- příklady
 - binomické rozdělení
 - rovnoměrné rozdělení – diskrétní – (hody kostkou)
 - Poissonovo rozdělení
- intervalu hodnot spojité veličiny přiřadíme pravděpodobnost
- příklady
 - rovnoměrné rozdělení
 - normální rozdělení
 - Studentovo rozdělení
 - exponenciální rozdělení

Pravděpodobnost

- **Rozdělení pravděpodobnosti**

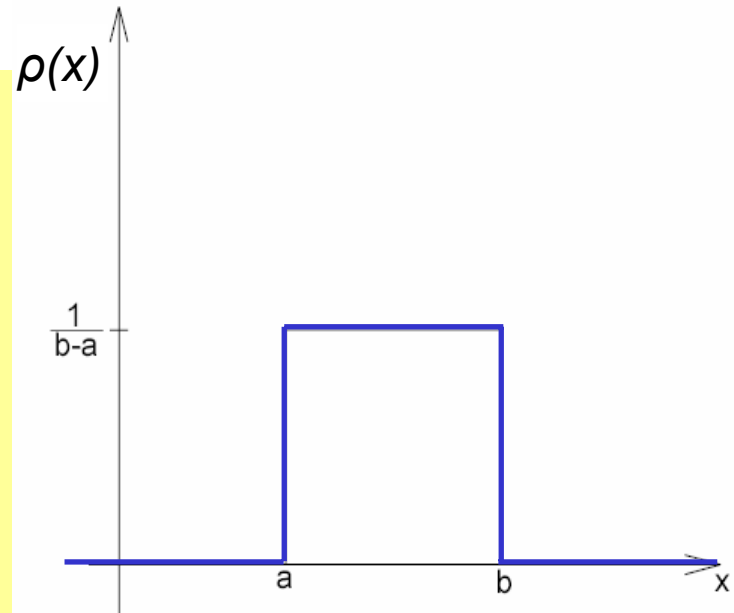
– rovnoměrné rozdělení

$$\rho(x) = 0, \text{ pro } x \notin (a, b)$$

$$\rho(x) = \frac{1}{b-a}, \text{ pro } x \in (a, b)$$

střední hodnota

$$\mu(x) = \frac{a+b}{2}$$



Pravděpodobnost

• Rozdělení pravděpodobnosti

– Normální rozdělení

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

střední hodnota

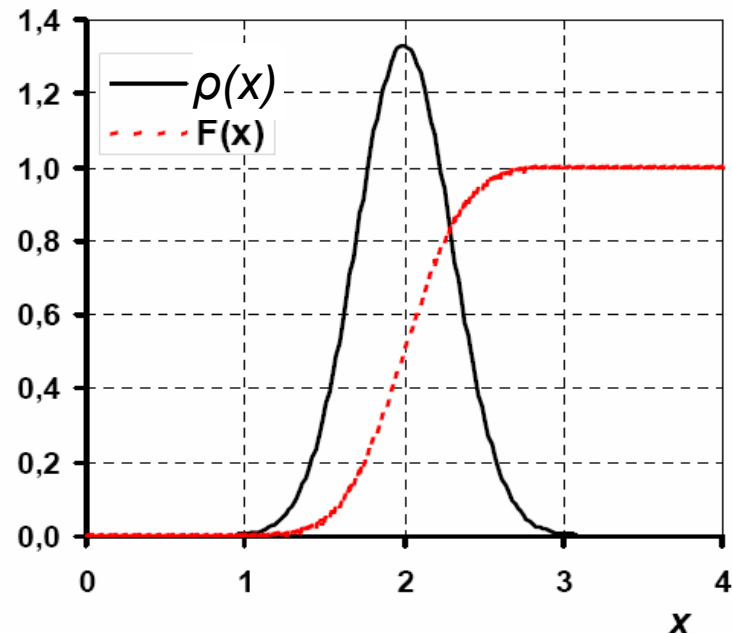
μ

rozptyl

σ^2

směrodatná odchylka

σ



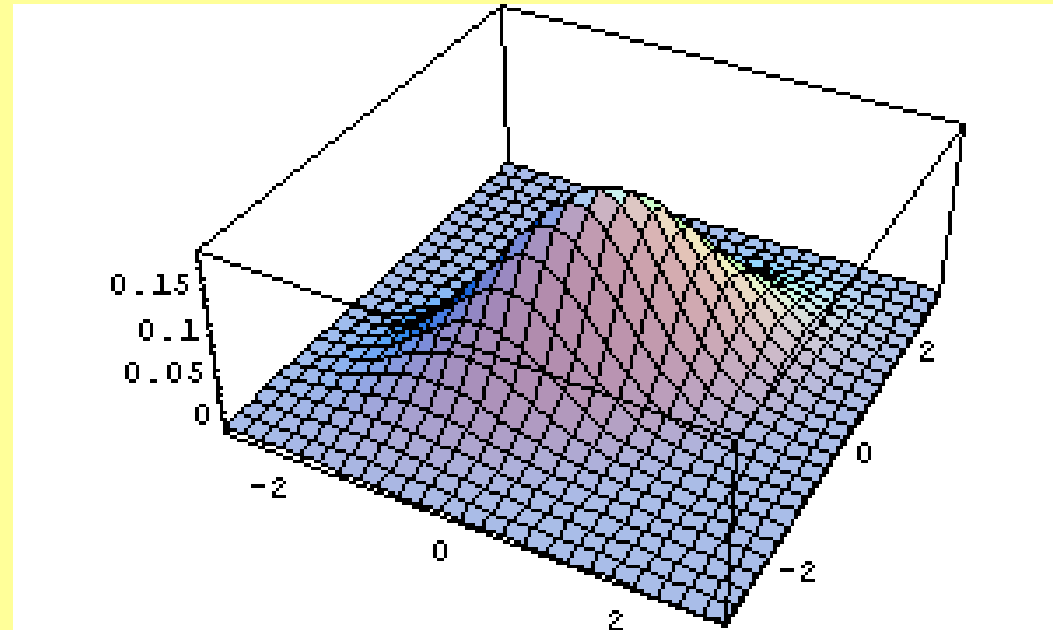
Obr. 1: Průběh hustoty pravděpodobnosti $\rho(x)$ distribuční funkce $F(x)$ náhodné veličiny $\xi(x)$ s normálním rozdělením $N(2,0; 0,09)$

Pravděpodobnost

- **Rozdělení pravděpodobnosti**
 - Vícerozměrné normální rozdělení

$$\rho(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu)\right),$$

$$\mu \in R^d, \Sigma \in R^{d \times d}$$



Pravděpodobnost

- **Centrální limitní věta**

Mějme n nezávislých náhodných veličin X_i .

Jejich součet $S = X_1 + \dots + X_n$ je také náhodná veličina se střední hodnotou $\mu = \mu_1 + \dots + \mu_n$ a rozptylem $\sigma^2 = \sigma_1^2 + \dots + \sigma_n^2$.

Centrální limitní věta:

S rostoucím n se distribuce $F(S)$ blíží normálnímu rozdělení $N(\mu, \sigma^2)$.

Pravděpodobnost

- **Rozdělení pravděpodobnosti**

- **Normální rozdělení** - jedno z nejdůležitějších rozdělení pravděpodobnosti spojité náhodné veličiny
- Tímto rozdělením pravděpodobnosti se sice neřídí velké množství veličin, ale jeho význam spočívá v tom, že za určitých podmínek dobře aproximuje řadu jiných pravděpodobnostních rozdělení (spojitých i diskrétních).
- V souvislosti s normálním rozdělením jsou často zmiňovány **náhodné chyby**, např. ***chyby měření, způsobené velkým počtem neznámých a vzájemně nezávislých příčin.***

Pravděpodobnost

- **Kvantily**

- Kvantily jsou body, ve kterých distribuční funkce náhodné proměnné dosahuje dané hodnoty.
- Přesněji řečeno je kvantil Q_p (nejmenší) takové číslo, pro které platí, že

$$P(X < Q_p) \leq p,$$

tzn.

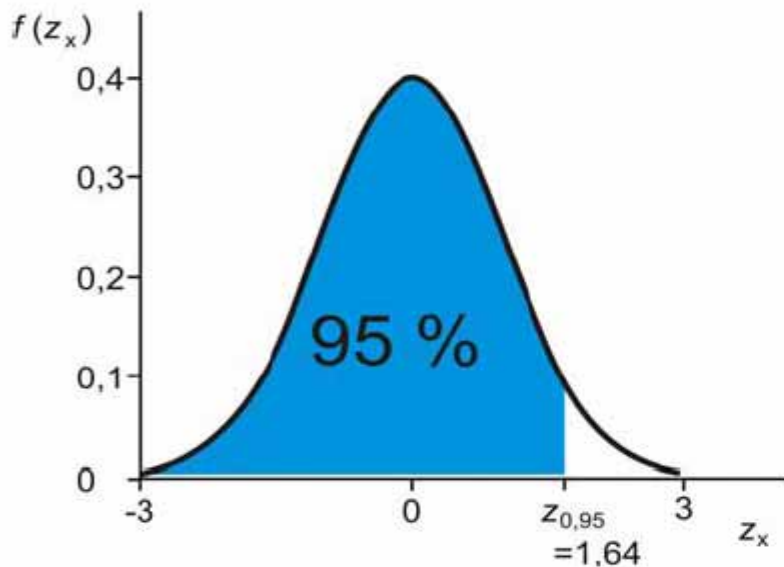
$$F(Q_p) \leq p,$$

kde $F(x)$ je distribuční funkce daného rozdělení.

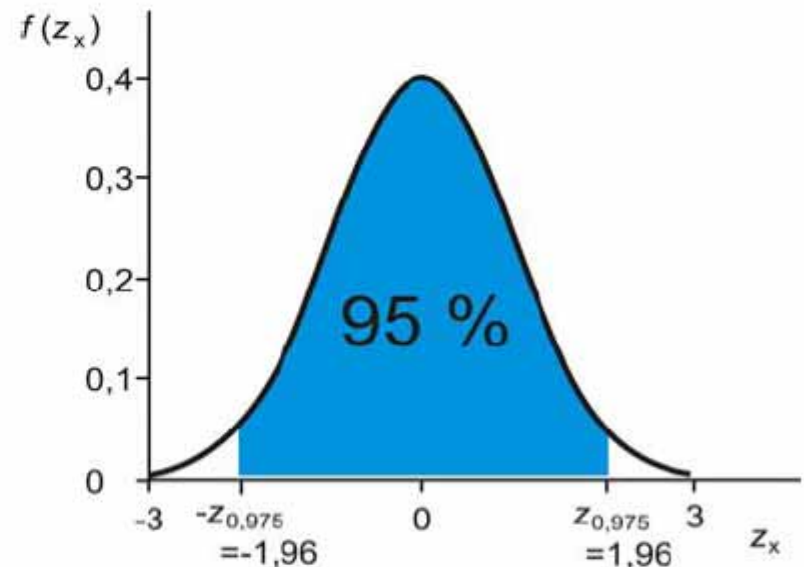
Pravděpodobnost

- **Kvantily**

– 95%procentní kvantil ($P=0,95$) je tedy hodnota, kterou 95 % výsledků analýzu nepřekročí



A



B

Obr. 3: Křivky hustoty pravděpodobnosti rozdělení $N(0,1)$ s vyznačenými 95% kvantily pro jednostrannou (A) a oboustrannou (B) pravděpodobnost.

Pravděpodobnost

- **Kvantily**

- 95%procentní kvantil ($P=0,95$) je tedy hodnota, kterou 95 % výsledků analýzu nepřekročí
- Tomu odpovídá 95 % **hladina spolehlivosti**
- Tomu odpovídá 5 % **hladina významnosti**

- **Běžně voleny 95 %, 99 %, příp. 90 % hladina spolehlivosti**

Pravděpodobnost

• Rozdělení pravděpodobnosti

– Normální rozdělení

- Charakteristiky - střední hodnota, rozptyl

- Nelze zjistit, jen odhadnout.

- ODHADY –

- Aritmetický průměr – odhad střední hodnoty

- Výběrová směrodatná odchylka – odhad směrodatné odchylky σ je s_{n-1}

- POZOR – aritmetický průměr není absolutní veličina – má své rozdělení $N(\bar{x}, s_p^2)$

- Směrodatná odchylka průměru s_p

$$s_p = \frac{s_{n-1}}{\sqrt{n}}$$

Statistické testy

- **Stanovení apriorního předpokladu**
 - nulové hypotézy – její testování na určité hladině významnosti
 - na příslušné hladině významnosti se hypotéza buď vyvrátí nebo potvrdí
 - Může být vyřčena alternativní hypotéza
 - **Chyby 1. druhu** – zamítnutí ve skutečnosti platné hypotézy (malá, odpovídá hladině významnosti)
 - **Chyby 2. druhu** – potvrzení neplatné hypotézy, velikost chyby obvykle neznáme – podrobný rozbor nulové a alternativní hypotézy
- **Statistické testy opakovaných měření**

Bilancování nejistot

Desatero bilancování nejistoty podle GUM

1. definuj měřenou veličinu
2. popiš měřící postup modelem měření
3. identifikuj všechny zdroje nejistoty
4. odhadni hodnoty všech vstupních veličin (modelu měření)
5. odhadni standardní nejistoty každé ze vstupních veličin
6. vypočti hodnotu měřené veličiny (pomocí rovnice měření)
7. vypočti kombinovanou nejistotu výsledku
8. vypočti rozšířenou nejistotu (s vybraným koeficientem rozšíření)
9. analyzuj jednotlivé příspěvky k celkové nejistotě a přemýšlej o nich
10. dokumentuj všechny kroky a písemně je zaznamenej a uchovej

Šum - „statistický“

ŠUM

- signál bez žádaného informačního obsahu
- rušivý signál
- signál náhodného (*pseudonáhodného*) charakteru

Šum - „statistický“

ŠUM

- **náhodná proměnná**
s nulovou střední hodnotou

POTLAČENÍ

Nejdokonalejší

- **mnohonásobné opakování experimentu**
za DOKONALE STEJNÝCH podmínek,
resp. za podmínek, které se liší POUZE
„NÁHODNÝM“ („statistickým“) šumem

Šum - potlačení

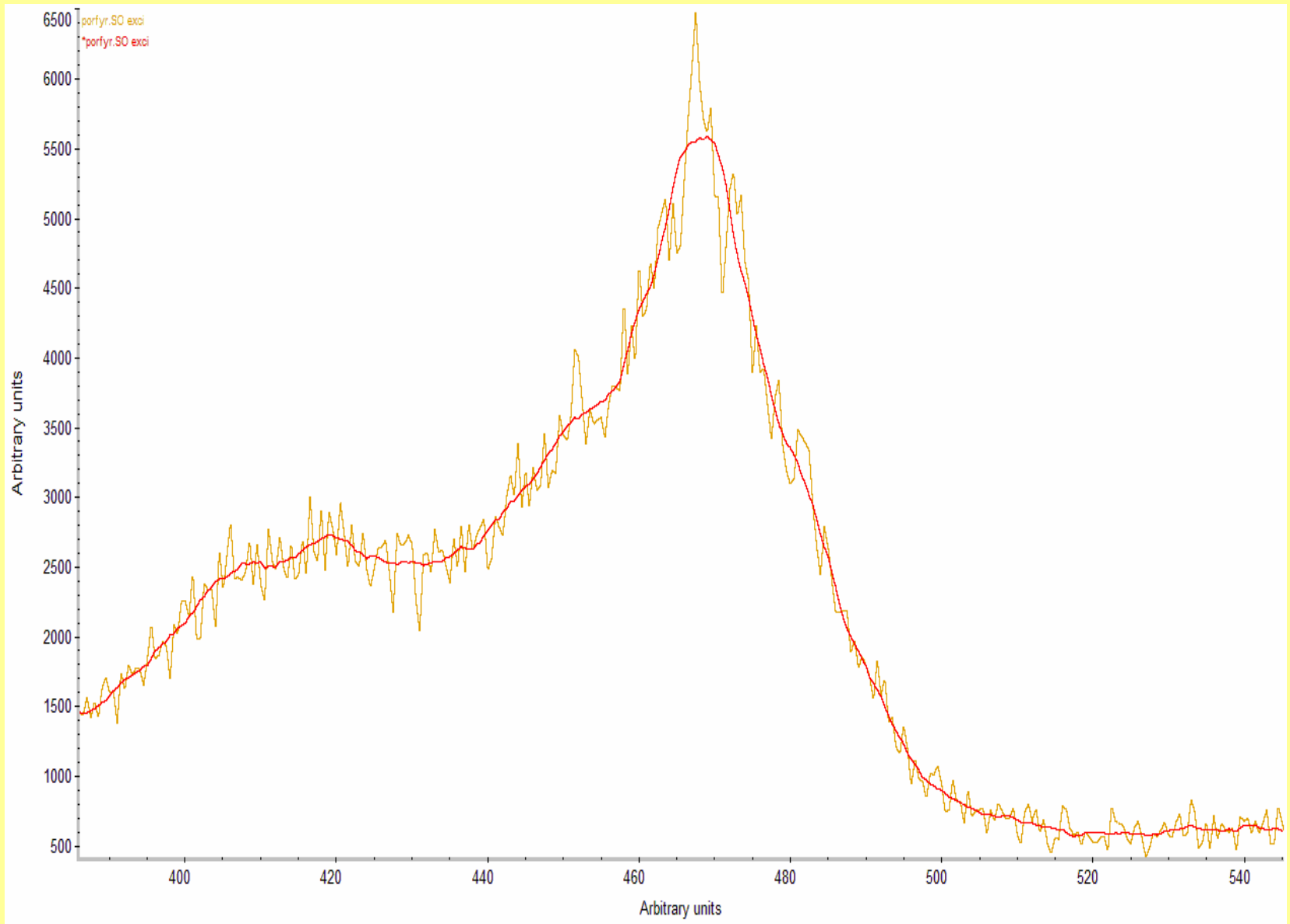
GRAFICKÉ VYHLAZENÍ ŠUMU

- proložení šumových oscilací hladkou křivkou

NUMERICKÉ VYHLAZENÍ ŠUMU

- proložení šumových oscilací
 - polynomem
 - segmentovou funkcí (spline function)

Šum - potlačení



Šum - potlačení

NUMERICKÉ VYHLAZENÍ ŠUMU

- proložení polynomem dostatečně vysokého řádu m

Data dvojice hodnot (x_i, y_i) - výpočet $y_{i,opr}$

1. volba lichého počtu bodů $(2n+1)$ $\{2n > m\}$ v intervalu, kde se bude vyhlazovat
2. volba hodnot postupně klesajících statistických vah w_0 až w_n , kterými se řídí vliv sousedních bodů na vyhlazovanou hodnotu
3. pro bod (x_i, y_i) se zahájí vyhlazování dle dalších bodů

NUMERICKÉ VYHLAZENÍ ŠUMU

Data dvojice hodnot (x_i, y_i) - výpočet $y_{i,opr}$

4. vyřeší se vzhledem k neznámým koeficientům
interpoláčního polynomu $p_k, k \in \langle 0, m \rangle$ soustava rovnic

$$\sum_{k=0}^m p_k (x_j - x_i)^k = y_j \quad j \in \langle i-n; i+n \rangle$$

a to se statistickými váhami w_{i-j} metodou vážených
nejmenších čtverců

5. jako opravená (vyhlazená) hodnota se dosadí

$$y_{i,opr} = p_0$$

6. i se změní o jednotku a postup se opakuje od 4

NUMERICKÉ VYHLAZENÍ ŠUMU

Prokládání polynomem

- metodou vážených nejmenších čtverců

- největší statistická váha studovanému bodu
- váha se snižuje se vzdáleností od tohoto bodu
- interpolovaná hodnota polynomické funkce pro pořadnici studovaného bodu je pak odhadem skutečné funkční hodnoty s eliminovaným šumem
- DOCHÁZÍ KE KONVOLUCI DAT - *nutná vysoká hustota bodů*

NUMERICKÉ VYHLAZENÍ ŠUMU

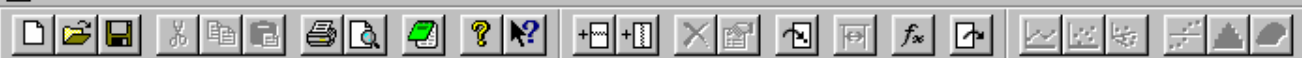
Prokládání polynomem

Savitzky A., Golay M.J.E.: *Smoothing and Differentiation of Data by Simplified Least Squares Procedures*, **Anal.Chem.** 36, 1627 (1964).

Steiner J., Termonia Y., Deltour J.: *Comments on Smoothing and Differentiation of Data by Simplified Least Squares Procedures*, **Anal.Chem.** 44, 1906 (1972).

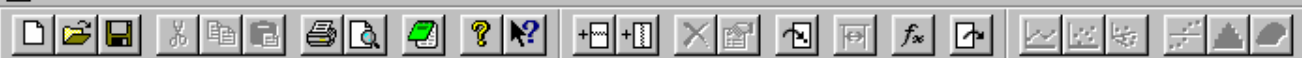
Multivariační analýza dat

- ✓ použití mnoha proměnných zároveň
- ✓ základem tabulka - matice dat
 - ↳ řádky - vzorky
 - ↳ sloupce - proměnné
- proměnné - závislé - např. koncentrace
- proměnné - nezávislé - spektrální data
- proměnné - klasifikační („category“)
 - třídění vzorků do skupin



		Methanol	Ethanol	Propanol	1100	1105	1110	1115	1120	1125	1130	1135	1140
		1	2	3	4	5	6	7	8	9	10	11	12
A1	1	0.0000	0.0000	100.0000	-0.1524	-0.1537	-0.1544	-0.1520	-0.1484	-0.1448	-0.1359	-0.1231	-9.6799e-01
A2	2	100.0000	0.0000	0.0000	-0.1266	-0.1281	-0.1287	-0.1264	-0.1208	-0.1102	-9.1172e-02	-6.2867e-02	-2.5467e-01
A3	3	0.0000	100.0000	0.0000	-0.1192	-0.1035	-0.1170	-0.1140	-0.1100	-0.1051	-9.7731e-02	-8.2392e-02	-3.8356e-01
A4	4	33.4000	33.0000	33.6000	-0.1470	-0.1485	-0.1484	-0.1459	-0.1403	-0.1347	-0.1218	-0.1016	-6.5365e-01
A5	5	49.8000	24.9000	25.3000	-0.1400	-0.1411	-0.1421	-0.1392	-0.1345	-0.1272	-0.1128	-9.1414e-02	-5.4544e-01
A6	6	24.9000	25.3000	49.8000	-0.1519	-0.1524	-0.1525	-0.1499	-0.1460	-0.1396	-0.1272	-0.1089	-7.5750e-01
A7	7	25.0000	50.0000	25.0000	-0.1492	-0.1482	-0.1484	-0.1456	-0.1408	-0.1365	-0.1225	-0.1039	-6.6432e-01
A8	8	50.1000	0.0000	49.9000	-0.1461	-0.1466	-0.1465	-0.1430	-0.1393	-0.1333	-0.1187	-9.7768e-02	-6.5721e-01
A9	9	66.6000	33.4000	0.0000	-0.1373	-0.1387	-0.1388	-0.1356	-0.1296	-0.1212	-0.1052	-8.1013e-02	-4.1380e-01
A10	10	0.0000	33.4000	66.6000	-0.1601	-0.1598	-0.1606	-0.1584	-0.1547	-0.1491	-0.1391	-0.1254	-9.3138e-01
A11	11	75.0000	0.0000	25.0000	-0.1375	-0.1390	-0.1404	-0.1381	-0.1317	-0.1231	-0.1068	-8.3593e-02	-4.6619e-01
A12	12	25.0000	0.0000	75.0000	-0.1547	-0.1553	-0.1555	-0.1530	-0.1482	-0.1429	-0.1304	-0.1142	-8.4702e-01
A13	13	33.4000	66.6000	0.0000	-0.2614	-0.2616	-0.2573	-0.2519	-0.2448	-0.2367	-0.2241	-0.2011	-0.1600
A14	14	0.0000	66.7000	33.3000	-0.2751	-0.2735	-0.2687	-0.2649	-0.2581	-0.2508	-0.2401	-0.2233	-0.1800
A15	15	33.4000	33.0000	33.6000	-0.2659	-0.2661	-0.2614	-0.2565	-0.2508	-0.2430	-0.2305	-0.2096	-0.1700
A16	16	75.0000	0.0000	25.0000	-0.1375	-0.1390	-0.1404	-0.1381	-0.1317	-0.1231	-0.1068	-8.3593e-02	-4.6619e-01
B1	17	100.0000	0.0000	0.0000	-0.1077	-0.1100	-0.1092	-0.1068	-0.1013	-9.3201e-02	-7.5410e-02	-4.7124e-02	-8.9750e-01
B2	18	0.0000	100.0000	0.0000	-0.1368	-0.1347	-0.1345	-0.1318	-0.1292	-0.1250	-0.1162	-9.7305e-02	-5.4578e-01
B3	19	0.0000	0.0000	100.0000	-0.1468	-0.1462	-0.1189	-0.1448	-0.1421	-0.1372	-0.1290	-0.1174	-9.0114e-01
B4	20	50.0000	50.0000	0.0000	8.2190e-02	7.9466e-02	8.2398e-02	8.4666e-02	8.3832e-02	9.1474e-02	0.1068	0.1291	0.1600
B5	21	50.0000	0.0000	50.0000	-0.1285	-0.1299	-0.1276	-0.1274	-0.1226	-0.1165	-0.1046	-8.5415e-02	-5.2313e-01
B6	22	0.0000	49.9000	50.1000	-0.1432	-0.1440	-0.1430	-0.1407	-0.1350	-0.1258	-0.1228	-0.1104	-7.4456e-01
B7	23	75.0000	0.0000	25.0000	-0.1208	-0.1218	-0.1209	-0.1191	-0.1149	-0.1060	-9.1533e-02	-6.7456e-02	-3.1735e-01
B8	24	25.7000	0.0000	74.3000	-0.1380	-0.1388	-0.1377	-0.1360	-0.1335	-0.1285	-0.1188	-0.1016	-7.0787e-01
B9	25	50.0000	25.0000	25.0000	-0.1279	-0.1285	-0.1286	-0.1263	-0.1230	-0.1151	-0.1027	-8.0337e-02	-4.3663e-01
B10	26	25.0000	50.0000	25.0000	-0.1326	-0.1327	-0.1319	-0.1296	-0.1265	-0.1213	-0.1099	-9.1401e-02	-5.2486e-01
B11	27	25.0000	25.0000	50.0000	-0.1349	-0.1348	-0.1340	-0.1329	-0.1287	-0.1238	-0.1127	-9.5241e-02	-6.1134e-01
P1	28	50.1000	0.0000	49.9000	-0.1491	-0.1497	-0.1496	-0.1461	-0.1423	-0.1363	-0.1218	-0.1008	-6.8768e-01
P2	29	66.6000	33.4000	0.0000	-0.1568	-0.1582	-0.1583	-0.1551	-0.1491	-0.1407	-0.1247	-0.1005	-6.0881e-01
P3	30	0.0000	33.4000	66.6000	-0.1348	-0.1346	-0.1353	-0.1331	-0.1294	-0.1239	-0.1138	-0.1002	-6.7886e-01
P4	31	25.0000	0.0000	75.0000	-0.1407	-0.1413	-0.1415	-0.1390	-0.1343	-0.1290	-0.1164	-0.1002	-7.0710e-01
P5	32	33.4000	66.6000	0.0000	-0.1656	-0.1658	-0.1615	-0.1561	-0.1490	-0.1409	-0.1283	-0.1053	-6.4390e-01
P6	33	0.0000	66.7000	33.3000	-0.1553	-0.1537	-0.1489	-0.1450	-0.1382	-0.1309	-0.1203	-0.1035	-6.6051e-01





	znak	Additivated	Type	octane	1100.0	1102.0	1104.0	1106.0	1108.0	1110.0	1112.0	1114.0	
	1	2	3	4	5	6	7	8	9	10	11	12	
MO1	1	M	No	Medium	88.6000	-1.8345e-03	-1.4965e-03	-1.0088e-03	-3.6036e-04	4.6595e-04	1.5021e-03	2.7976e-03	4.4103e-03
MO2	2	M	No	Medium	88.8000	-1.6270e-03	-1.2608e-03	-7.4818e-04	-6.5377e-05	8.1759e-04	1.9427e-03	3.3538e-03	5.0884e-03
MO5	3	M	No	Medium	89.4000	-4.6487e-04	-3.3003e-04	1.8072e-04	1.1213e-03	2.4647e-03	4.1399e-03	6.0749e-03	8.2380e-03
LO6	4	L	No	Low	86.7000	-1.4426e-03	-1.0941e-03	-5.9577e-04	8.1742e-05	9.7191e-04	2.1136e-03	3.5445e-03	5.2987e-03
H11	5	H	No	High	91.2000	-1.6075e-03	-1.1549e-03	-5.4838e-04	2.4382e-04	1.2589e-03	2.5514e-03	4.1770e-03	6.1885e-03
H12	6	H	No	High	91.3000	-5.0201e-04	-5.3970e-06	6.5686e-04	1.5220e-03	2.6312e-03	4.0366e-03	5.7912e-03	7.9449e-03
L13	7	L	No	Low	87.4000	-1.1307e-03	-8.0107e-04	-3.2307e-04	3.1966e-04	1.1486e-03	2.2039e-03	3.5410e-03	5.2250e-03
L14	8	L	No	Low	87.1000	-1.9691e-03	-1.6273e-03	-1.1357e-03	-4.6888e-04	3.9858e-04	1.5020e-03	2.8750e-03	4.5478e-03
L15	9	L	No	Low	87.0000	-1.2943e-03	-9.5174e-04	-4.7322e-04	1.7785e-04	1.0386e-03	2.1510e-03	3.5508e-03	5.2641e-03
H17	10	H	No	High	91.8000	-9.8381e-04	-5.4232e-04	4.6605e-05	8.2156e-04	1.8227e-03	3.0996e-03	4.7021e-03	6.6691e-03
M18	11	M	No	Medium	89.1000	-1.2279e-03	-8.4948e-04	-3.2593e-04	3.6318e-04	1.2407e-03	2.3428e-03	3.7161e-03	5.4172e-03
H20	12	m	No	High	91.8000	-1.5178e-03	-1.1265e-03	-5.8910e-04	1.1130e-04	9.9650e-04	2.1085e-03	3.5048e-03	5.2608e-03
L21	13	L	No	Low	86.9000	-1.6707e-03	-1.3024e-03	-7.7863e-04	-6.9812e-05	8.5402e-04	2.0345e-03	3.5182e-03	5.3478e-03
H24	14	H	No	High	91.7000	-1.5543e-03	-1.0224e-03	-3.3135e-04	5.5202e-04	1.6620e-03	3.0513e-03	4.7768e-03	6.8948e-03
H27	15	H	No	High	91.7000	-1.5807e-03	-1.1741e-03	-6.1728e-04	1.0519e-04	1.0117e-03	2.1404e-03	3.5527e-03	5.3292e-03
L29	16	L	No	Low	87.0000	-1.2918e-03	-9.2109e-04	-4.0504e-04	2.8513e-04	1.1825e-03	2.3282e-03	3.7648e-03	5.5238e-03
L31	17	L	No	Low	87.0000	-1.6437e-03	-1.2955e-03	-8.0905e-04	-1.4881e-04	7.2328e-04	1.8474e-03	3.2591e-03	4.9862e-03
H32	18	m	No	High	90.8000	-3.8350e-04	2.1294e-05	5.8253e-04	1.3267e-03	2.2817e-03	3.4862e-03	4.9827e-03	6.8108e-03
L35	19	L	No	Low	87.2000	-8.8136e-04	-4.9750e-04	3.6174e-05	7.5037e-04	1.6775e-03	2.8577e-03	4.3325e-03	6.1324e-03
H36	20	H	No	High	91.4000	-1.4990e-03	-1.0743e-03	-4.8915e-04	2.8680e-04	1.2870e-03	2.5557e-03	4.1320e-03	6.0541e-03
L37	21	L	No	Low	87.2000	-1.5063e-03	-1.1689e-03	-6.7099e-04	2.5389e-05	9.5641e-04	2.1617e-03	3.6798e-03	5.5443e-03
H38	22	H	No	High	92.2000	-1.5216e-03	-1.0764e-03	-4.8081e-04	2.9840e-04	1.2978e-03	2.5669e-03	4.1509e-03	6.0884e-03
H39	23	H	No	High	91.8000	-1.7410e-03	-1.2751e-03	-6.3829e-04	1.9978e-04	1.2739e-03	2.6287e-03	4.3237e-03	6.4285e-03
L40	24	L	No	Low	87.0000	-1.4102e-03	-1.0815e-03	-6.1008e-04	2.8631e-05	8.5912e-04	1.9147e-03	3.2390e-03	4.8770e-03
M52	25	M	Yes	Medium	89.0000	-3.8799e-03	-3.5395e-03	-3.0391e-03	-2.3566e-03	-1.4673e-03	-3.2818e-04	1.1077e-03	2.8825e-03
H59	26	H	Yes	High	92.4000	-4.4693e-03	-4.0828e-03	-3.5466e-03	-2.8335e-03	-1.9128e-03	-7.4202e-04	7.2229e-04	2.5238e-03
S.003	27	S	No	m	88.6000	1.5370e-03	1.8477e-03	2.2950e-03	2.9058e-03	3.7118e-03	4.7558e-03	6.0798e-03	7.7240e-03
S.004	28	S	No	m	88.8000	-8.1778e-04	-4.5066e-04	6.2463e-05	7.4953e-04	1.6390e-03	2.7725e-03	4.1896e-03	5.9301e-03
S.010	29	S	No	m	91.2000	-1.9827e-03	-1.5751e-03	-1.0175e-03	-2.7478e-04	6.9399e-04	1.9363e-03	3.4989e-03	5.4213e-03
S.016	30	S	No	m	91.8000	-1.6016e-03	-1.1583e-03	-5.6356e-04	2.1848e-04	1.2261e-03	2.5086e-03	4.1197e-03	6.1020e-03
S.019	31	S	No	m	89.0000	-1.9282e-03	-1.6082e-03	-1.1418e-03	-5.1407e-04	2.9262e-04	1.3125e-03	2.5958e-03	4.2059e-03
S.022	32	S	No	m	91.4000	-1.8353e-03	-1.4577e-03	-9.3005e-04	-2.1984e-04	7.0978e-04	1.8994e-03	3.3892e-03	5.2133e-03
S.025	33	S	No	m	88.6000	-1.0111e-03	-5.8962e-04	-1.1423e-05	7.5970e-04	1.7601e-03	3.0378e-03	4.6402e-03	6.6052e-03



Označení metod

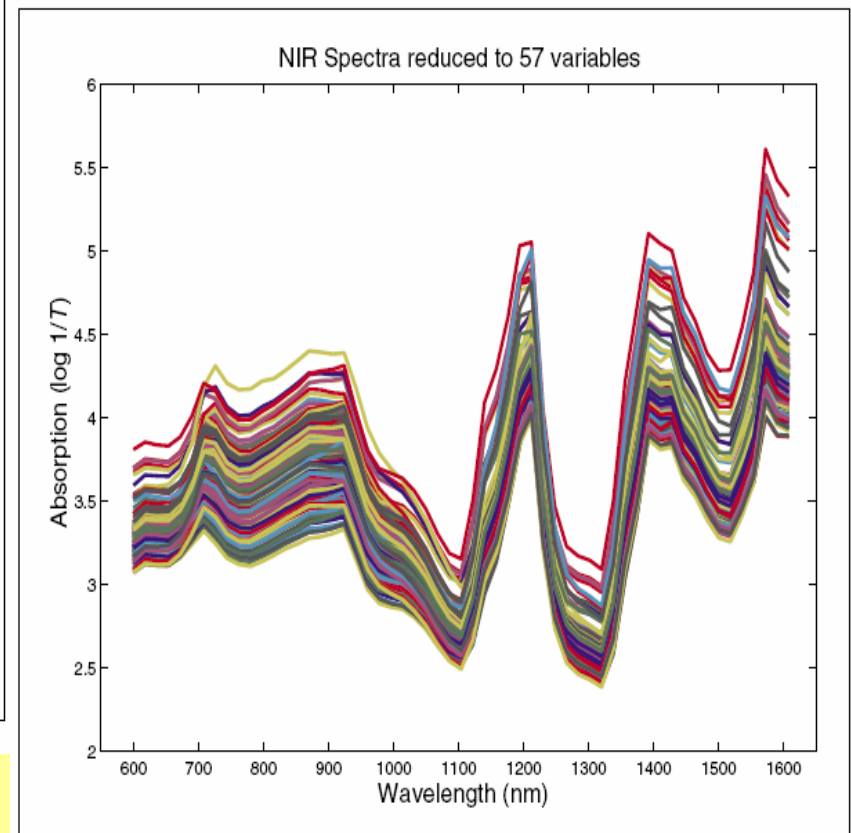
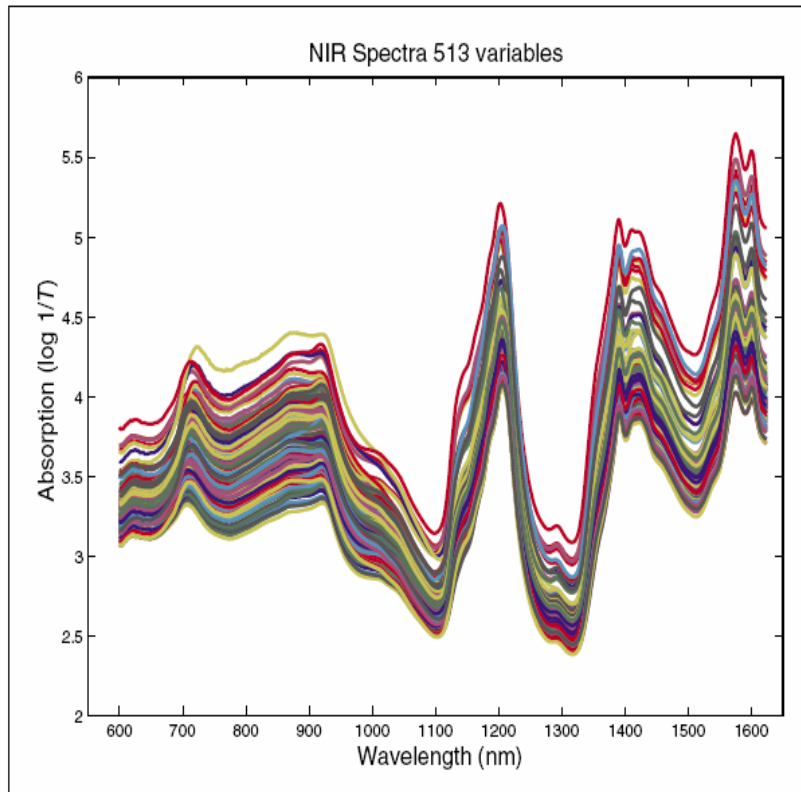
- **CLS** - classical least squares
- **ILS** - inverse least squares
- **MLR** - multiple linear regression
- **PCA** - principal component analysis
- **LDA** - linear discriminant analysis
- **PCR** - principal component regression
- **PLS1** - partial least squares 1
- **PLS2** - partial least squares 2

Analýza hlavních komponent

- reorganizace informace v souboru dat – hledání struktury dat – NIKOLI REGRESE, NIKOLI KVANTITA
- analýza samotných spektrálních dat
- analýza souboru dat s velkým počtem proměnných
- redukce počtu proměnných na relevantní proměnné popisující variabilitu dat, zachování informačního obsahu

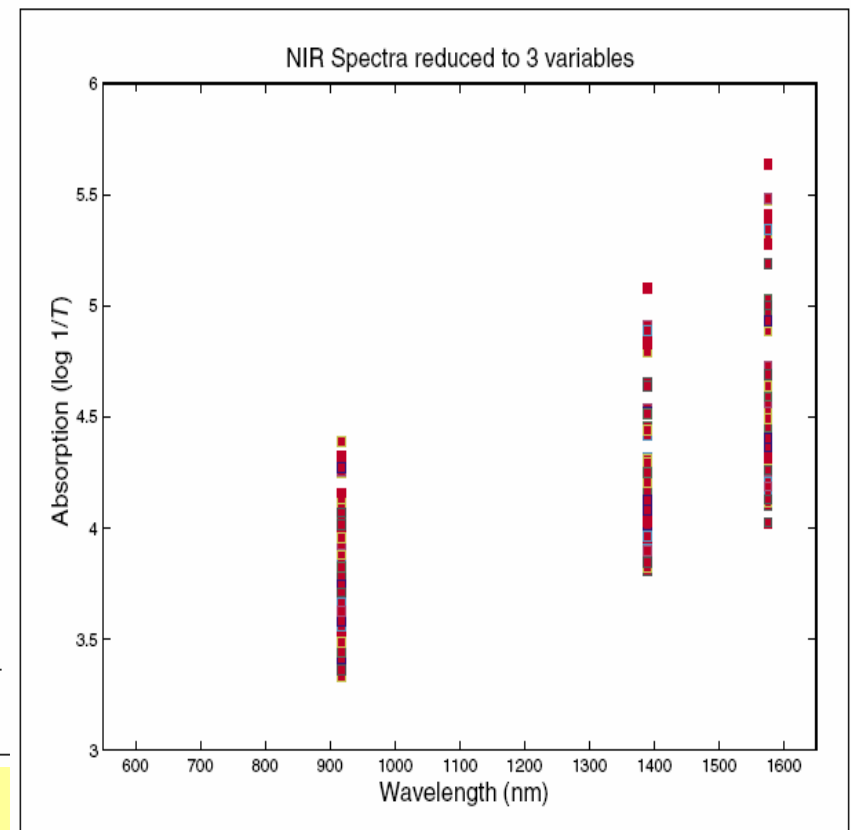
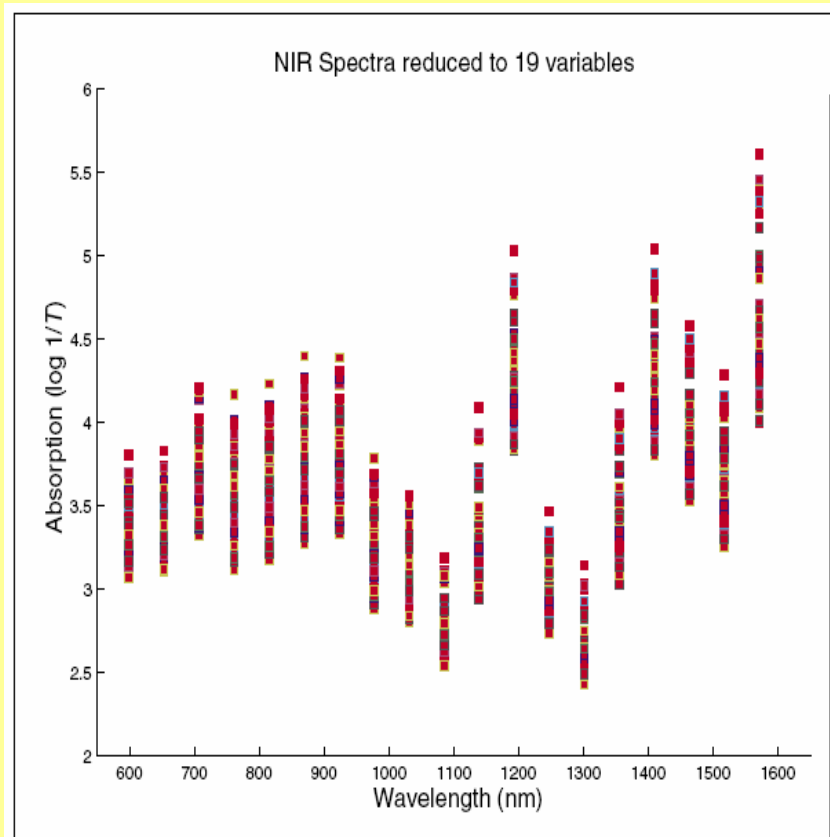
Analýza hlavních komponent

- PŘÍKLAD



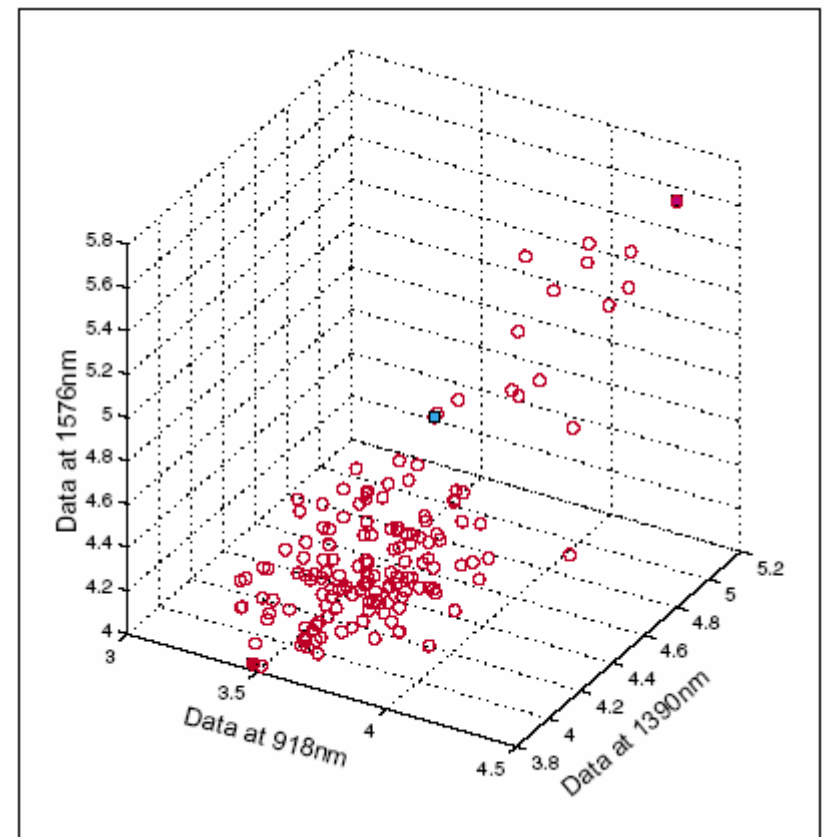
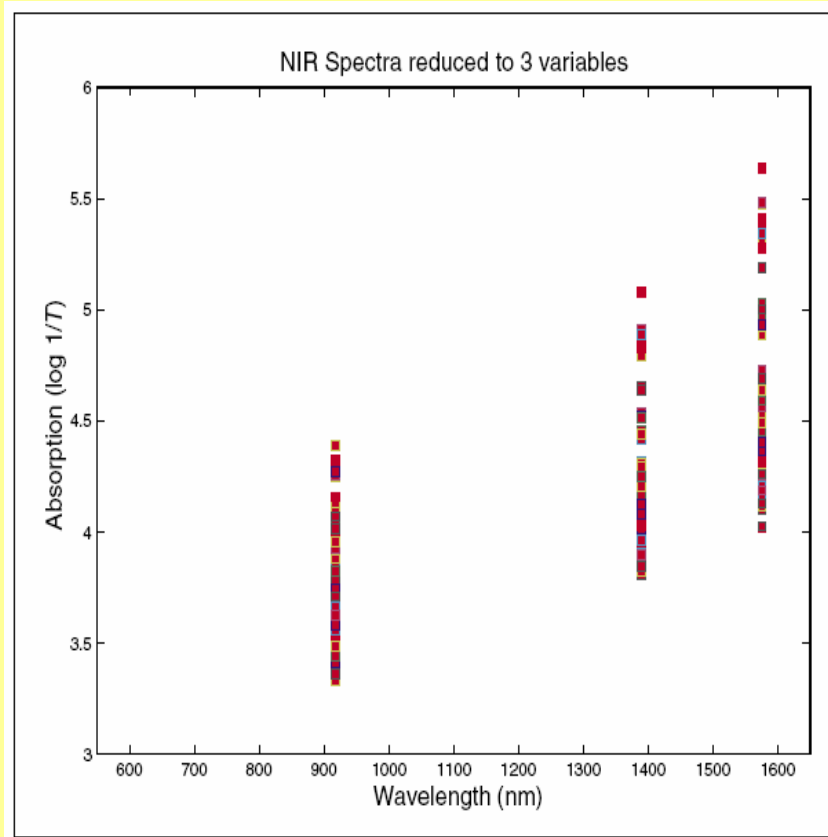
Analýza hlavních komponent

- PŘÍKLAD



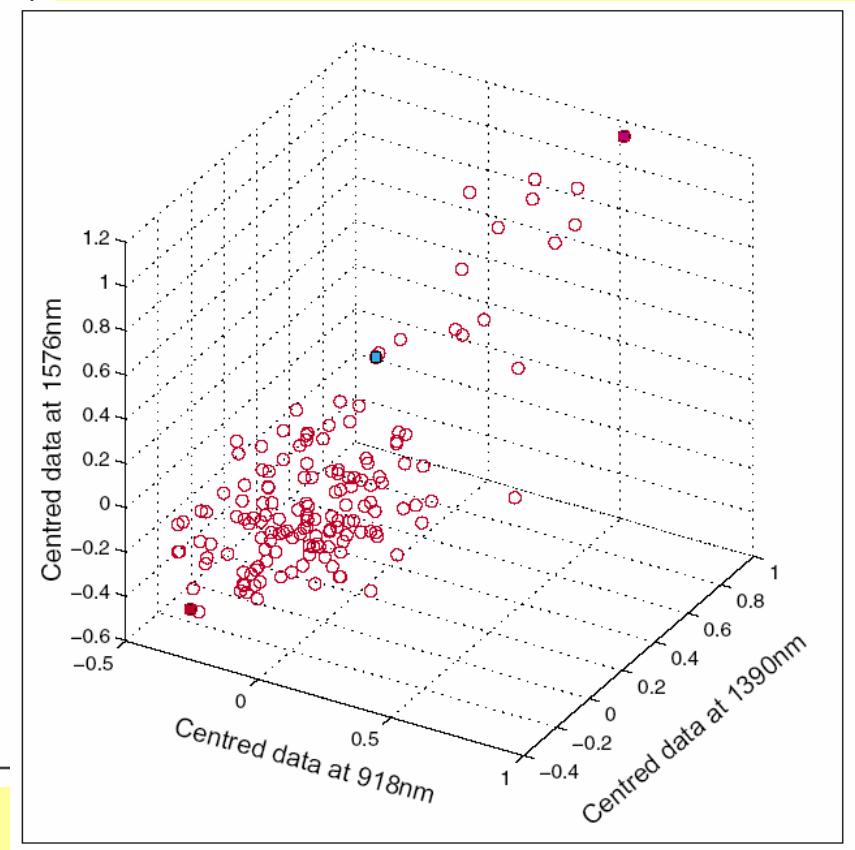
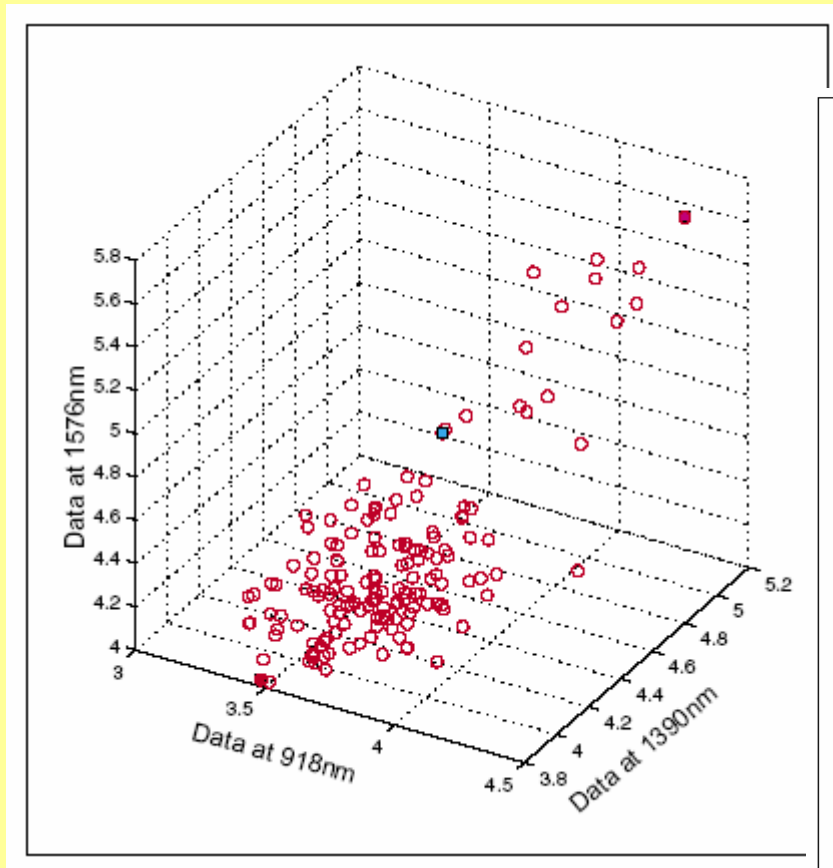
Analýza hlavních komponent

- PŘÍKLAD



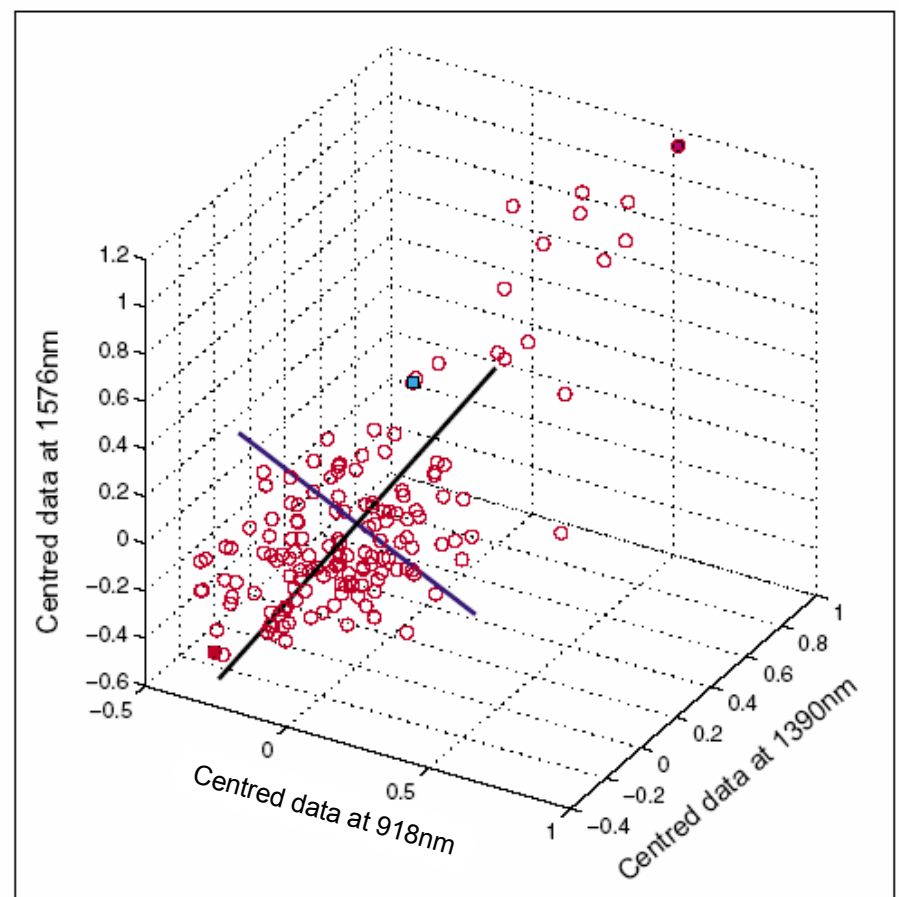
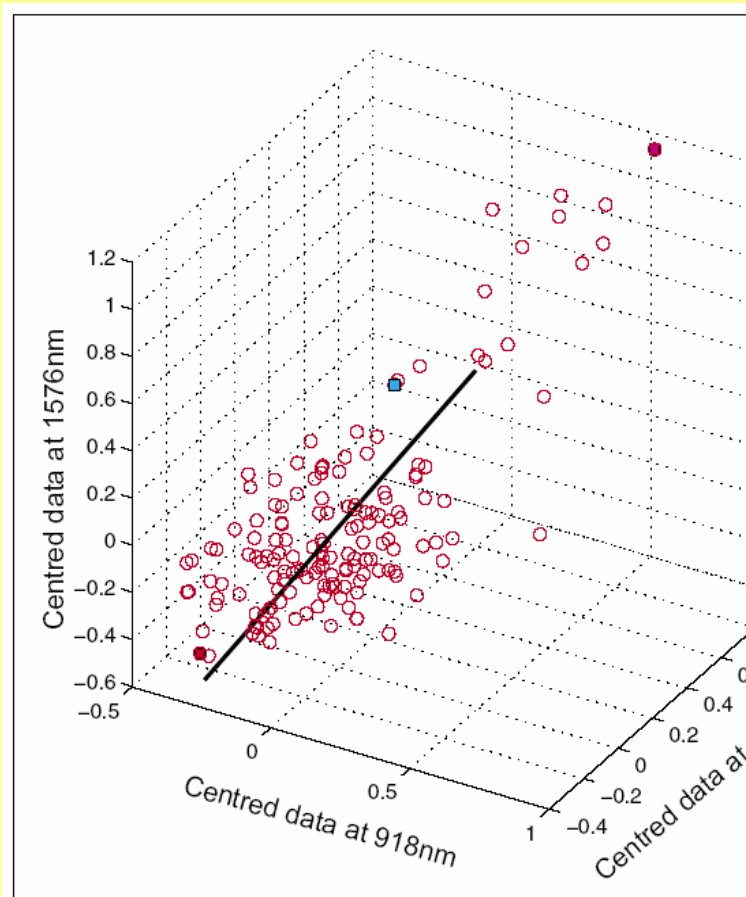
Analýza hlavních komponent

- PŘÍKLAD



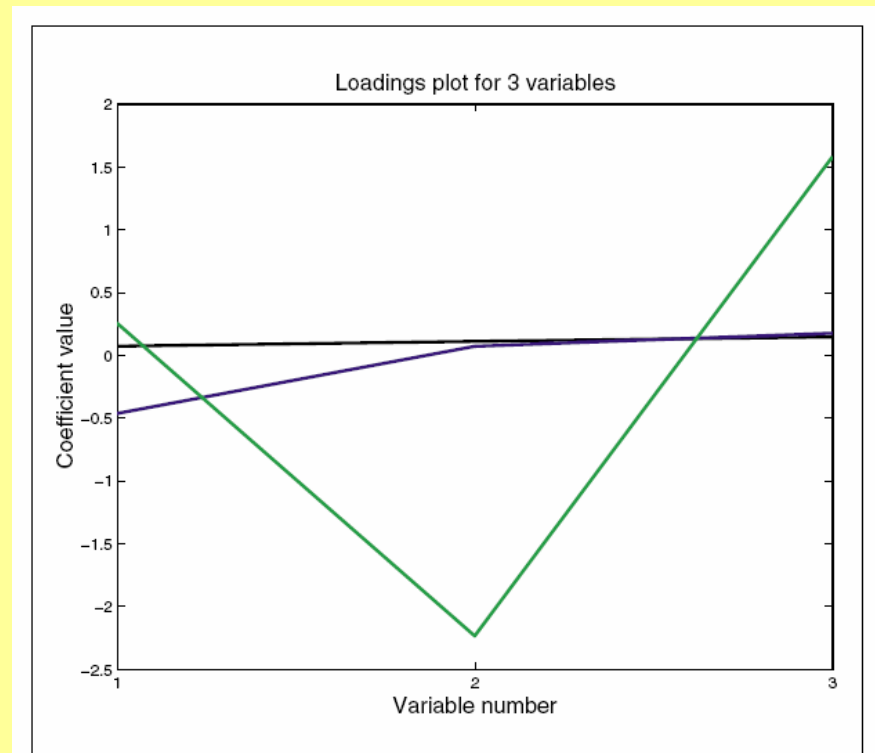
Analýza hlavních komponent

- PŘÍKLAD



Analýza hlavních komponent

- PŘÍKLAD – zátěže, váhy, loadings
 - Velikost absolutní hodnoty
 - Znaménko
 - Příspěvek dané proměnné k jednotlivým hlavním komponentám



Plot of the weights for each PC.
— 1st PC; — 2nd PC; — 3rd PC.

Analýza hlavních komponent

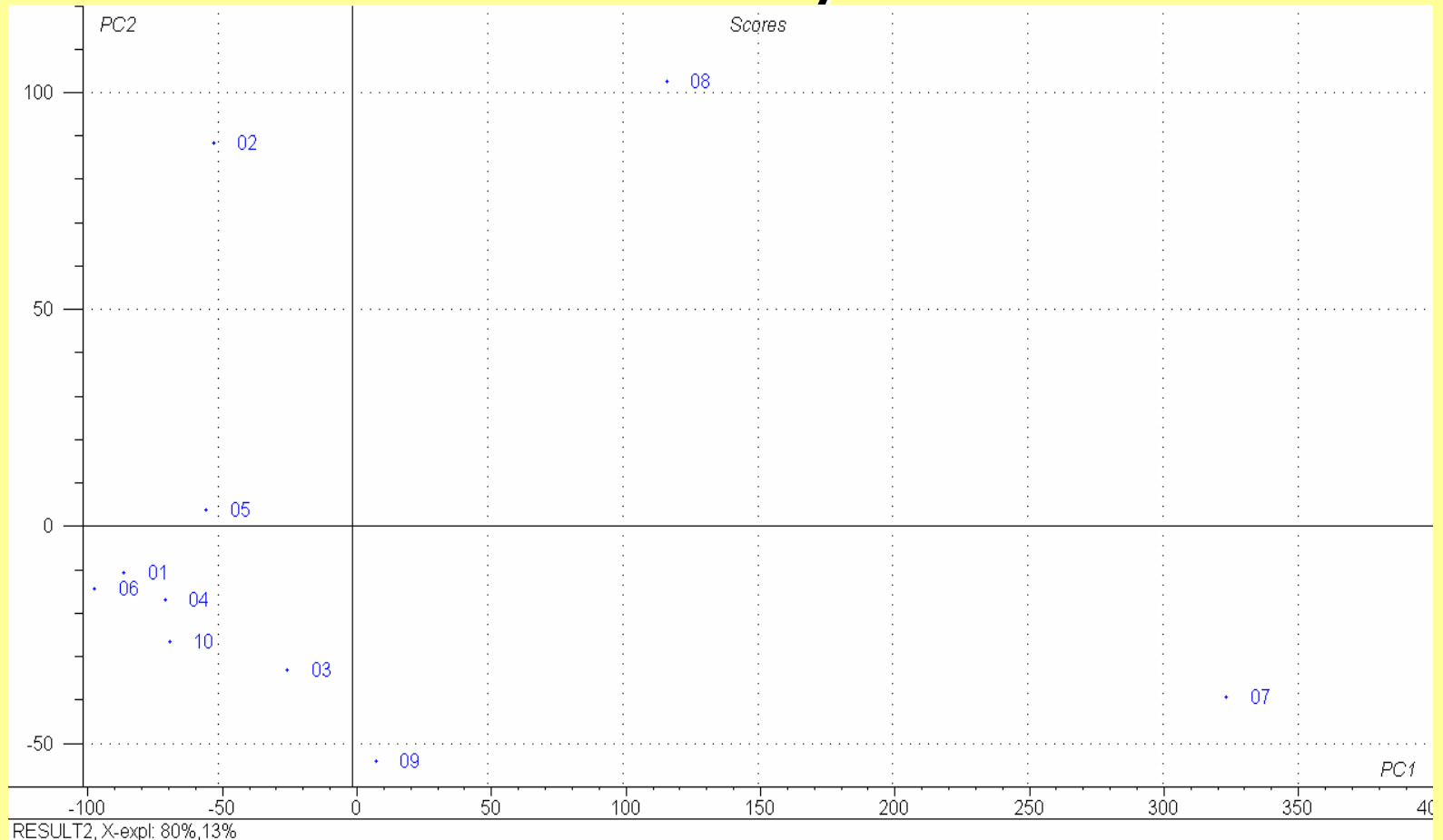
- Příklad – složení vody

The screenshot shows the software interface for 'The Unscrambler - [PCA-vody]'. The menu bar includes File, Edit, View, Plot, Modify, Task, Results, Window, and Help. The toolbar contains various icons for file operations, editing, and analysis. Below the toolbar is a data table with the following structure:

		Ca	Mg	Na	Cl-	SO4
		1	2	3	4	5
voda01	1	36.6000	6.2600	1.0400	5.0000	25.0000
voda02	2	38.6000	144.0000	6.8000	18.2000	24.5000
voda03	3	69.7000	20.4000	60.8000	12.6000	42.4000
voda04	4	45.9000	14.3000	18.1000	2.9400	36.4000
voda05	5	83.9000	19.0000	1.5000	3.0000	19.4000
voda06	6	10.0000	8.6000	10.0000	0.5000	2.7000
voda07	7	270.0000	71.4000	277.8000	185.4000	0.2500
voda08	8	185.2000	107.1000	41.2000	144.2000	0.1850
voda09	9	84.1000	30.6000	103.0000	10.5000	64.7000
voda10	10	42.8000	10.5000	25.4000	4.4600	51.7000

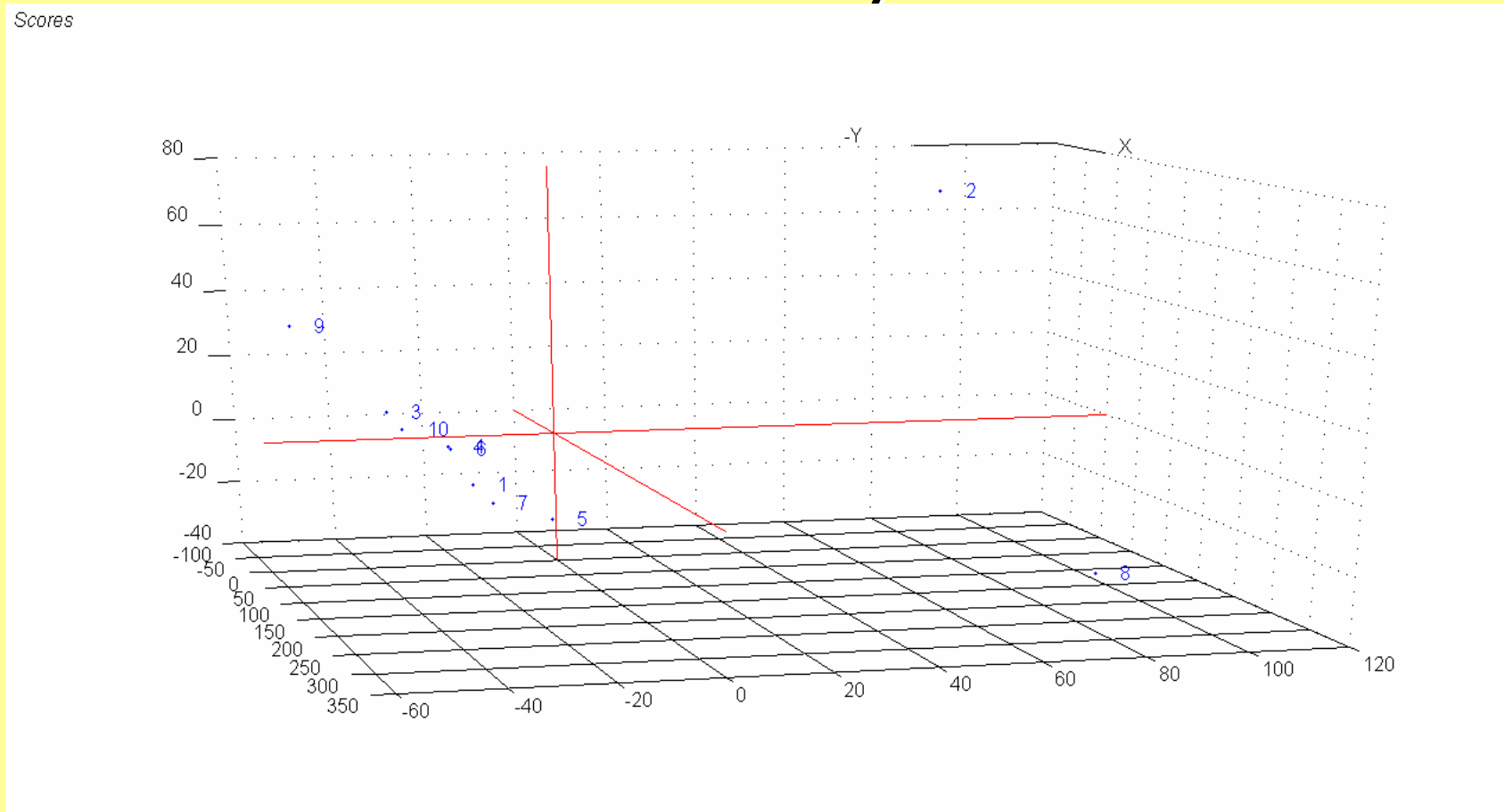
Analýza hlavních komponent

- Příklad – složení vody



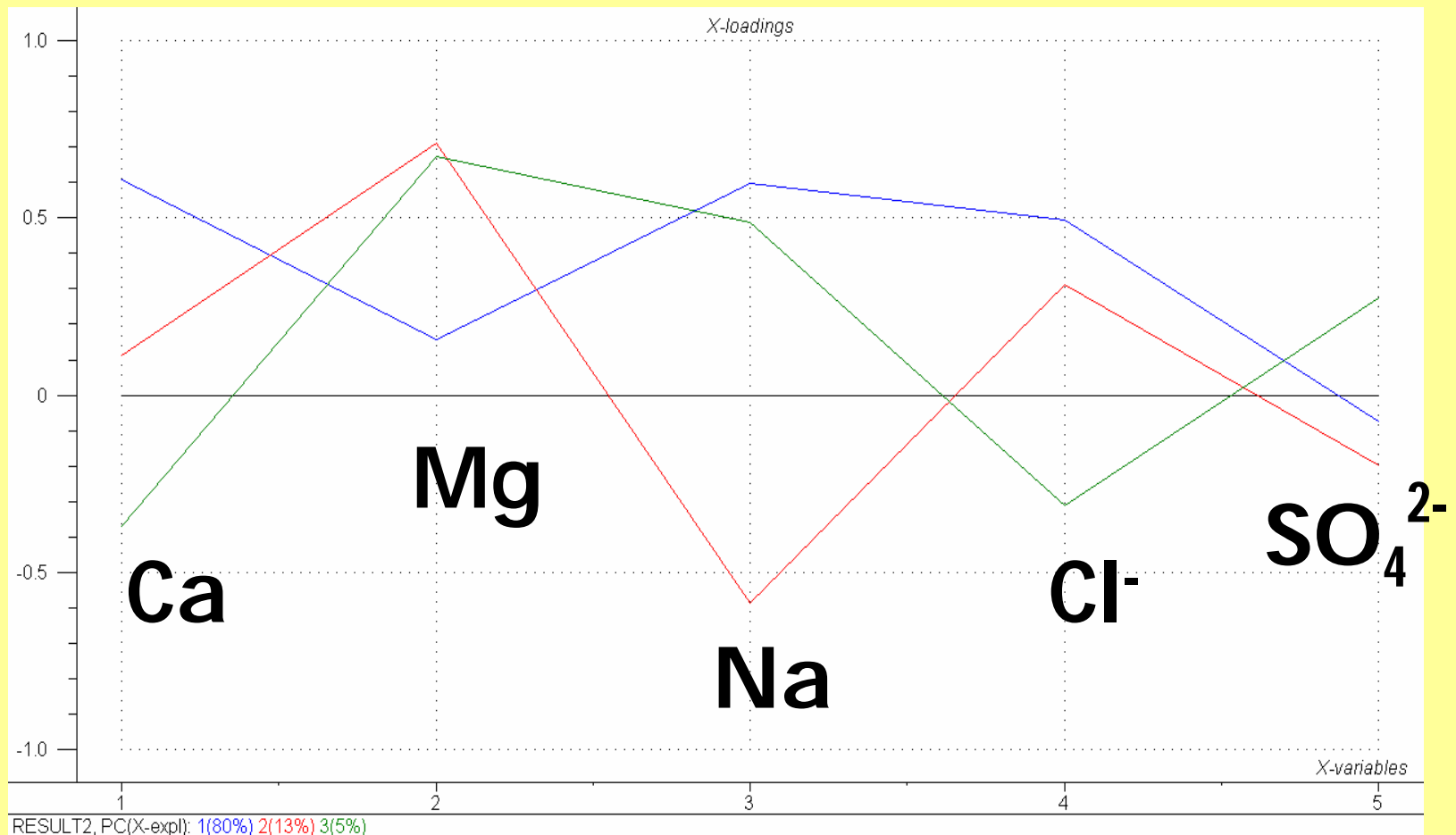
Analýza hlavních komponent

- Příklad – složení vody



Analýza hlavních komponent

- Příklad – složení vody



Analýza hlavních komponent

- Příklad – složení vody

