

Optimalizace parametrů hmotnostního detektoru (MS/MS) s využitím vícerozměrných statistických analýz

Kamil Šťastný
říjen 2020

**Licenční studium PYTHAGORAS
Statistické zpracování experimentálních dat**

**Univerzita Pardubice
Fakulta chemicko-technologická, Katedra analytické chemie**

Obsah

1. Úvod
2. Zadání
3. Naměřená data
4. Analýza a interpretace dat
 - 4.1 Explatorní analýza dat, EDA
 - 4.2 Analýza hlavních komponent, PCA
 - 4.3 Porovnání grafických výstupů PCA a Shlukové analýzy CLU
 - 4.4 PCA pro Beta-agonisty a Nitrofurany
 - 4.5 Faktorová analýza, FA
5. Závěr

1. Úvod

Hmotnostní detektory na bázi kvadrupolů v současnosti nejčastěji tandemově spřažených, označovaných jako „triple-Q“ nebo MS/MS detektory, se jeví jako nejvhodnější pro organickou stopovou analýzu rezidui farmaceutik a tzv. zakázaných látek v biologických matricích a to pro svou poměrně vysokou citlivost a selektivitu. Techniky hmotnostní spektrometrie s kvadrupolovými detektory MS/MS mají pro laboratoře, kde jedním z hlavních cílů je prokázat přítomnost nebo nepřítomnost zakázané látky v testovaném vzorku, ještě jednu nezanedbatelnou výhodu oproti metodám chromatografickým, elektromigračním či spektrálním a to rozšířením analýzy o možnost confirmace, tj. potvrzení identity stanovovaného analytu. Hmotnostní detektor, který díky své konstrukci v sobě zahrnuje všechny tři výše popsané výhody (vysoká citlivost, selektivita a možnost confirmace) je zařízení TSQ Vantage firmy Thermo-scientific, USA viz. Obr.1,2 a 3.

1. Úvod



2. Zadání

- Software Excaliber v.2.1 hmotnostního detektoru TSQ Vantage od firmy Thermo Scientific provádí automatickou optimalizaci všech parametrů detektoru ale vždy jen pro jeden analyt, tedy potřebujeme-li stanovovat dva a více analytů v jednom nástřiku (v jednom měření), musíme na základě experimentálních zkušeností s tímto zařízením najít a nastavit parametry detektoru tak, aby společně vyhovovaly více analytům. Jinými slovy, analytik ve své každodenní práci provádí „optimalizaci“ parametrů MS/MS detektoru pro více než jeden analyt na základě svých experimentálních zkušeností a naměřených parametrů pro jednotlivé analyty. Ale trendem v dnešní analytické chemii ve stopové analýze při kontrole potravin na zakázané látky je získat z jednoho vzorku co nejvíce informací v krátkém čase, tedy využívání multireziduálních metod stanovení co největšího počtu analytů v jednom nástřiku (měření).
- Statistické metody PCA, FA a CLU by měly odpovědět na otázky, jest-li mají sledované analyty podobné fyzikálně-chemické vlastnosti (tvoří shluky) vhodné pro MS/MS stanovení? Existují vnitřní vztahy mezi optimalizovanými parametry detektoru? A jak spolu souvisí nebo nesouvisí (kladně, záporně korelují nebo nekorelují) parametry MS/MS detektoru?

3. *Naměřená data*

Experimentální data jednotlivých parametrů detektoru byla získána při automatické optimalizaci na zařízení TSQ Vantage s ESI iontovým zdrojem a při nástřiku jednotlivých standardů analytů s konstantním průtokem.

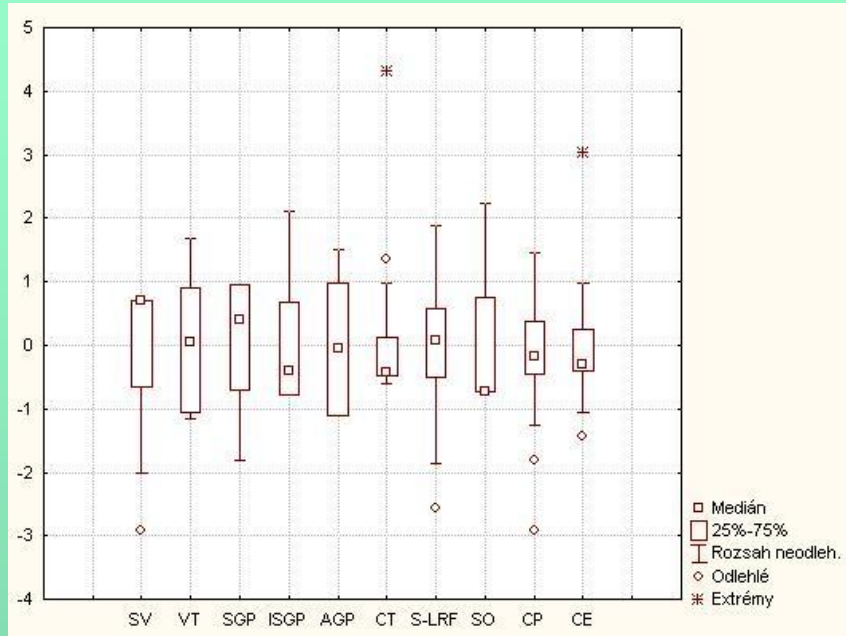
1. Byly naměřeny parametry optimálního nastavení MS/MS detektoru pro jednotlivé analyty ze skupiny Beta-agonistů v počtu 20 standardů a 9 izotopově značených interních standardů IS.
2. Pro další testování byly na zkoušku přidány do optimalizace další 3 analyty z jiné skupiny zakázaných látek a to Nitrofurany, pro které by mohl být použit stejný postup přípravy zpracování vzorku jako pro skupinu Beta-agonistů.

3. *Naměřená data*

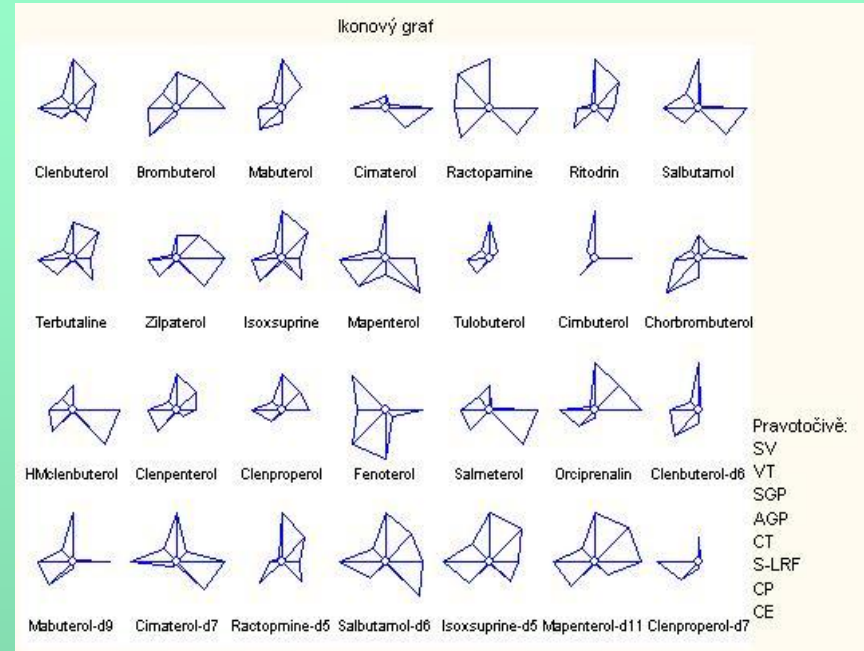
10 parametrů (znaků), které optimalizuje MS/MS detektor TSQ Vantage:

- **SV** Spray Voltage (V)
- **VT** Vaporizer temperature (C)
- **SGP** Steath Gas Pressure (psi)
- **SGP** Ion Sweep Gas Pressure (psi)
- **AGP** Aux Gas Pressure (psi)
- **CT** Capillary Temperature (C)
- **S-LRF** S-Lens RF Amplitude voltage (V)
- **SO** Skimmer Offset (V)
- **CP** Collision Pressure (mTorr)
- **CE** Collision Energy (V)

4.1 Exploratorní analýza dat, EDA



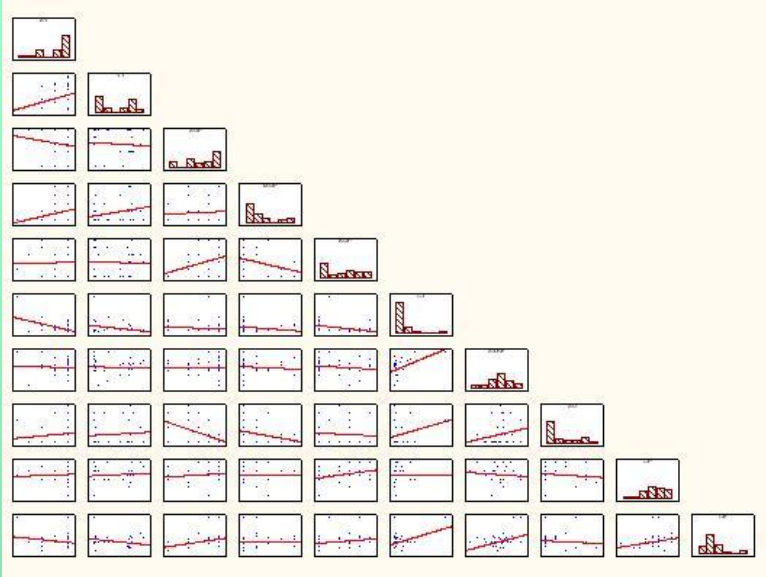
Obr.1.1 Krabicový graf pro všechny znaky a všechny objekty (STATISTICA)



Obr.1.2 Symbolové grafy – hvězdicový

4.1 Explorativní analýza dat, EDA

Obr.1.3 Maticový graf korelací dat a histogramy (STATISTICA)



Obr.1.3 v grafické podobě a Tabulka v číselné podobě ukazuje na korelace nízkého významu mezi jednotlivými dvojicemi znaků zdrojové matice dat a proto další zpracování dat pomocí analýzy PCA a FA může vést k potížím s hledáním latentních proměnných a určením jejich odpovídajícího počtu.

Tabulka -Korelační matice párových korelačních koeficientů (STATISTICA)

	SV	VT	SGP	ISGP	AGP	CT	S-LRF	SO	CP	CE
SV	1,0000	0,3457	-0,2000	0,2809	0,0239	-0,4696	-0,0626	0,1141	0,0750	-0,1938
VT	0,3457	1,0000	-0,0852	0,2535	-0,0201	-0,2614	-0,0636	0,0885	0,1162	-0,2412
SGP	-0,2000	-0,0852	1,0000	0,0805	0,3992	-0,0363	0,0002	-0,5329	0,1510	0,3640
ISGP	0,2809	0,2535	0,0805	1,0000	-0,2966	-0,1815	-0,1023	-0,2605	-0,0240	-0,0315
AGP	0,0239	-0,0201	0,3992	-0,2966	1,0000	-0,3082	-0,1497	-0,0878	0,3609	0,1973
CT	-0,4696	-0,2614	-0,0363	-0,1815	-0,3082	1,0000	0,5620	0,2586	0,0079	0,4176
S-LRF	-0,0626	-0,0636	0,0002	-0,1023	-0,1497	0,5620	1,0000	0,2262	-0,1079	0,3938
SO	0,1141	0,0885	-0,5329	-0,2605	-0,0878	0,2586	0,2262	1,0000	-0,1435	-0,1143
CP	0,0750	0,1162	0,1510	-0,0240	0,3609	0,0079	-0,1079	-0,1435	1,0000	0,2401
CE	-0,1938	-0,2412	0,3640	-0,0315	0,1973	0,4176	0,3938	-0,1143	0,2401	1,0000

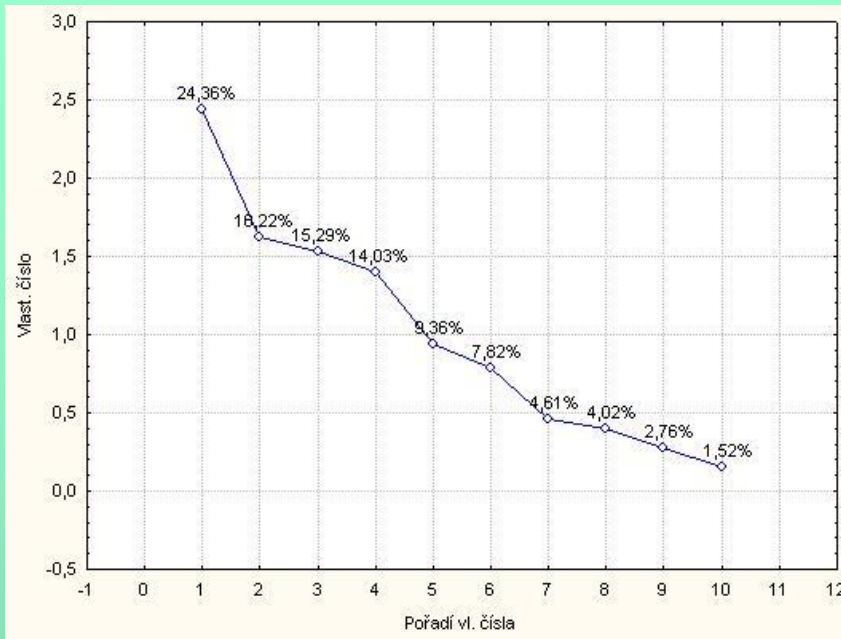
4.1 Explatorní analýza dat, EDA

Řešení:

Symbolové grafy (hvězdicový) pro porovnání objektů neukazují na žádné výrazně podobné objekty ale ukazují rozdílné objekty, v našem případě analyty. Všechny použité diagnostické grafy pro identifikaci vlivných bodů indikují výrazně odlehlý jeden bod (outlier) – Fenoterol, který výrazně negativně ovlivňuje vícerozměrnou normalitu a měl by být ze statistického hlediska z matice naměřených dat pro další analýzy vypuštěn.

Pro korelace nízkého významu mezi jednotlivými dvojicemi znaků zdrojové matice dat může další zpracování dat pomocí analýzy PCA a FA vést k potížím s hledáním latentních proměnných a určením jejich odpovídajícího počtu.

4.2 Analýza hlavních komponent, PCA



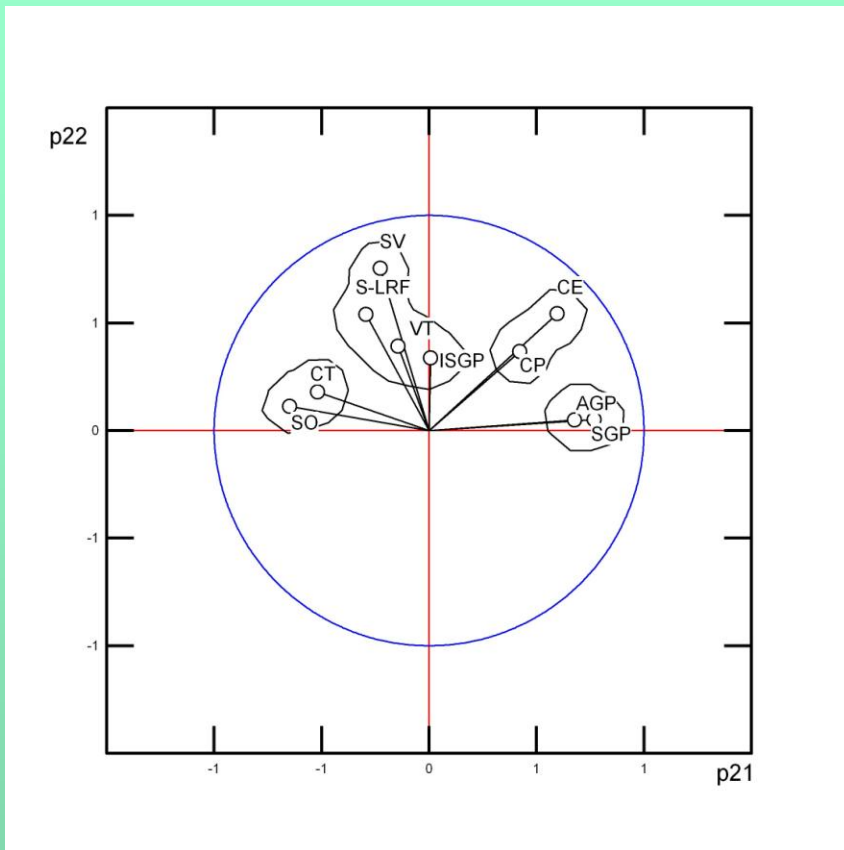
Obr 1.4 Cattelův indexový graf úpatí vlasních čísel bez Fenoterolu (STATISTICA)

Test of Validity of Latent Variables

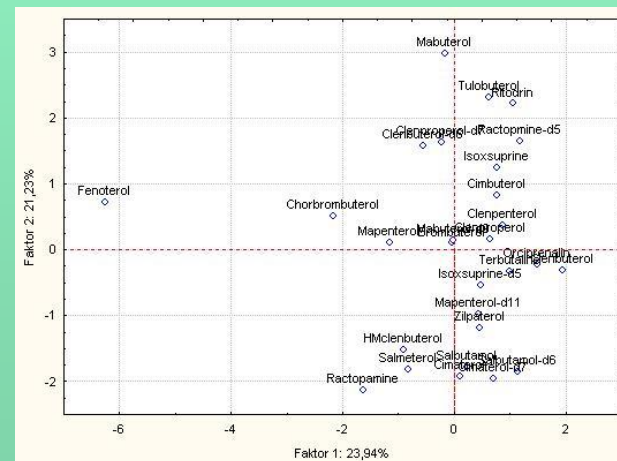
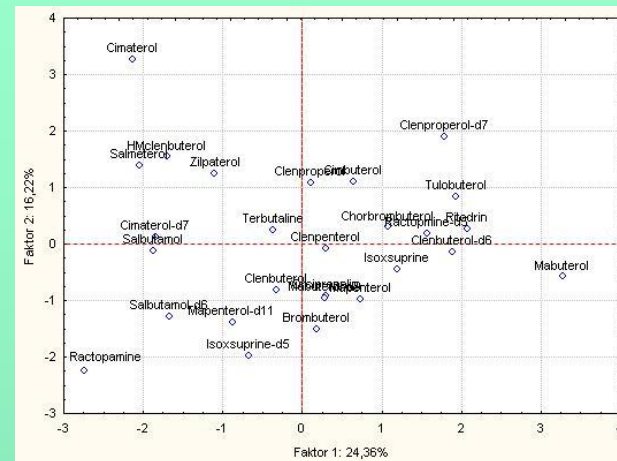
P	RSC(P)	ni(P-1)-ni(P)	ni(P)	F	F(0.95)	MEP(P)
1	1.538E+02					6.878E+00
2	1.039E+02	36	144	1.917	1.499	5.201E+00
3	7.208E+01	36	108	1.326	1.526	4.138E+00
4	4.966E+01	36	72	0.903	1.581	3.251E+00
5	2.996E+01	36	36	0.658	1.743	2.322E+00

Tabulka -Test validity hlavních komponent (data objektů bez Fenoterolu) OPstat

4.2 Analýza hlavních komponent, PCA



Obr 1.5 Graf komponentních vah ploch PC1 a PC2



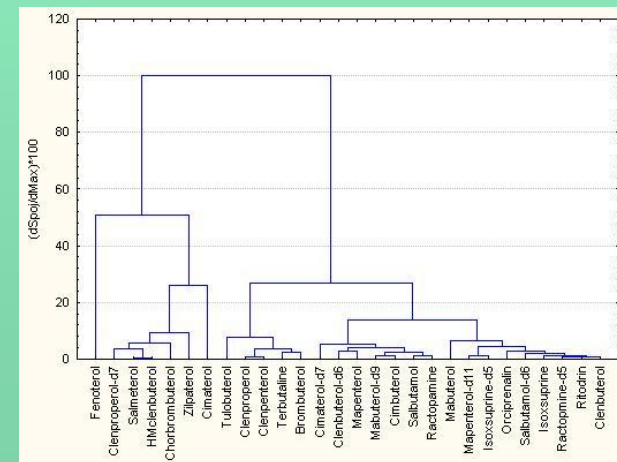
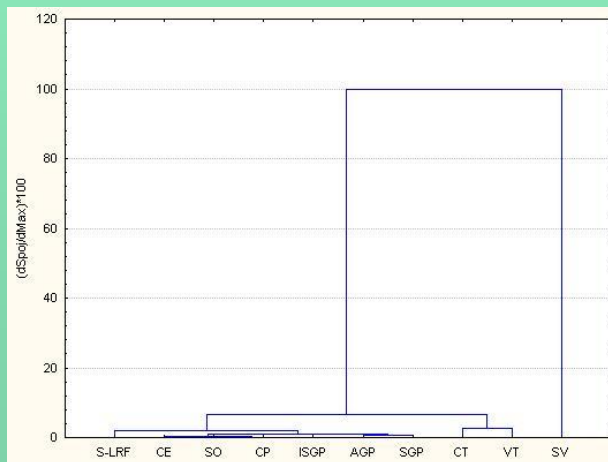
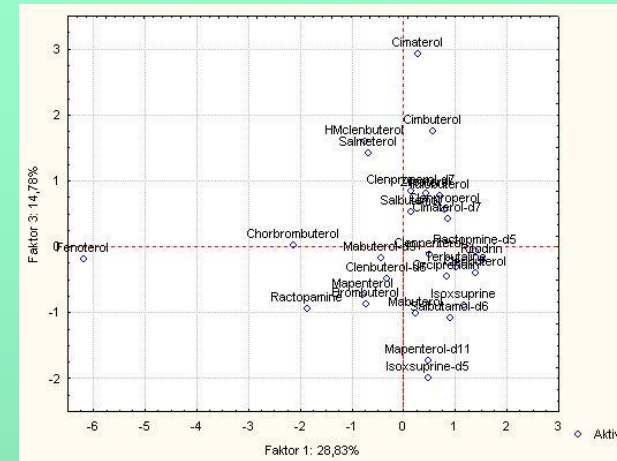
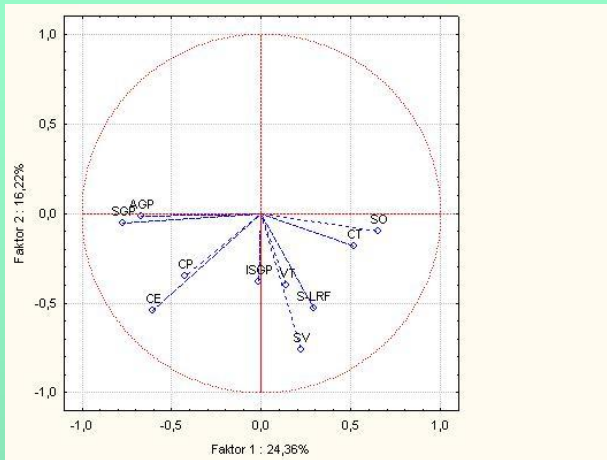
Obr 1.6 Rozptylový diagram komponentního skóre pro PC1 a PC2 (STATISTICA)

4.2 Analýza hlavních komponent, PCA

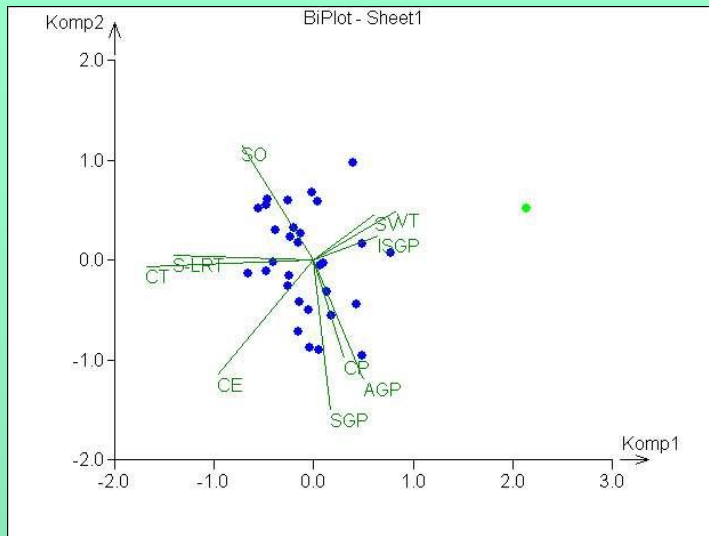
Řešení:

- Graf komponentních vah pro PC1 a PC2 ukazují na čtyři shluky znaků zřetelně od sebe oddělených, které naznačují na zajímavé korelace, například pozitivní korelaci mezi znaky SO a CT, která je v negativní korelaci ke znakům SGP a AGP. Znaky SO a CT souvisí s prostorem v detektoru Q0 těsně před stupem ionizovaných molekul do kvadrupolu Q1. Pozitivní korelace mezi znaky SGP a AGP (jsou to znaky, které přímo souvisí s kvalitou „spreje“ v iontovém zdroji) jednoznačně odpovídají experimentálním zkušenostem. Pokud zvyšujeme tlak proudu dusíku SGP a AGP, tak můžeme snižovat teplotu ve vstupní kapiláře CT a napětí na skimmeru SO.
- Znaky SV a VT se silnou korelací, s kterými korelují slaběji znaky S-LRT a ISGP tvoří další významný shluk. Jsou to právě ty znaky, které přímo souvisí s iontovým zdrojem (schopností molekul ionizovat se) a vytvářet odpovídající počet iontů v ESI zdroji a pak se v daném počtu „zaostřit“ do jemného paprsku na vstupu do prvního kvadrupolu Q1.
- Posledním shlukem znaků se silnou korelací jsou znaky CP a CE, která se projevila když byl ze souboru dat vyřazen outlier objekt Fenoterolu. Jako velmi zajímavé se jeví porovnání grafů komponentních vah PC1 a PC2 pro všechny objekty, kdy nebyl vyřazen Fenoterol, tady už spolu znaky CP a CE nekorelují a naopak se dostaly do vzájemné ortogonální polohy. Znaky CP a CE přímo souvisí s kolizní celou (označovanou v detektoru jako Q2) a se stabilitou molekul měřených analytů a jejich schopností se štěpit na „dceřiné“ produkty.
- Toto porovnání ukazuje na to, že optimalizujeme-li podmínky v kolizní cele Q2 detektoru pro skupinu analytů, které mají podobnou molekulu a mají stejný nebo velmi podobný mechanismus štěpení, pak znaky CP a CE spolu pozitivně korelují a naopak budou-li se optimalizovat podmínky v kolizní cele pro různé skupiny analytů (různé skupiny molekul) nebo budou-li mít analyty odlišný mechanismus štěpení, pak znaky CP a CE spolu nebudou korelovat, jak je vidět z grafu pro všechny objekty jsou v ortogonální poloze. Tento dílčí závěr je možné rovněž experimentálně pozorovat při hledání optimálních podmínek v kolizní cele Q2 v MS/MS detektoru.
- Rozptylový diagram komponentního skóre PC1 a PC2 pro všechny objekty ukazuje, že se významně oddělil jeden objekt – Fenoterol a ostatní objekty vytvořily jeden „mrak“, který je „protažený“ vertikálně, tedy více rozdělený podél osy PC2. V případě, že byl objekt Fenoterol ze skupiny dat analytů odstraněn tyto ostatní objekty vytvoří rovnoměrný mrak rozprostřený rovnoměrně podél obou os PC1 a PC2. Tato skutečnost jasně odpovídá na otázku, že bude možné nastavit jednotné podmínky všech parametrů MS/MS detektoru, které budou optimálně vyhovovat všem analytům ze skupiny Beta agonistů, kromě jediného analytu Fenoterolu.

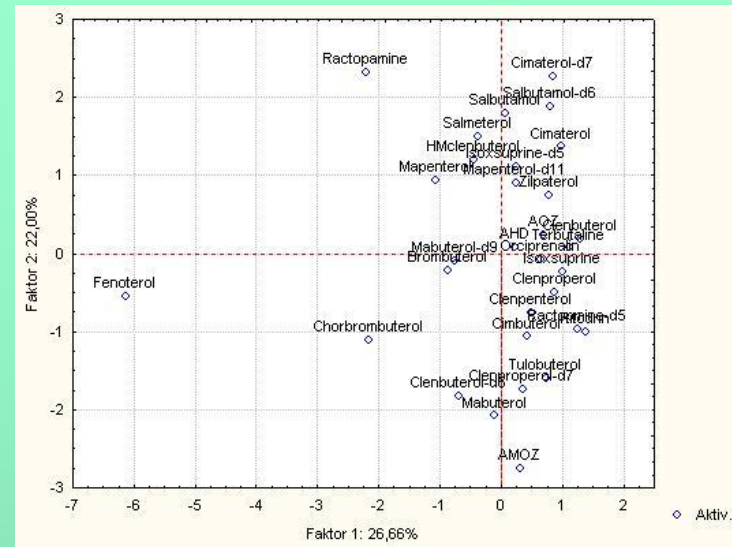
4.3 Porovnání grafických výstupů z PCA a z Shlukové analýzy CLU



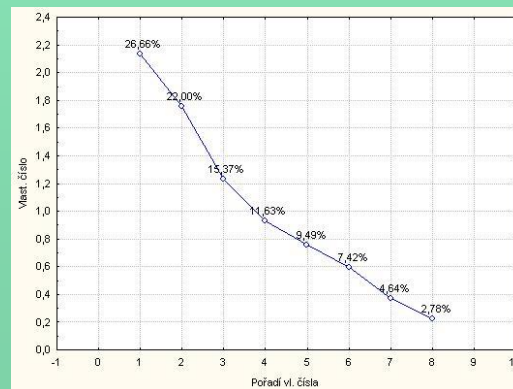
4.4 PCA pro Beta-agonisty a Nitrofurany



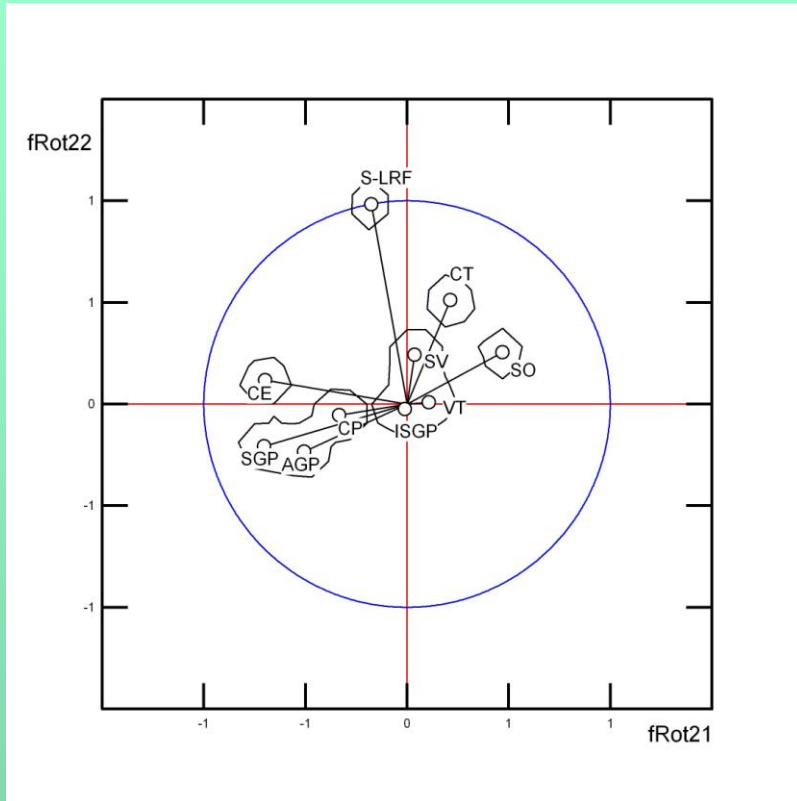
Obr 2.1 Biplot pro Beta-agonisty a Nitrofurany (QC.Extert)



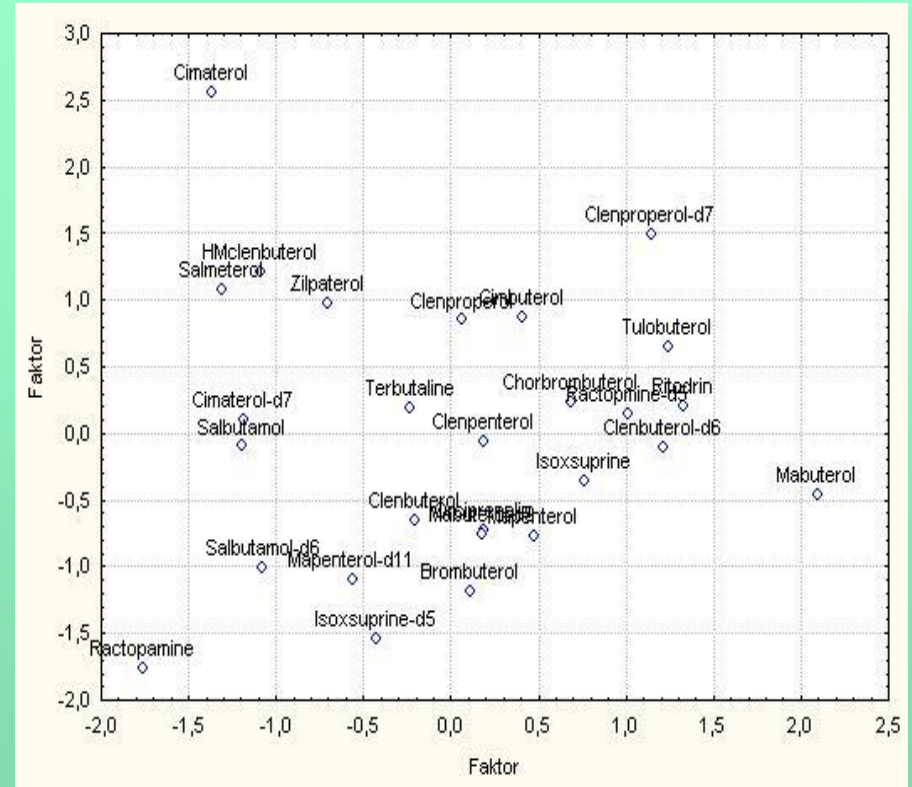
Obr 2.2 Rozptylový diagram komponentního skóre PC1 a PC2 (STATISTICA)



4.5 Faktorová analýza, FA



Obr 2.3 Graf faktorových zátěží F1 a F2, po rotaci (OPstat)



Obr 2.4 Graf faktorového skóre jednotlivých objektů (STATISTICA)

5. Závěr

- Použité statistické metody PCA, FA i CLU prokázaly, že se dají nastavit jednotné parametry MS/MS detektoru TSQ Vantage, které budou optimální pro všechny analyty ze skupiny Beta-agonistů kromě jediného analytu Fenoterolu. Tento závěr potvrzuje i experimentální měření, kde nastavení optimálních podmínek a měření všech analytů Beta-agonistů včetně Fenoterolu v jednom nástřiku jde na úkor citlivosti pro Fenoterol. Všechny analyty mají Limit detekce (LOD) v rozmezí 0,1 – 0,5 µg/l v moči jen Fenoterol má LOD 1,5 µg/l v moči.
- Obecně se dá říci, že statistická metoda PCA nabízí velký potenciál pro řešení problematiky optimalizace parametrů (znaků) mass detektorů MS/MS pro rozdílné skupiny analytů, pro tzv. “multi-reziduální metody”.

Otázka č.1

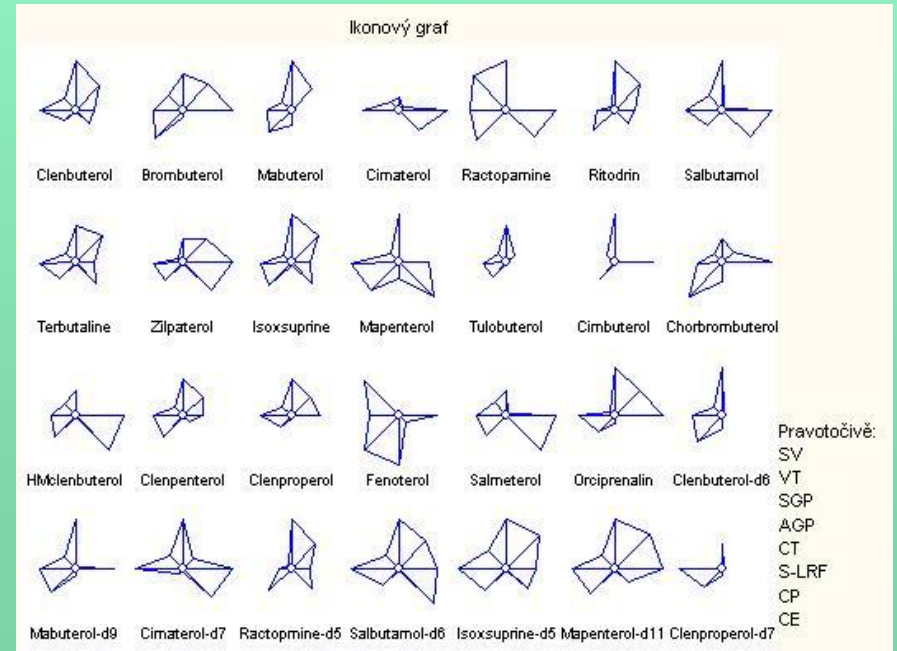
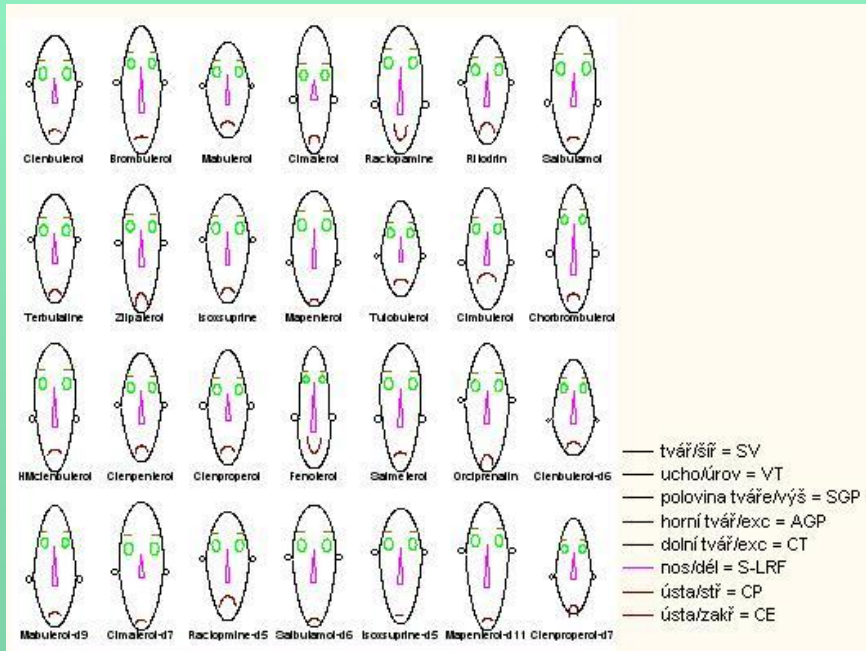
Pro zobrazení vícerozměrných dat se používají obecně dvě základní skupiny grafických technik:

- ✓ Symbolové grafy
- ✓ Zobecněné rozptylové grafy

Otázka č.1

Symbolové grafy – jednotlivé znaky jsou „kódovány“ s ohledem na jejich konkrétní hodnoty do určitých geometrických tvarů či symbolů. Vlastnosti dat se pak posuzují s ohledem na vizuální rozdíly mezi obrázky či symboly a vyhodnocují se objekty x_i , které mají stejné (velmi podobné) vlastnosti.

Mezi základní typy patří: profily, polygony, tváře, křivky a stromy.



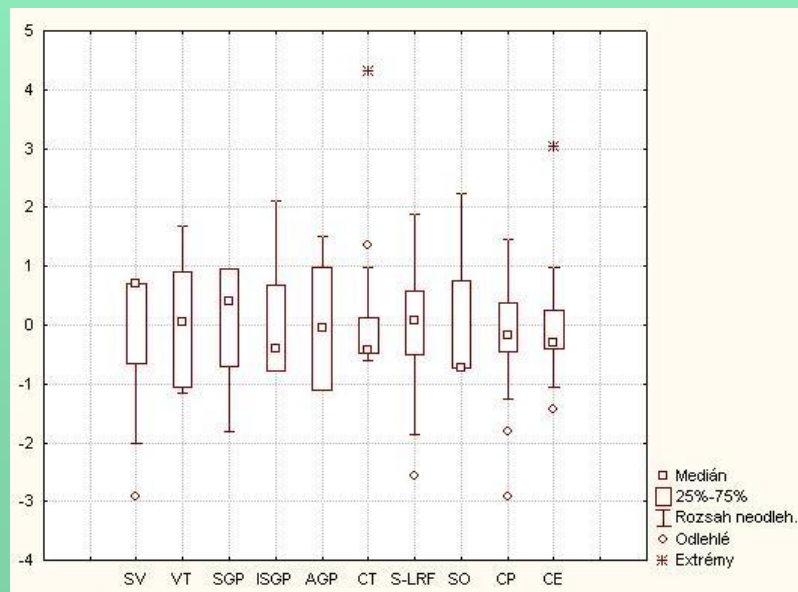
Otázka č.1

Zobecněné rozptylové grafy – charakterizují znaky v jednotlivých objektech a vynášejí se většinou dva znaky x_{i1} a x_{i2} proti sobě vždy na osu x a osu y.

Z těchto párových rozptylových diagramů lze snadno identifikovat:

- Vybočující hodnoty (body),
- Struktury v datech (shluky bodů) – jež charakterizují heterogenitu použitého výběru nebo přítomnost různých dílčích výběrů s odlišným chováním,
- Míru párové závislosti – mezi dvěma znaky objektů.

Pro posouzení variability jednotlivých znaků se ještě využívají krabicové grafy.



Otázka č.2

Standardizace dat – znamená odstranění závislosti na jednotkách a na parametru polohy (popřípadě i rozptýlení). Standardizace tvoří často první krok v předúpravě vícerozměrných dat. Pokud mají znaky rozdílné jednotky (měřítka na osách), musí se standardizace provést vždy.

Nejběžněji se používá Autoškálování tzv.“studentizace“ – jedná se o kombinaci sloupcového centrování a sloupcové standardizace dle vztahu:

$$y_{ij} = (x_{ij} - \bar{x}_j) / s_j$$

Průzkumová analýza vícerozměrných dat EDA – používají se grafické techniky ve 2-rozměrném nebo 3-rozměrném souřadnicovém systému a umožňují:

- Identifikovat vektory \mathbf{x}_i nebo jejich složky, které se jeví jako vybočující,
- Indikovat různé shluky v datech (heterogenita nebo přítomnost různých výběrů),
- Testování a posouzení vícerozměrné normality.

Otázka č.3

Faktorová analýza se pokouší objasnit kovariance (a korelace) původních znaků pomocí lineárních kombinací několika málo společných faktorů a jediného specifického faktoru pro každý znak. Je to analýza struktury vnitřních vztahů mezi velkým počtem původních znaků, pomocí souboru menšího počtu latentních proměnných, zvaných faktory.

Jednotlivé objekty v modelu FA můžeme vyjádřit vztahem:

$$x_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1p}F_p + e_1$$

$$x_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2p}F_p + e_2$$

$$x_m = a_{m1}F_1 + a_{m2}F_2 + \dots + a_{mp}F_p + e_m$$

Kde $F_1 \dots F_p$ jsou společné faktory, které vyvolávají korelace mezi m původními znaky.
 $e_1 \dots e_m$ jsou specifické (chybové) faktory, které přispívají k rozptylu jednotlivých znaků.
 a_{ij} jsou koeficienty, která nazýváme faktorové zátěže i -tého znaku na j -tém faktoru F a jsou to prvky matice faktorových zátěží.

Otázka č.3

V maticovém tvaru pro n – tici objektů a m -tici znaků lze model FA vyjádřit ve tvaru:

$$X = TF^T + E$$

Kde: T je matice latentních proměnných (score matice, matice faktorových skorů),
 F je matice zátěží (matice společných faktorových koeficientů, matice loading),
 E je matice nevysvětlitelné variability (matice reziduí).

Model FA rozdělí každý původní znak na dvě části, tak i rozptyl rozdělí na dvě části:

$$s_i^2 = h_i^2 + u_i^2$$

Kde: h_i se nazývá komunalita, u_i se nazývá jedinečnost (zbytková variabilita) chybového členu a může se ještě dále rozdělit na specifitu a nespolehlivost (chybu měření).

Otázka č.3

Faktorová rotace – hlavním cílem je odvodit z dat snadno vysvětlitelné a pojmenovatelné společné faktory. Počáteční odhady faktorů bývají často obtížně vysvětlitelné, protože většina faktorů je korelována s mnoha znaky. Rotace (otočení) faktorů transformuje faktory do nové „podoby“, která se dá snadněji vysvětlit a přiřadit faktor určitému původnímu znaku, jinými slovy, zátěže většinou u jednoho znaku byly maximalizovány 1 a zbývající zátěže ostatních znaků byly co nejnižší, téměř na 0 Znak „faktorově čistý“.

Rotace je:

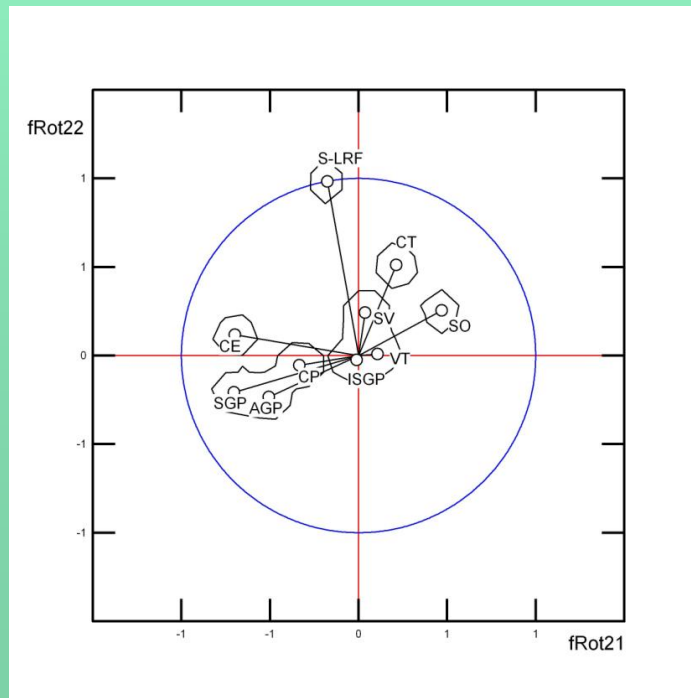
- Ortogonální – osy faktorů F_1 a F_2 se otočí o 90^0 (faktory pak jsou vzájemně nekorelované)
- Neortogonální – otočení o jiný úhel než 90^0 (smysluplnější faktory)

Metody faktorové rotace:

- Metoda varimax – minimalizuje počet znaků, které vykazují vysokou zátěž,
- Metoda quartimax – minimalizuje počet faktorů, potřebných k popisu znaků,
- Metoda equimax – je kombinací předchozích metod.

Otázka č.4

Graf faktorových zátěží – slouží pro posouzení původních znaků a jejich přiřazení k jednotlivým faktorům vhodnou faktorovou rotací. Dá se říci, že shluky znaků, které leží blízko konce souřadnicové osy jednoho z faktorů, mají vysokou zátěž pouze na přiřazené ose (faktoru) mají velmi podobné vlastnosti a označují se jako znaky faktorově čisté. Naopak leží-li shluky znaků mezi osami v jednom ze čtyř kvadrantů, pak jsou to znaky faktorově nečisté (nejasné) a leží-li znaky uprostřed os, mají takové znaky 0 zátěž pro vybrané dva faktory.



Otázka č.5

Graf faktorového skóre – slouží pro vybraný pár faktorů k posouzení a odhalení výjimečných hodnot objektů, především odlehlých pozorování.

