

# KRITIKA METOD FAKTOROVÉ ANALÝZY SPEKTER PŘI STANOVENÍ POČTU SVĚTLO-ABSORBUJÍCÍCH ČÁSTIC

CRITICAL COMPARISON OF FACTOR ANALYSIS METHODS IN SPECTRA ANALYSIS  
DETERMINING THE NUMBER OF LIGHT-ABSORBING SPECIES

Milan Meloun, Petr Mikšík a Karel Kupka  
Katedra analytické chemie,  
Univerzita Pardubice  
532 10 Pardubice, Česká republika

**KEY WORDS:** *factor analysis, spectra analysis, rank of matrix, the number of species, Kankare method, methods of real error, extracted error, mean error and method of logarithm of eigenvalues, instrumental error of spectrophotometer*

## SUMMARY

*Critical comparison of 14 numerical methods in the factor analysis of absorbance matrix was performed on 20 various sets of UV/VIS simulated and experimental spectra sets and leads to conclusion that the Kankare method, methods of real error, extracted error, mean error and method of logarithm of eigenvalues brings true value of the number of light-absorbing species and true value of the instrumental error of spectrophotometer used. Validity of algorithm for comparison of 14 methods was proved.*

## ÚVOD

**1. Rozklad absorbanční matice v metodě hlavních komponent:** výchozí datovou formou pro metodu hlavních komponent (zkráceně PCA) je datová matice, jejíž sloupce představují měřenou vlastnost a řádky příslušné objekty, na nichž je tato vlastnost měřena. V našem případě se jedná o absorbanční matici  $A$  rozměru  $M \times N$ , kde  $M$  řádků představuje spektra chemických směsí složených z určitého počtu chemických individuí, aktivních v studované oblasti spektra a digitalizovaných do  $N$  bodů, čili měřených při  $N$  vlnových délkách. V metodě hlavních komponent, jakožto v metodě s latentními proměnnými, se předpokládá, že variabilitu zdrojové matice lze vyčerpávajícím způsobem popsat pomocí určitého malého počtu latentních (skrytých) proměnných, které v sobě v kondenzované formě obsahují podstatnou část informace (variability) zdrojové matice. Pro absorbanční matici je počet signifikantních latentních proměnných roven počtu světlo-absorbujících částic.

Aplikace metody hlavních komponent spočívá v rozkladu zdrojové *absorbanční matice*  $A$  na součin dvou matic  $T$  a  $P^T$  a *matici nepopsané variability*  $E$  podle vztahu

$$A = T P^T + E$$

Matice  $T$  rozměru  $M \times k$  se nazývá *matice latentních proměnných* a obsahuje  $k$  sloupcových vektorů označovaných jako hlavní komponenty. Matice  $P$  rozměru  $N \times k$  je nazývána *maticí zátěží*. Jednotlivé sloupcové vektory této matice představují míru příspěvku příslušné latentní proměnné k popisu variability sloupců zdrojové matice. Jeden ze způsobů výpočtu matic  $T$  a  $P$  je založen na rozkladu kovarianční matice  $Z$  definované vztahem

$$Z = A^T A$$

Rozklad se provede diagonalizací (řešením vlastního problému) matice  $Z$ . Tímto postupem se získají tzv. *vlastní čísla* (vlastní hodnoty)  $g_a$ , která vyjadřují rozptyl příslušné latentní proměnné a jsou mírou izolované informace (variability), a matice  $Q$  jim příslušejících *vlastních vektorů*  $q_a$ , totožná totiž s maticí  $P$ . Matici  $T$  je potom možno dopočítat podle vztahu

$$T^T = (P^T P)^{-1} P^T A^T$$

2. *Krátký cyklus*: za krátký cyklus se v PCA označuje rekonstrukce zdrojové matice  $A$  s použitím prvních  $k$  latentních proměnných (hlavních komponent), popisujících podstatnou část variability zdrojové matice bez zahrnutí experimentální chyby. Rekonstrukce predikované

matice  $A^{\text{pred}}$  je pak popsána vztahem  $A^{\text{pred}} = T P^T$

3. *Teorie chyb*: v hypotetickém případě, kdy absorbanční matice  $A$  je nezatížená experimentálními chybami, by byla její hodnota rovna přímo počtu latentních proměnných (v chemické interpretaci počtu světlo-absorbujících částic) a souhlasila by s počtem nenulových vlastních čísel matice  $Z$ . *Reálná absorbanční matice*  $A$  je však formálně tvořena součtem *matice čistých dat*  $A^{\text{pure}}$  a *matice experimentálních chyb*  $E$  podle vztahu

$$A = A^{\text{pure}} + E$$

Rozměr všech těchto matic je  $M \times N$ , jejich hodnoty jsou však různé. Hodnota matice čistých dat  $A^{\text{pure}}$  je rovna počtu aktivních složek ve směsi, kdežto hodnoty matic  $A$  a  $E$  jsou rovny menšímu z čísel  $M$  a  $N$ .

Zahrnutí experimentálních chyb do matice  $A$  vede při diagonalizaci její kovarianční matice  $Z$  k většímu počtu nenulových vlastních čísel, než by příslušelo kovarianční matici čistých dat. Použití kompletního souboru hlavních komponent v krátkém cyklu pak povede k plné reprodukci experimentálních chyb v datech. Výběr takového počtu hlavních komponent pro krátký cyklus, který by popisoval prostor čistých dat, může vést k eliminaci určitého podílu chyb v rekonstruovaných datech.

Malinowski<sup>1</sup> vypracoval teorii chyb v PCA a ukázal, jak experimentální chyby vstupují do jejího schématu. Na základě této teorie je možno vlastní čísla  $g_a$  a jim příslušející vlastní vektory  $q_a$  rozdělit na primární a sekundární. *Primární vlastní čísla* jsou největší, přímo souvisí s chemickou informací obsaženou v absorbanční matici  $A$ , ale odrážejí i část nepřesností. *Sekundární vlastní čísla* jsou menší a souvisí pouze s experimentální chybou. Primárním vlastním číslem přísluší *primární vlastní vektory*, které definují bázi faktorového prostoru rozměru  $k$ . Použití prvních  $k$  latentních proměnných vede při krátkém cyklu k optimální rekonstrukci absorbanční matice a eliminuje část experimentálních chyb.

Podle Malinovského vystupují v PCA tři chyby:

- 1) *Reálná chyba (RE)*, která je mírou rozdílu mezi naměřenými (hrubými) daty a ideálními čistými daty.
- 2) *Extrahovaná chyba (XE)*, která je mírou rozdílu mezi hrubými daty a daty získanými rekonstrukcí v krátkém cyklu pomocí daného počtu latentních proměnných. Extrahovaná chyba tedy charakterizuje podíl chyb eliminovaných v krátkém cyklu ze spektrálních dat.
- 3) *Zabudovaná chyba (IE)*, která je mírou rozdílu mezi rekonstruovanými a čistými daty, čili charakterizuje podíl chyb, který zůstává zabudovaný v rekonstruovaných datech.

Pro tyto tři chyby platí vztah  $RE^2 = XE^2 + IE^2$

Z teorie chyb v PCA plyne význam správného určení počtu signifikantních latentních proměnných, neboť podíl eliminovaných experimentálních chyb úzce souvisí s tímto počtem. Přecenění počtu latentních proměnných způsobí zmenšení extrahované chyby a zvětšení zabudované chyby, která potom zatěžuje konečné výsledky. Opačná situace vede k fyzikálně-chemicky nekorektnímu popisu absorbanční matice.

4. *Určení počtu latentních proměnných*: určení počtu latentních proměnných je klíčovým problémem všech metod s latentními proměnnými. V podstatě jde o určení nejmenšího signifikantního počtu latentních proměnných, které popisují variabilitu zdrojové matice bez

zahrnutí experimentální chyby. V případě absorbanční matice je možno tento počet ztotožnit s počtem světlo-absorbujících částic v studovaném systému.

Schopnost všech metod PCA zjistit správný počet aktivních složek ve směsi je podmíněna splněním dvou základních předpokladů:

- 1) Žádné ze spekter čistých složek není lineární kombinací spekter ostatních složek.
- 2) Koncentraci žádné ze složek není možné vyjádřit pomocí konstantního poměru koncentrací jiných složek, čili relativní poměr koncentrací složek v jednotlivých roztocích není konstantní.

Každé nesplnění těchto podmínek snižuje hodnotu absorbanční matice čistých dat  $A^{pure}$  a v konečném důsledku vede k určení menšího počtu světlo-absorbujících částic, než je ve směsi skutečně přítomno.

Metody používané k určení počtu latentních proměnných lze rozdělit do dvou základních skupin:

- 1) Metody založené na odhadu velikosti experimentální chyby v datech.
- 2) Aproximativní metody, nepožadující znalost velikosti experimentální chyby.

Pokud je znám odhad experimentální chyby zatěžující měřená data, upřednostňuje se první skupina metod.

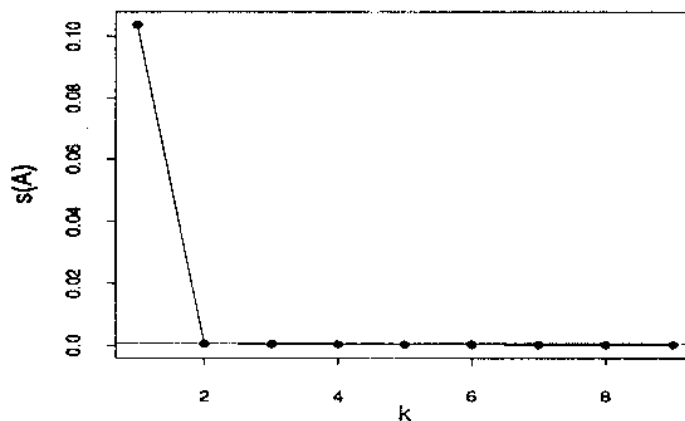
## METODICKÁ ČÁST

### A. Metody, založené na odhadu velikosti experimentální chyby

#### 1. KANKAREHO METODA

Kankareho metoda<sup>2</sup> vychází z druhého momentu  $M$  absorbanční matice  $A$  vypočteného podle vztahu

$$M = \frac{1}{M} A^T A$$



**Obr. 1** Příklad určení počtu světlo-absorbujících částic *Kankareho metodou* ze simulované absorbanční matice (10 roztoků, 31 vlnových délek) acidobazické rovnováhy bromkrezolové zeleně ( $c_{BKZ} = 5.444 - 5.379 \times 10^{-4} \text{ mol.l}^{-1}$ ) se dvěma částicemi L a HL v oblasti  $\text{pH} = 6.49 - 3.49$ . Na data je superponován výběr náhodných chyb o směrodatné odchylce  $s_{\text{svt}}(A) = 0.0006$ . Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose směrodatné odchylce  $s_k(A)$ . Vodorovná linie označuje předpokládanou velikost  $s_{\text{svt}}(A)$ .

**Závěr:** Hodnota  $s_k(A)$  aproximuje nejlépe  $s_{\text{svt}}(A)$  pro  $k = 2$ . Z grafu je zřejmé, že vyšší hodnota  $k$  nepřináší výrazné snížení hodnoty  $s_k(A)$

Pomocí vlastních čísel  $r_a$  matice  $M$  se určí *reziduální směrodatná odchylka absorpance*  $s_k(A)$  podle vztahu

$$s_k(A) = \sqrt{\frac{\text{tr}(M) - \sum_{a=1}^k r_a}{N - k}}$$

kde  $\text{tr}(M)$  je stopa matice  $M$  a  $k$  je určovaný počet aktivních komponent.

Hodnoty  $s_k(A)$  pro různý počet předpokládaných aktivních komponent  $k$  se vynesou graficky jako funkce

$$s_k(A) = f(k)$$

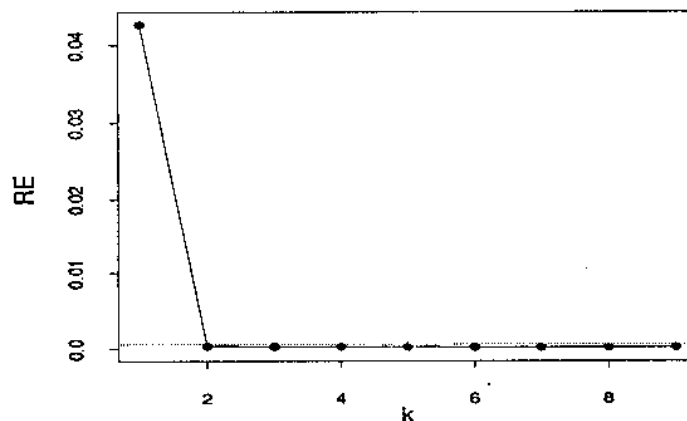
a jako hledaný počet světlo-absorbujících částic ve směsi se přijme ta hodnota  $k$ , při které  $s_k(A)$  nejlépe aproximuje předpokládanou směrodatnou odchylku absorbance  $s_{\text{inst}}(A)$  pro daný spektrofotometr.

## 2. METODA REÁLNÉ CHYBY

Reálná chyba  $RE$  je měřítkem difference mezi skutečnými hrubými daty a ideálními čistými daty a je definována výrazem

$$RE = \sqrt{\frac{\sum_{a=k+1}^M g_a}{N(M-k)}}$$

kde  $g_a$  jsou vlastní čísla kovarianční matice  $Z$  a  $k$  je hledaný počet signifikantních latentních proměnných.



Obr. 2 Příklad určení počtu světlo-absorbujících částic metodou reálné chyby ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose reálné chybě  $RE$ . Vodorovná linie označuje velikost experimentální chyby  $s_{\text{inst}}(A)$ .

Závěr: Hodnota  $RE$  aproximuje nejlépe experimentální chybu pro  $k = 2$ . Z grafu je zřejmé, že vyšší hodnota  $k$  nepřináší výrazné snížení hodnoty  $RE$ .

Hodnoty  $RE$  se mohou opět vynést graficky jako funkce počtu latentních proměnných

$$RE = f(k)$$

a srovnat s velikostí odhadu experimentální chyby pro daný spektrofotometr. Hodnotu  $k$ , při které  $RE$  experimentální chybu uspokojivě reprodukuje, je možno ztotožnit s počtem světlo-absorbujících komponent ve směsi.

## 3. METODA EXTRAHOVANÉ CHYBY

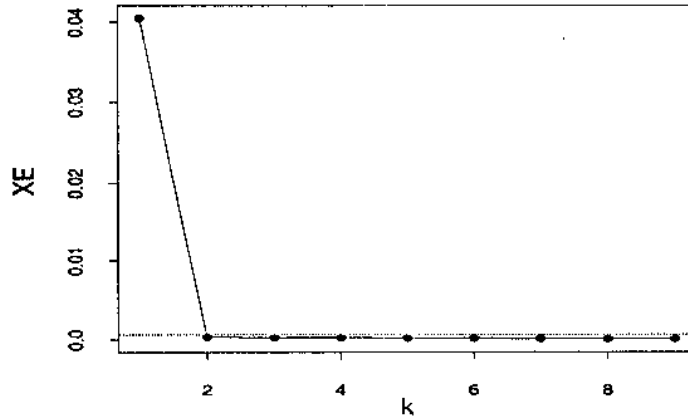
Extrahovaná chyba  $XE$  jako měřítko difference mezi hrubými daty a daty získanými rekonstrukcí v krátkém cyklu pomocí prvních  $k$  hlavních komponent je definována vztahem

$$XE = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^M (A_{ij} - \hat{A}_{ij})^2}{NM}}$$

kde  $A_{ij}$  je příslušný prvek absorbanční matice  $A$  a  $\hat{A}_{ij}$  je tento prvek rekonstruovaný pomocí prvních  $k$  latentních proměnných podle vzorce

$$\hat{A}_{ij} = \sum_{a=1}^k t_{ia} p_{aj}$$

kde  $t_{ia}$  je prvek matice latentních proměnných  $T$  a  $p_{aj}$  je prvek matice zátěží  $P$ .



**Obr. 3** Příklad určení počtu světlo-absorbujících částic *metodou extrahované chyby* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose reálné chybě  $XE$ . Vodorovná linie označuje předpokládanou velikost experimentální chyby  $s_{\text{inst}}(A)$ .

**Závěr:** Hodnota  $XE$  aproximuje nejlépe experimentální chybu pro  $k = 2$ . Z grafu je zřejmé, že vyšší hodnota  $k$  nepřináší výrazné snížení hodnoty  $XE$

Alternativní způsob výpočtu  $XE$  představuje vztah

$$XE = \sqrt{\frac{\sum_{a=k+1}^M g_a}{NM}}$$

kde  $g_a$  jsou vlastní čísla kovarianční matice  $Z$  a  $k$  je hledaný počet signifikantních latentních proměnných. Obdobně jako v předešlé metodě lze hodnoty  $XE$  vynést graficky jako funkci počtu latentních proměnných  $XE = f(k)$  a srovnat s velikostí odhadu experimentální chyby pro daný spektrofotometr. Hodnotu  $k$ , při které  $XE$  experimentální chybu uspokojivě reprodukuje, je možno ztotožnit s počtem světlo-absorbujících částic ve směsi. Srovnáním výrazů pro výpočet  $RE$  a  $XE$  obdržíme vztah

$$XE = \sqrt{\frac{M-N}{M}} RE$$

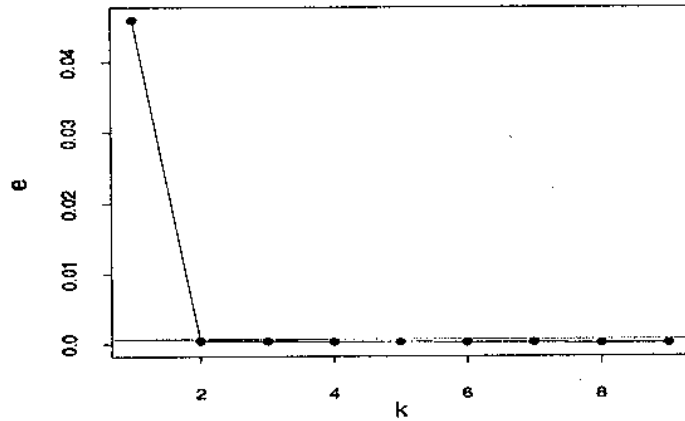
Přestože jsou tyto chyby vzájemně spjaty, jsou tvořeny na rozdílných základech. Pro určení signifikantního počtu latentních proměnných bývá preferována metoda reálné chyby.

#### 4. METODA PRŮMĚRNÉ CHYBY

**Průměrná chyba** absorbance  $\bar{\epsilon}$  je definována jako aritmetický průměr z absolutních hodnot diferencí mezi naměřenými daty a daty rekonstruovanými v krátkém cyklu pomocí  $k$  prvních latentních proměnných vztahem

$$\bar{e} = \frac{\sum_{i=1}^N \sum_{j=1}^M |A_{ij} - \hat{A}_{ij}|}{NM}$$

kde význam symbolů  $A_{ij}$  a  $\hat{A}_{ij}$  je stejný jako při definici  $XE$ .



**Obr. 4** Příklad určení počtu světlo-absorbujících částic *metodou průměrné chyby* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose reálné chybě  $\bar{e}$ . Vodorovná linie označuje předpokládanou velikost experimentální chyby  $\bar{e}_{inst} = s_{inst}(A)$ .

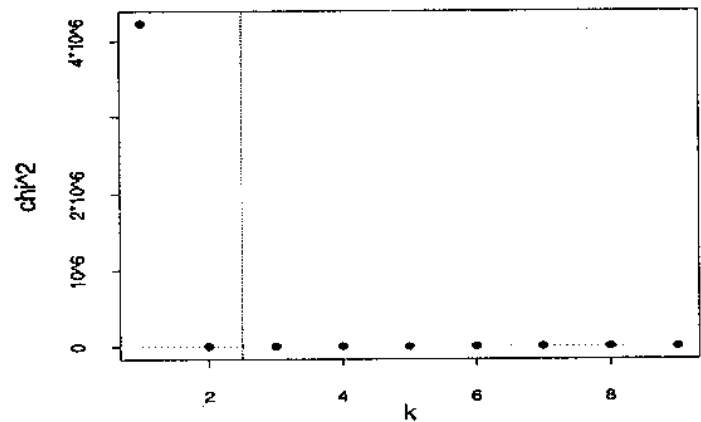
**Závěr:** Hodnota  $\bar{e}$  aproximuje nejlépe experimentální chybu pro  $k = 2$ . Z grafu je zřejmé, že vyšší hodnota  $k$  nepřináší výrazné snížení hodnoty  $\bar{e}$

Hodnoty průměrné chyby se vynesou do grafu proti počtu latentních proměnných  $k$  jako funkce

$$\bar{e} = f(k)$$

a srovnají se s hodnotou odhadu průměrné chyby  $\bar{e}_{inst}$  pro daný spektrofotometr. Jako počet světlo-absorbujících částic ve směsi se označí ta hodnota  $k$ , při které příslušná hodnota  $\bar{e}$  nejlépe aproximuje hodnotu  $\bar{e}_{inst}$ .

### 5. METODA $\chi^2$ KRITÉRIA



**Obr. 5** Příklad určení počtu světlo-absorbujících částic *metodou  $\chi^2$  kritéria* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose testovacímu kritériu  $\chi^2$ , čerchovaná čára označuje velikost  $\chi^2_{krit}$  a svislá linie odděluje hodnoty  $k$ , pro něž byla  $H_0$  přijata.

**Závěr:** Nejvyšší  $k$  splňující  $H_0$  je  $k = 2$

Schopnost matice  $A^{\text{pred}}$  aproximovat skutečnou matici  $A$  závisí na počtu latentních proměnných  $k$  použitých k její rekonstrukci v krátkém cyklu. S rostoucím  $k$  se stupeň aproximace zvyšuje. První matici  $A^{\text{pred}}$ , která reprodukuje matici  $A$  v rámci experimentální chyby, je možno považovat za její nejlepší aproximaci a příslušnou hodnotu  $k$  ztotožnit s počtem světlo-absorbujících částic ve směsi.

Na posouzení, zda matice  $A^{\text{pred}}$  reprodukuje matici  $A$  v rámci experimentální chyby, byl navržen  $\chi^2$  test<sup>3,4</sup>. Hodnota  $\chi^2$  kritéria je definována vztahem

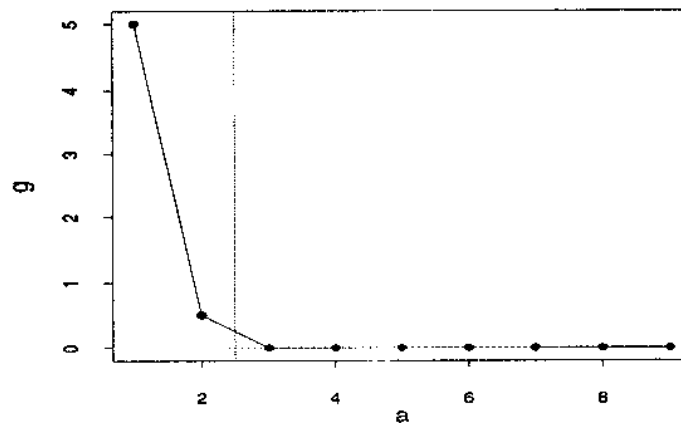
$$\chi^2 = \sum_{i=1}^N \sum_{j=1}^M \left( \frac{A_{ij} - \hat{A}_{ij}}{\sigma_{ij}} \right)^2$$

kde  $\sigma_{ij}$  je odhad směrodatné odchytky příslušného prvku absorbanční matice  $A_{ij}$  a  $\hat{A}_{ij}$  je odpovídající prvek matice  $A^{\text{pred}}$ . Hodnota  $\chi^2$  pro  $k = 1, 2, 3, \dots, M$  se porovnává s kritickou hodnotou

$$\chi^2_{\text{krit}} = (N - k)(M - k)$$

Počet světlo-absorbujících částic ve směsi odpovídá první hodnotě  $k$ , pro kterou je hodnota  $\chi^2$  menší než kritická hodnota.

## 6. METODA SMĚRODATNÝCH ODCHYLEK VLASTNÍCH ČÍSEL



Obr. 6 Příklad určení počtu světlo-absorbujících částic *metodou směrodatných odchylek vlastních čísel* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose vlastním číslům  $g_a$ , čerchovaná linie označuje velikost směrodatných odchylek vlastních čísel  $\sigma_{g_a}$  a svislá linie odděluje ty  $g_a$ , které jsou větší než odpovídající směrodatná odchylna.

**Závěr:** První dvě vlastní čísla jsou větší než jejich směrodatné odchylky

Hugus a El-Awady<sup>5</sup> odvodili pro směrodatnou odchylnu vlastního čísla  $g_a$  kovarianční matice  $Z$  výraz

$$\sigma_{g_a} = \sqrt{\sum_{l=1}^M \sum_{m=1}^M q_{la}^2 q_{ma}^2 \sigma_{lm}^2}$$

kde  $q_{la}$  a  $q_{ma}$  jsou prvky matice vlastních vektorů  $Q$  a  $\sigma_{lm}$  jsou směrodatné odchylky prvků matice  $Z$  dané výrazy

$$\sigma_{lm}^2 = \sum_{i=1}^N (A_{il}^2 \sigma_{im}^2 + A_{im}^2 \sigma_{il}^2)$$

pro  $l \neq m$  a

$$\sigma_{ll}^2 = 4 \sum_{i=1}^N (A_{il}^2 \sigma_{il}^2)$$

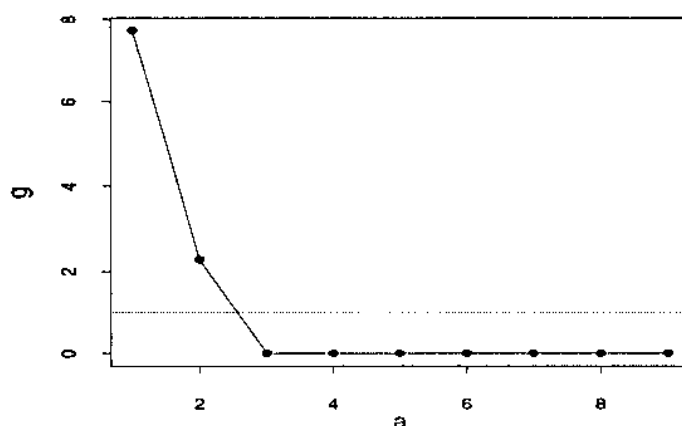
pro  $l = m$ , kde  $\sigma_{ii}$  a  $\sigma_{im}$  jsou odhady směrodatných odchylek příslušných prvků  $A_{ii}$  a  $A_{ij}$  absorbanční matice  $A$ .

Podle metody odhadu směrodatných odchylek vlastních čísel<sup>4</sup> je počet světlo-absorbujících komponent ve směsi roven počtu vlastních čísel  $g_a$ , která jsou větší než jejich směrodatné odchylky  $\sigma_{g_a}$ .

## B. Metody, nepožadující informaci o experimentální chybě

### 7. METODA PRŮMĚRU VLASTNÍCH ČÍSEL

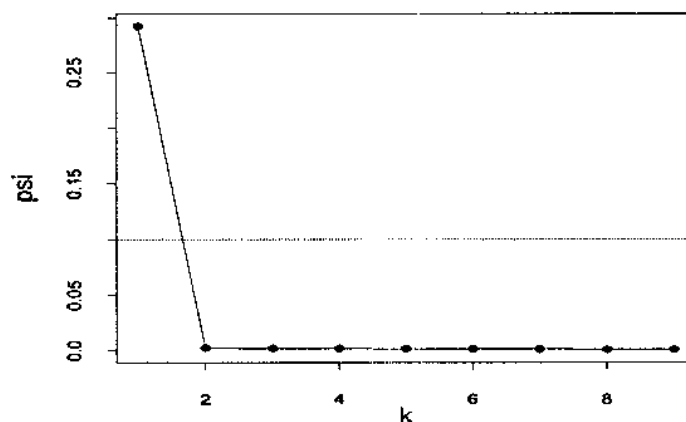
Metoda průměru vlastních čísel<sup>3</sup> označuje jako signifikantní pouze ty latentní proměnné, pro které příslušná vlastní čísla  $g_a$  kovarianční matice  $Z$  převyšují velikost jejich aritmetického průměru. Pokud jsou vlastní čísla počítána z korelační matice, porovnává se jejich velikost s jedničkou. Pro tento případ se tato metoda označuje také jako „eigenvalue-one“ kritérium. Počet signifikantních latentních proměnných je roven počtu světlo-absorbujících částic ve směsi.



Obr. 7 Příklad určení počtu světlo-absorbujících částic *metodou průměru vlastních čísel* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose vlastní číslům  $g_a$ , vypočteným z korelační matice. Vodorovná linie odděluje ty  $g_a$ , která jsou větší než jejich aritmetický průměr (pro korelační matici roven jedné).

Závěr: První dvě vlastní čísla jsou větší než aritmetický průměr vlastních čísel

### 8. METODA EXNEROVY FUNKCE



Obr. 8 Příklad určení počtu světlo-absorbujících částic *metodou Exnerovy funkce* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose hodnotám Exnerovy funkce  $\psi$ . Vodorovná linie označuje hodnotu  $\psi = 0.1$ , odpovídající excellentnímu proložení.

Závěr: Hodnoty  $\psi \leq 0.1$  je dosaženo pro  $k = 2$ . Vyšší hodnota  $k$  nepřináší výrazné snížení hodnoty  $\psi$



Exnerova  $\psi$  funkce<sup>3</sup> je definována výrazem

$$\psi = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^M (A_{ij} - \hat{A}_{ij})^2}{\sum_{i=1}^N \sum_{j=1}^M (A_{ij} - \bar{A})^2}} \cdot \frac{NM}{(NM) - k}$$

kde  $\bar{A}$  představuje aritmetický průměr prvků absorbanční matice  $A$  a  $\hat{A}_{ij}$  prvek matice  $A^{\text{pred}}$ , získané v krátkém cyklu pomocí  $k$  prvních hlavních komponent. Exnerova funkce může obecně nabývat hodnot od 0 do  $\infty$  a se zlepšující se aproximací matice  $A$  maticí  $A^{\text{pred}}$  se blíží k nule. Hodnota  $\psi = 1.0$  představuje horní hranici pro významnost rekonstrukce dat a odpovídá tvrzení, že data rekonstruovaná pomocí prvních  $k$  latentních proměnných jsou stejně dobrou aproximací matice  $A$  jako matice, jejíž každý prvek je roven  $\bar{A}$ . Se snižujícím se  $\psi$  se významnost rekonstrukce dat zvyšuje. Pro excelentní korelaci mezi původními a rekonstruovanými daty navrhuje Exner  $\psi = 0.1$ . Odpovídající hodnotu  $k$ , kdy  $\psi \leq 0.1$ , je možno brát jako počet světlo-absorbujících komponent ve směsi.

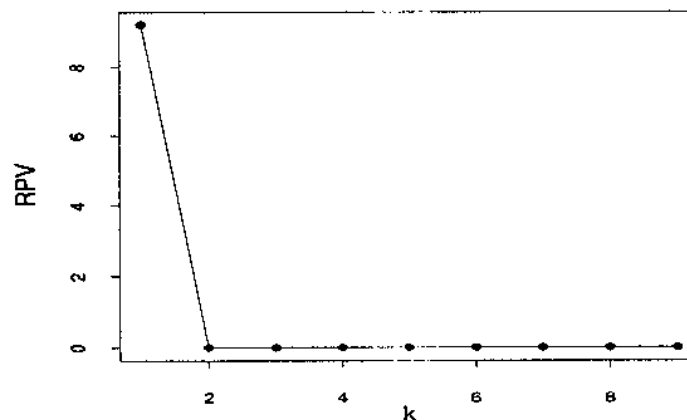
### 9. Metoda reziduální variance

Reziduální variance<sup>3</sup>  $RV$  asociovaná s rekonstruovanými daty je definována vztahem

$$RV = \frac{\sum_{i=1}^N \sum_{j=1}^M (A_{ij} - \hat{A}_{ij})^2}{NM}$$

a je tedy rovna čtverci extrahované chyby. Často se vyjadřuje v procentech jako *reziduální procentuální variance*  $RPV$  podle vzorce

$$RPV = 100 \left( \frac{\sum_{i=1}^N \sum_{j=1}^M (A_{ij} - \hat{A}_{ij})^2}{\sum_{i=1}^N \sum_{j=1}^M A_{ij}^2} \right)$$



Obr. 9 Příklad určení počtu světlo-absorbujících částic *metodou reziduální variance* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose hodnotám reziduální procentuální variance  $RPV$ .

Závěr: Směrnice závislosti  $RPV = f(k)$  je prakticky nulová od  $k = 2$

Pomocí vlastních čísel  $g_s$  kovarianční matice  $Z$  je možno  $RPV$  vyjádřit vztahem

$$RPV = 100 \left( \frac{\sum_{a=k+1}^M g_a}{\sum_{a=1}^M g_a} \right)$$

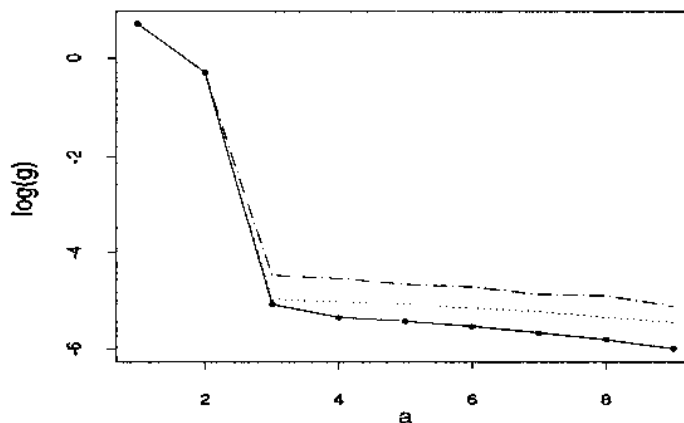
Pro určení počtu latentních proměnných se graficky vynese závislost  $RPV$  na počtu latentních proměnných  $k$  a za signifikantní počet latentních proměnných, odpovídající počtu světlo-absorbujících částic, se přijme ta hodnota  $k$ , při které směrnice funkce

$$RPV = f(k)$$

prudce klesá k nule. Od tohoto bodu se křivka jakoby „vyrovnává“ a hodnota  $RPV$  se dále významně nemění.

### 10. METODA LOGARITMŮ VLASTNÍCH ČÍSEL

Metoda logaritmů vlastních čísel<sup>3</sup> vychází z předpokladu, že primární vlastní čísla kovarianční matice  $Z$  se svou velikostí významně odlišují od sekundárních vlastních čísel, jejichž velikost je přibližně stejná.



Obr. 10 Příklad určení počtu světlo-absorbujících částic *metodou logaritmů vlastních čísel* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose hodnotám  $\log(g_a)$ . Čerchované čáry označují velikost vlastních čísel pro danou absorbanční matici s dalším superponovaným výběrem náhodných chyb o směrodatné odchylce  $2 \times s_{\text{int}}(A)$  a  $4 \times s_{\text{int}}(A)$ .

**Závěr:** Pouze pro  $a = 1$  a  $2$  nezávisí hodnota  $\log(g_a)$  na velikosti superponovaného šumu. Tyto hodnoty  $\log(g_a)$  jsou také výrazně větší než ostatní

Graficky se tedy dají primární a sekundární vlastní čísla odlišit vynesáním závislosti

$$\log(g_a) = f(a)$$

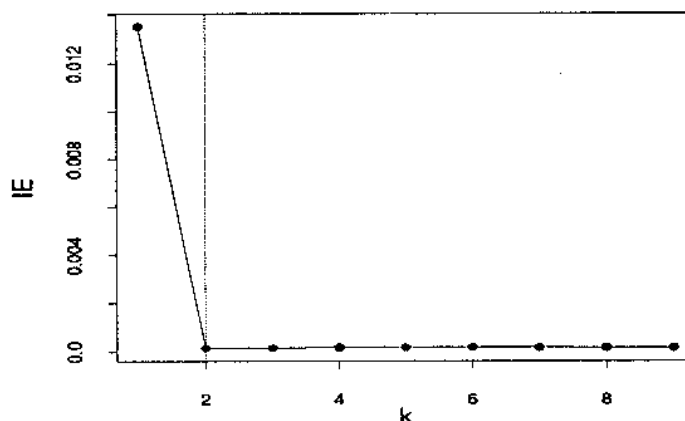
kde  $a$  je pořadí daného vlastního čísla v jejich sestupně uspořádaném souboru. Tento test však není dostatečně citlivý na přítomnost světlo-absorbujících složek v relativně malých množstvích vzhledem k ostatním složkám, kdy rozlišení primárních a sekundárních vlastních čísel podle hodnot jejich logaritmů nemusí být jednoznačné. Je proto výhodné doplnit tuto metodu o poznatek<sup>8</sup>, že velikost primárních vlastních čísel v podstatě nezávisí na úrovni šumu, kterým jsou zatížena spektrální data, kdežto velikost sekundárních vlastních čísel se s rostoucí hladinou šumu zvyšuje. Když tedy do jednoho grafu znázorníme výše uvedenou funkci pro tutéž absorbanční matici, ale s různou úrovní experimentální chyby (získanou např. superponováním syntetického šumu různé úrovně na původní matici  $A$ ), bude rozlišení primárních a sekundárních vlastních čísel výraznější. Počet primárních vlastních čísel je pak možno ztotožnit s počtem světlo-absorbujících částic ve směsi.

## 11. METODA ZABUDOVANÉ CHYBY

Zabudovaná chyba<sup>3</sup>  $IE$  definovaná výrazem

$$IE = \sqrt{\frac{k \sum_{a=k+1}^M g_a}{NM(M-k)}}$$

je měřítkem difference mezi rekonstruovanými a čistými daty a charakterizuje tedy podíl chyb, který zůstává zabudovaný v rekonstruovaných datech.



**Obr. 11** Příklad určení počtu světlo-absorbujících částic *metodou zabudované chyby* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose zabudované chybě  $IE$ . Svislá linie označuje hodnotu  $k$ , pro kterou funkce  $IE = f(k)$  dosahuje minima.

**Závěr:** Funkce  $IE = f(k)$  dosahuje minima pro  $k = 2$

Tento podíl se snižuje, pokud se počet latentních proměnných  $k$  pro rekonstrukci dat v krátkém cyklu přibližuje skutečnému počtu světlo-absorbujících částic ve směsi. V bodě, kdy  $k$  je právě rovno počtu světlo-absorbujících komponent, vykazuje funkce

$$IE = f(k)$$

minimum.

V reálných situacích je však často funkce v okolí minima velmi plytká nebo minima vůbec nenabývá. Tato situace nastává zejména když rozdělení experimentálních chyb není náhodné a rovnoměrné, data jsou zatížena systematickou chybou a nebo se v nich vyskytují hrubé chyby.

## 12. METODA FAKTOROVÉ INDIKÁTOROVÉ FUNKCE

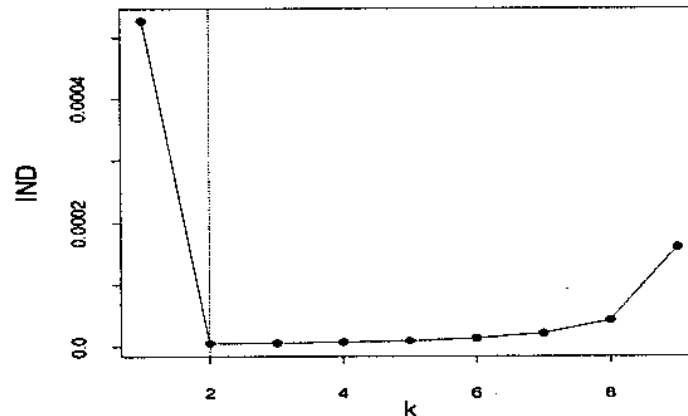
Faktorová indikátorová funkce<sup>3</sup>  $IND$  daná vztahem

$$IND = \sqrt{\frac{\sum_{a=k+1}^M g_a}{N(M-k)}} = \frac{RE}{(M-k)^2}$$

je empirická funkce, jejíž vlastnosti jsou pro určení počtu signifikantních latentních proměnných lepší než u  $IE$ .

$IND$  nabývá minima pro  $k$  rovno počtu světlo-absorbujících částic ve směsi, a to často i tehdy, když  $IE$  toto minimum nevykazuje.

Hrubé chyby v absorbanční matici  $A$  však mohou způsobit vznik sekundárního lokálního minima nebo rozšíření funkce v okolí minima na plytké plató, takže určení minima je nejednoznačné.



Obr. 12 Příklad určení počtu světlo-absorbujících částic *metodou faktorové indikátorové funkce* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose hodnotám  $IND$ . Svislá linie označuje hodnotu  $k$ , pro kterou funkce  $IND = f(k)$  dosahuje minima.

Závěr: Funkce  $IND = f(k)$  dosahuje minima pro  $k = 2$

### 13. MALINOWSKÉHO F-TEST

Malinowski<sup>6, 7</sup> navrhl pravděpodobnostní distribuční funkci  $P$  sekundárních vlastních čísel kovarianční matice  $Z$ , která přísluší datům zatíženým rovnoměrně rozdělenou chybou, ve tvaru

$$P = B (N - a + 1) (M - a + 1)$$

kde  $B$  je normovací konstanta a  $a$  index vlastního čísla  $g_a$ . Tzv. redukováná vlastní čísla jsou definována vztahem

$$g_a^{\text{red}} = \frac{g_a}{(N - a + 1) (M - a + 1)}$$

Rozdíly ve velikosti sekundárních  $g_a^{\text{red}}$  jsou statisticky málo významné, oproti tomu primární redukováná vlastní čísla jsou významně větší, neboť kromě příspěvku experimentální chyby obsahují i příspěvky reálných složek a jejich rozdělení neodpovídá uvedené distribuční funkci.

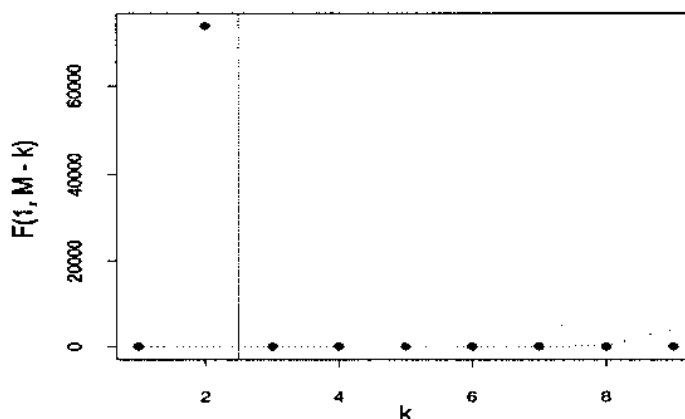
Pro rozlišení primárních a sekundárních redukovanych vlastních čísel navrhl Malinowski<sup>6,7</sup> testovat nulovou hypotézu  $H_0 : g_a^{\text{red}} = \bar{g}_a^{\text{red}}$  oproti alternativní  $H_A : g_a^{\text{red}} > \bar{g}_a^{\text{red}}$ , kde  $\bar{g}_a^{\text{red}}$  je vážený průměr sekundárních  $g_a^{\text{red}}$ .

V případě platnosti  $H_0$  má testovací kritérium

$$F(1, M - k) = \frac{\sum_{a=k+1}^M (N - a + 1) (M - a + 1)}{(N - k + 1) (M - k + 1)} \cdot \frac{g_a}{\sum_{a=k+1}^M g_a}$$

$F$ -rozdělení s 1 a  $M - k$  stupni volnosti. Při testování se postupuje od nejmenších vlastních čísel, čili  $k$  se mění v pořadí  $M - 1, M - 2, \dots, 1$ . První  $k$ -té redukováné vlastní číslo, pro které je  $F(1, M - k)$  větší než tabelovaná hodnota pro zvolenou hladinu významnosti, je

považováno za nejmenší primární a hodnota  $k$  pak určuje počet světlo-absorbujících částic ve směsi.

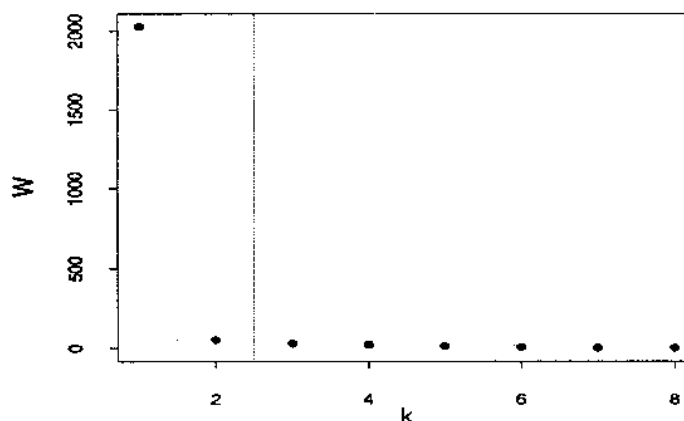


Obr. 13 Příklad určení počtu světlo-absorbujících částic *Malinowského F-testem* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose testovacímu kritériu  $F(1, M - k)$ , čerchovaná čára označuje velikost odpovídající tabelované hodnoty pro hladinu významnosti 0.99 a svislá linie odděluje nejvyšší hodnotu  $k$ , pro kterou byla  $H_0$  přijata. Závěr: Nejvyšší  $k$  splňující  $H_0$  je  $k = 2$

#### 14. BARTLETTŮV TEST IZOTROPIE

Bartlettův test<sup>3</sup> izotropie testuje hypotézu o rovnosti  $M - k$  nejmenších vlastních čísel kovařianční matice  $Z$  ve tvaru

$$H_0 : g_{k+1} = g_{k+2} = \dots = g_M$$



Obr. 14 Příklad určení počtu světlo-absorbujících částic *Bartlettovým testem izotropie* ze simulované absorbanční matice z obr. 1. Hodnoty na x-ové ose odpovídají počtu světlo-absorbujících částic ve směsi, hodnoty na y-ové ose testovacímu kritériu  $W$ , čerchovaná čára označuje velikost odpovídajícího  $\chi^2$ -kvantilu pro hladinu významnosti 0.99 a svislá linie odděluje hodnoty  $k$ , pro něž nebyla  $H$  akceptována. Závěr: Nejnižší  $k$ , pro které  $H$  není ještě splněna, je  $k = 2$

Testovací kritérium  $W$  má tvar

$$W = \left( N - \frac{2M+11}{6} \right) (M-k) \ln \left( \frac{\bar{g}_k^a}{\bar{g}_k^g} \right)$$

kde  $\bar{g}_k^a$  je aritmetický a  $\bar{g}_k^g$  geometrický průměr  $M - k$  nejmenších vlastních čísel matice  $Z$ . Za předpokladu platnosti  $H_0$  má  $W$  přibližně  $\chi^2$  rozdělení s  $0.5 (M - k + 2) (M - k - 1)$  stupni volnosti.

Při testování se postupuje v řadě  $k = 0, 1, 2, \dots$  až po první hodnotu  $k$ , která je menší než tabelovaná hodnota  $\chi^2$  pro zvolenou hladinu významnosti. Od této hodnoty je hypotéza  $H_0$  akceptována. Počet světlo-absorbujících částic ve směsi pak bude roven  $k - 1$ , čili počtu vlastních čísel, pro něž  $H_0$  akceptována nebyla.

## DISKUSE

### Odhad instrumentální chyby spektrofotometru

Pro stanovení odhadu instrumentální chyby používaného spektrofotometru byla užitá metoda navržená Wernimontem<sup>9</sup>, spočívající v analýze spekter koncentrační závislosti spektrálního standardu. Spektrální standard poskytuje v roztoku jedinou světlo-absorbující částici a je tedy možné vyčíslit reziduální směrodatnou odchylku absorpance  $s_k(A)$  pro  $k = 1$  výše popsanou Kankareho metodou a položit ji rovnu instrumentální chybě přístroje  $s_{inst}(A)$ , tzv. *přístrojové chybě*. Stejně tak je možné vyčíslit odhad *experimentální reálné chyby RE*, *extrahované chyby XE* a *průměrné chyby  $\bar{e}$*  absorpance pro daný přístroj dle vztahů pro  $k = 1$ .

Tabulka 1 Stanovení instrumentální chyby spektrofotometru,  $s_{inst}(A)$ , tzv. přístrojové chyby

Spektrální standard	$K_2CrO_4$		$CoSO_4 \cdot 7H_2O$		$CuSO_4 \cdot 5H_2O$	
	1.	2.	1.	2.	1.	2.
Kank, $k = 1$	0.00070	0.00069	0.00038	0.00031	0.00036	0.00024
RE, $k = 1$	0.00048	0.00040	0.00018	0.00015	0.00022	0.00018
XE, $k = 1$	0.00047	0.00038	0.00018	0.00015	0.00021	0.00017
$\bar{e}$ , $k = 1$	0.00024	0.00021	0.00019	0.00016	0.00018	0.00015

Protože vlastní experimenty byly měřeny v oblasti  $\lambda = 380 - 650$  nm, byly za odhad experimentální chyby, používaný v metodách Kankareho (Kank), reálné chyby RE, extrahované chyby XE, průměrné chyby ( $\bar{e}$ ),  $\chi^2$  kriteria a v metodě směrodatných odchylek vlastních čísel, brány nejpesimističtější hodnoty, vycházející pro chroman draselný, tj.  $s_l(A) = 0.00070$ ,  $RE = 0.00048$ ,  $XE = 0.00047$  a  $\bar{e} = 0.00024$ .

### Kritika algoritmu hledání počtu světlo-absorbujících částic

Algoritmus hledání počtu světlo-absorbujících částic faktorovou analýzou absorbanční matice  $A$  byl ověřen na čtvero rozličných souborech dat: šestnácti souborech spekter acidobazických rovnováh sulfonftaleinů bromkrezolové zeleně, fenolové červeně a tymolové modři, čtyřech souborech spektrálních standardů dichromanu draselného, síranu kobaltnatého a síranu měďnatého. Grafická prezentace výsledků čtrnácti metod faktorové analýzy pro všechny uvedené soubory následuje.

#### 1. Spektra acidobazických rovnováh sulfonftaleinů (Tabulka 1)

Pro jednoduché protonační rovnováhy sulfonftaleinů  $HL \rightleftharpoons H^+ + L^-$  byla ze známých hodnot matice molárních absorpčních koeficientů a matice volných koncentrací  $L^-$  a  $HL$  dle Bouguer-Lambert-Beerova zákona vypočtena absorbanční matice  $A^{pure}$  na 9 platných ciferech.

K takto přesným hodnotám absorbanční matice byly připočteny náhodné chyby, vypočtené generátorem náhodných čísel pro předvolený šum přístroje  $s_{inst}(A)$ . Náhodné chyby vykazovaly Gaussovo rozdělení s nulovou střední hodnotou a směrodatnou odchylkou rovnou právě šumu  $s = s_{inst}(A)$ . Soubory spekter byly postupně simulovány pro hodnoty šumu  $s_{inst}(A) = 0.0003$ ,

0.0006, 0.0009, 0.0020 a 0.0040. Data byla simulována pro bromkrezolovou zeleň (soubory A1 až A5), fenolovou červeň (soubory A6 až A10) a tymolovou modř (soubory A11 až A15). Takto připravené soubory spekter popisovaly v podstatě dvě barevné částice rovnovážného systému, tj. L a HL. Poslední soubor spekter (A16) byl experimentálního charakteru a týkal se forem HL a L od směsi tří výše jmenovaných sulfonftaleinů.

Tabulka 2 ukazuje zajímavou skutečnost, kterou je třeba v další analýze respektovat, že totiž Kankareho metodou nalezená hodnota směrodatné odchylky absorbance  $s_k(A)$  je vždy nepatrně nižší než předvolená hodnota  $s_{inst}(A)$ .

Z testovaných čtrnácti metod faktorové analýzy prvních třináct naprosto bezpečně určilo správný počet světlo-absorbujících částic  $k = 2$  u všech patnácti souborů simulovaných dat, pouze Bartlettův test izotropie vyšel správně jen u dvou souborů a u třinácti selhal.

U experimentálních dat našlo skutečný počet světlo-absorbujících částic  $k = 6$  sedm metod, tj. *Kank*, *RE*, *XE*,  $\bar{e}$ ,  $\chi^2$ ,  $\sigma_g$  a  $\log(g)$ , zbylých sedm metod selhalo. Zajímavé je, že úspěšné metody jsou postaveny na znalosti hodnoty instrumentální chyby přístroje  $s_{inst}(A)$ . Vyjímkou je metoda  $\log(g)$ , která tuto informaci nepožaduje, i když je výhodné mít představu alespoň o řádové velikosti  $s_{inst}(A)$ , aby se při provádění této metody nesuperponoval na absorbanční matici příliš velký nebo naopak příliš malý šum.

## 2. Spektra spektrálních standardů (Tabulka 3)

K porovnání účinnosti metod faktorové analýzy, hledajících počet světlo-absorbujících částic, lze s výhodou použít spektra spektrálních standardů. Jedná se totiž o spektra vzniklá Beerovskou kalibrační závislostí absorbance na koncentraci spektrálního standardu.

Byly naměřeny čtyři experimentální soubory spekter (B1 až B4), ve kterých byly postupně kombinovány spektrální standardy  $K_2Cr_2O_7$ ,  $CoSO_4 \cdot 7H_2O$  a  $CuSO_4 \cdot 5H_2O$ . Až na soubor B1 byla u zbylých tří souborů spekter nalezena směrodatná odchylka absorbance  $s_k(A)$  výrazně nižší než dosazovaná hodnota  $s_{inst}(A)$  užívaného spektrofotometru. I tak lze však konstatovat, že  $s_k(A)$  i  $s_{inst}(A)$  dosahují mimořádně nízkých hodnot, svědčících o spolehlivosti (tj. správnosti a přesnosti) užívaného přístroje a experimentální techniky.

U souboru B1 sedm metod, tj. *RE*, *XE*,  $\bar{e}$ ,  $\chi^2$ ,  $\sigma_g$ ,  $g$  a *RPV*, našlo správně  $k = 2$ . U souboru B2 devět metod, tj. *Kank*, *RE*, *XE*,  $\bar{e}$ ,  $\chi^2$ ,  $\sigma_g$ , *RPV*,  $\log(g)$  a *F*, našlo správně  $k = 2$  světlo-absorbující částice. U souboru B3 našly správně  $k = 2$  tytéž metody jako u souboru B2. U tří spektrálních standardů ve směsi (soubor B4) správný počet světlo-absorbujících částic našlo devět metod, *Kank*, *RE*,  $\bar{e}$ ,  $\chi^2$ ,  $\sigma_g$ ,  $\log(g)$ , *IND*, *F* a *Bart*.

## 3. Spolehlivost testovaných metod (Tabulka 4)

Chceme-li kriticky zhodnotit spolehlivost testovaných čtrnácti metod, určujících počet světlo-absorbujících částic, je vhodné závěry analýzy souborů A a B vynést do přehledné tabulky (Tabulka 4). Z ní je dostatečně zřejmá procentuelní úspěšnost jednotlivých vyšetřujících metod.

Lze proto uzavřít, že za *spolehlivé* metody (skupina I, úspěšnost 91 až 100%) lze považovat Kankareho metodu, metodu reálné chyby, metodu extrahované chyby, metodu průměrné chyby, metodu směrodatných odchylek vlastních čísel a metodu logaritmu vlastních čísel. Za metody *občas selhávající* (skupina II, úspěšnost 81 až 90%) lze označit metodu  $\chi^2$  kriteria, metodu reziduální variance a Malinowského *F*-test. Jako *nedostatečně spolehlivé a nespolehlivé* (skupina III, úspěšnost 80 a méně %) se pak jeví metoda průměru vlastních čísel, metoda Exnerovy funkce, metoda zabudované chyby, metoda faktorové indikátorové funkce a Bartlettův test izotropie.

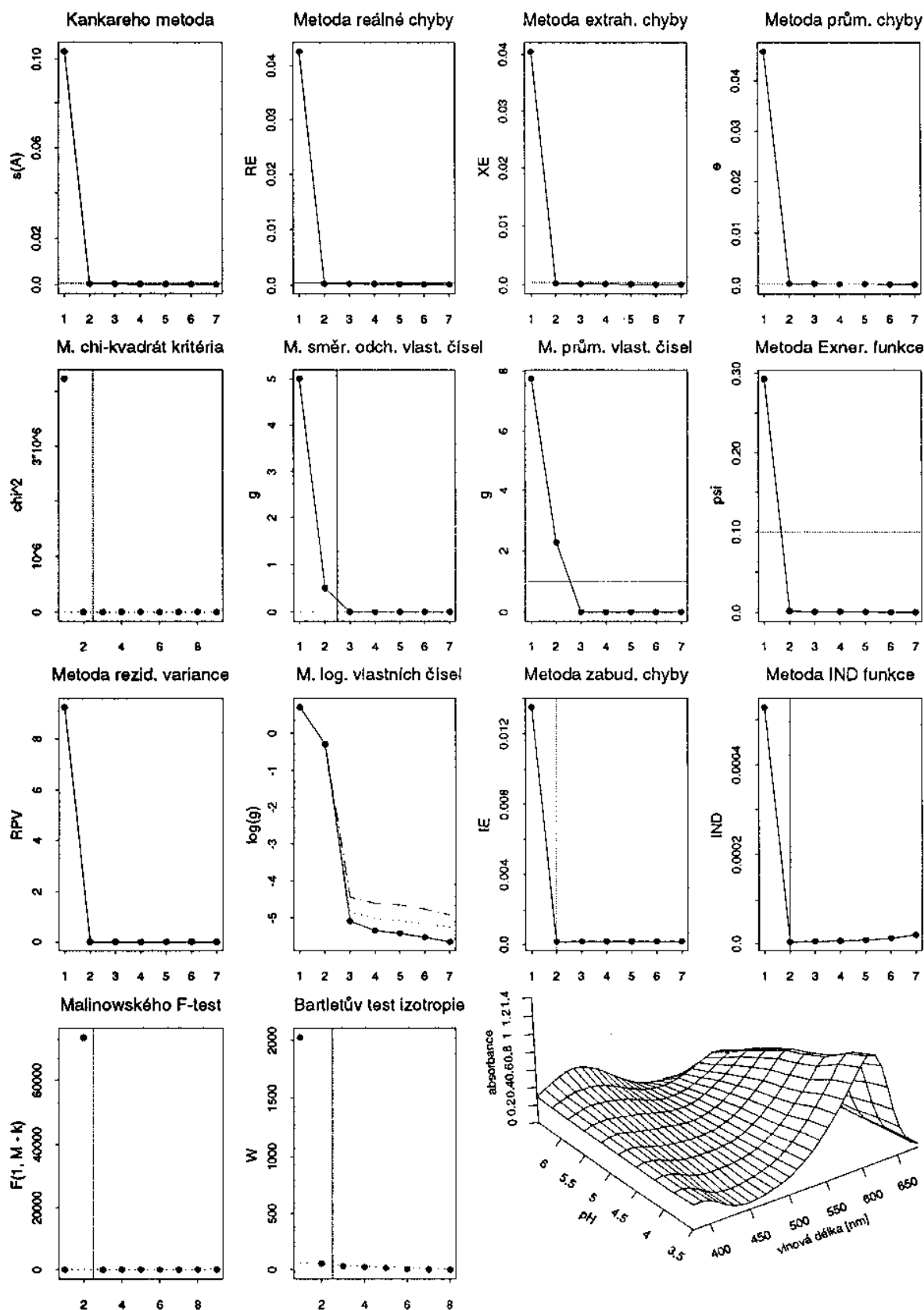
## ZÁVĚR

1. Analýzou spekter spektrálních standardů Wernimontovou metodou se ukázala **Kankareho metoda** jako nejvhodnější pro stanovení směrodatné odchylky absorbance  $s_{inst}(A)$  čili pro hladinu šumu užívaného spektrofotometru. Hodnota  $s_{inst}(A)$  bývá v literatuře označována jako přístrojová chyba a její velikost pro spektrofotometr SPECORD M400 (Carl Zeiss, Jena) ve viditelné oblasti 380 - 650 nm byla stanovena  $s_{inst}(A) = 0.0007$ .
2. Vyšetřením 14 metod faktorové analýzy na 20 souborech spekter o známém počtu světlo-absorbujících částic se ukázaly jako nejvěrohodnější Kankareho metoda, metoda reálné chyby, metoda extrahované chyby, metoda průměrné chyby, metoda směrodatných odchylek vlastních čísel a metoda logaritmů vlastních čísel.
3. Při testování 14 metod faktorové analýzy na 22 souborech spekter acidobazických a komplexotvorných rovnováh sulfoazoxinů o neznámém počtu světlo-absorbujících částic dospělo šest metod ke shodným závěrům. Jsou to Kankareho metoda, metoda reálné chyby, metoda extrahované chyby, metoda průměrné chyby, metoda  $\chi^2$  kritéria a metoda logaritmů vlastních čísel.
4. Průnikem množin metod závěru 3. a závěru 4. je výsledná množina pěti metod, které se ukázaly v této studii jako **nejspolehlivější** a jsou proto doporučovány k obecnějšímu použití: Kankareho metoda, metoda reálné chyby, metoda extrahované chyby, metoda průměrné chyby a metoda logaritmů vlastních čísel.
5. Obecně se spolehlivějšími jeví metody založené na <sup>z nejspolehlivější</sup> odhadu experimentální chyby v datech. Z metod tuto informaci nepožadujících lze doporučit pouze metodu logaritmů vlastních čísel, pro kterou je však výhodné znát alespoň řádový odhad velikosti instrumentální chyby  $s_{inst}(A)$ .
6. Při stanovování počtu světlo-absorbujících částic analýzou spekter je doporučené postupovat podle zásady, že konfrontace více kritérií zvyšuje spolehlivost.

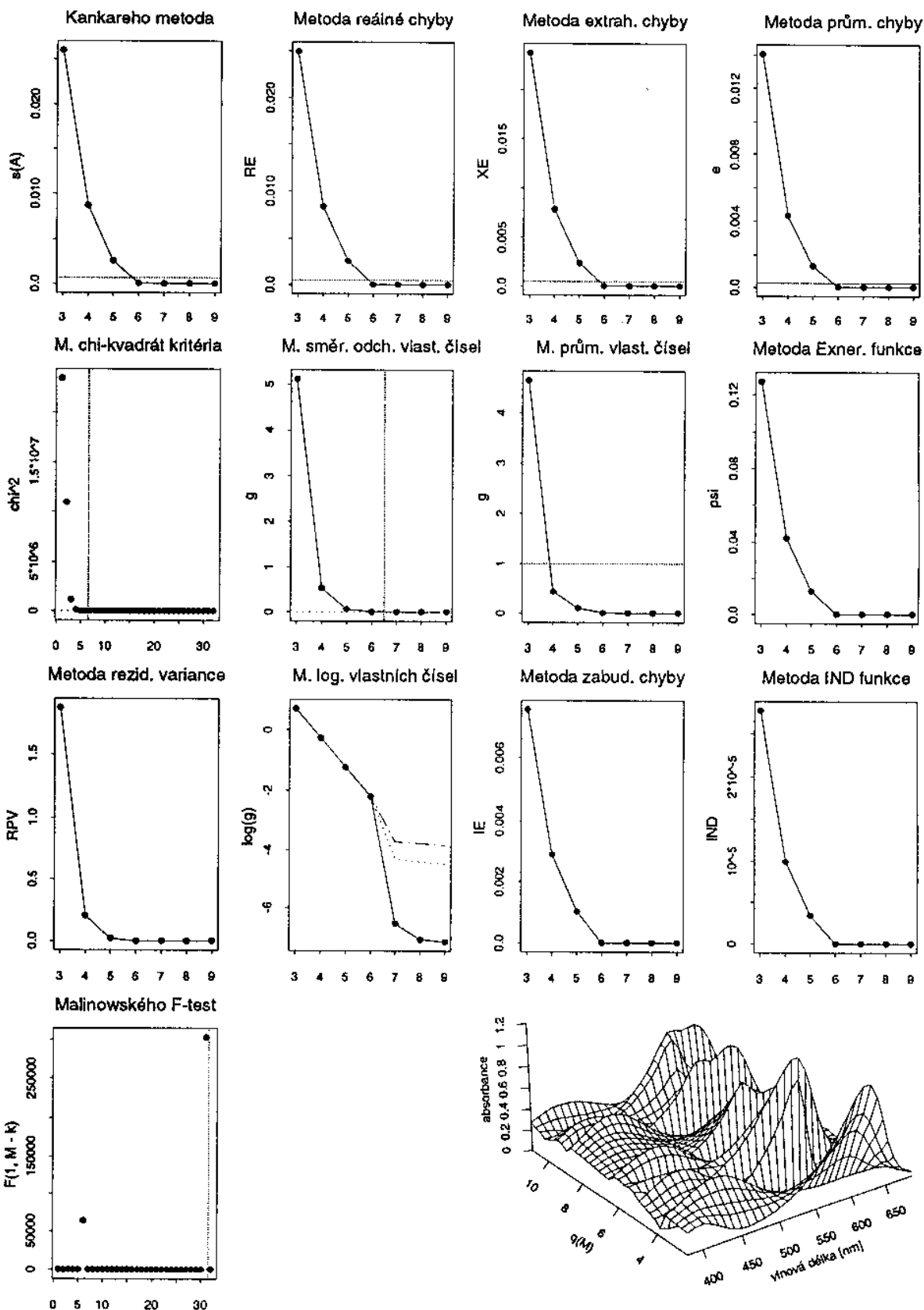
## LITERATURA

1. Malinowski E. R. (1977) *Anal. Chem.* 49: 606.
2. Kankare J. J. (1970) *Anal. Chem.* 42: 1322.
3. Brereton R. G., *Data Handling in Science and Technology - Volume 9: Multivariate Pattern Recognition in Chemometrics, Chapter 5*, J. M. Deane: *Data Reduction Using Principal Components Analysis*, Elsevier, Amsterdam - London - New York - Tokyo, 1992.
4. Čepan M., Pelikán P., Liška M.: *Metódy faktorovej analýzy v molekulovej spektrometrii I. Abstraktná faktorová analýza a určenie počtu aktívnych zložiek v spektrách zmesí*, (in press).
5. Hugus Z. Z. Jr., El-Awady A. A. (1971) *J. Phys. Chem.* 75: 2954.
6. Malinowski E. R. (1987) *J. Chemometrics* 1: 33.
7. Malinowski E. R. (1989) *J. Chemometrics* 3: 49.
8. Antoon M. K., D'Esposito L., Koenig J. L. (1979) *Appl. Spectrosc.* 33: 351.
9. Wernimont G. (1967) *Anal. Chem.* 39: 554.





**Soubor A2** Výsledky metod pro vyšetření počtu světlo-absorbujících částic pro simulovaná data acidobazické titrace bromkrezolové zeleně pro koncentrace  $c_{BKZ} = 5.444 - 5.379 \times 10^{-4} \text{ mol.l}^{-1}$  a  $\text{pH} = 6.49 - 3.49$ . Na přesná data je superponován výběr náhodných chyb s nulovou střední hodnotou a směrodatnou odchylkou 0.0006. Vpravo dole je zobrazena příslušná absorbanční matice



**Soubor A16** Výsledky metod pro vyšetření počtu světlo-absorbujících částic pro acidobazickou titraci směsi bromkrezolové zeleně, fenolové červeně a tymolové modří při rozsahu koncentrací  $c_{BKZ} = 0.000 - 2.407 \times 10^{-4}$  mol.l<sup>-1</sup>,  $c_{FC} = 0.000 - 1.407 \times 10^{-4}$  mol.l<sup>-1</sup>,  $c_{TM} = 0.000 - 3.205 \times 10^{-4}$  mol.l<sup>-1</sup> a pH = 2.52 - 11.00. Vpravo dole je zobrazena příslušná absorbanční matice

**Tabulka 2 Spektra acidobazických rovnovah sulfonftaleinů**

(BKZ: bromkrezolová zeleň, FC: fenolová červeň, TM: tymolová modř), Kank Kankareho metoda, RE metoda reálné chyby, XE metoda extrahované chyby,  $\bar{\epsilon}$  metoda průměrné chyby,  $\chi^2$  metoda  $\chi^2$  kritéria,  $\sigma_g$  metoda směrodatných odchylek vlastních čísel, g metoda průměru vlastních čísel, psi metoda Exnerovy funkce, RPV metoda reziduální variance, log(g) metoda logaritmu vlastních čísel, IE metoda zabudované chyby, IND metoda faktorové ind. funkce, F Malinowského F-test, Bart Bartlettův test izotropie, (-) značí neurčený odhad

Soubor	Spektra pro rovnováhu	Typ dat, $s_{\text{inf}}(A)$ přístroje, $s_g(A)$ nalezené (Kank)	Nalezený počet světlo-absorbujících částic (v závorce u každé metody)
A1	BKZ - H <sup>+</sup>	Simulace, 0.0003, 0.00027	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A2	BKZ - H <sup>+</sup>	Simulace, 0.0006, 0.00051	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (2)
A3	BKZ - H <sup>+</sup>	Simulace, 0.0009, 0.00078	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A4	BKZ - H <sup>+</sup>	Simulace, 0.0020, 0.00173	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A5	BKZ - H <sup>+</sup>	Simulace, 0.0040, 0.00376	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (3), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (-), IND (2), F (2), Bart (3)
A6	FC - H <sup>+</sup>	Simulace, 0.0003, 0.00024	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A7	FC - H <sup>+</sup>	Simulace, 0.0006, 0.00053	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A8	FC - H <sup>+</sup>	Simulace, 0.0009, 0.00081	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A9	FC - H <sup>+</sup>	Simulace, 0.0020, 0.00172	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)

A10	FC - H <sup>+</sup>	Simulace, 0.0040, 0.00347	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (-), IND (2), F (2), Bart (2)
A11	TM - H <sup>+</sup>	Simulace, 0.0003, 0.00027	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A12	TM - H <sup>+</sup>	Simulace, 0.0006, 0.00054	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (-), IND (2), F (2), Bart (1)
A13	TM - H <sup>+</sup>	Simulace, 0.0009, 0.00079	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A14	TM - H <sup>+</sup>	Simulace, 0.0020, 0.00195	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (3), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (-), IND (2), F (2), Bart (1)
A15	TM - H <sup>+</sup>	Simulace, 0.0040, 0.00324	Kank (2), RE (2), XE (2), $\bar{\epsilon}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (2), IE (2), IND (2), F (2), Bart (1)
A16	BKZ - FC - TM - H <sup>+</sup>	Experimentální, 0.0015, 0.00008	Kank (6), RE (6), XE (6), $\bar{\epsilon}$ (6), $\chi^2$ (6), $\sigma_g$ (6), g (3), psi (2), RPV (5), log(g) (6), IE (-), IND (-), F (31), Bart (-)

### Tabulka 3 Spektra spektrálních standardů

(dichroman draselný  $K_2Cr_2O_7$ , síran kobaltnatý  $CoSO_4 \cdot 7H_2O$  a síran měďnatý  $CuSO_4 \cdot 5H_2O$ ), Kank Kankareho metoda, RE metoda reálné chyby, XE metoda extrahované chyby,  $\bar{e}$  metoda průměrné chyby,  $\chi^2$  kritéria,  $\sigma_g$  metoda směřodatných odchylek vlastních čísel, g metoda průměru vlastních čísel, psi metoda Exnerovy funkce, RPV metoda reziduální variance, log(g) metoda logaritmu vlastních čísel, IE metoda zabudované chyby, IND metoda faktorové indikátorové funkce, F Malinowského F-test, Bart Bartlettův test izotropie, (-) značí neurčený odhad

Soubor	Spektra pro rovnováhu	Typ dat, $s_{inst}(A)$ přístroje, $s_x(A)$ nalezené (Kank)	Nalezený počet světlo-absorbujících částic (v závorce u každé metody)
B1	$K_2Cr_2O_7 - CoSO_4 \cdot 7H_2O$	Experimentální, 0.0007, 0.00089	Kank (3), RE (2), XE (2), $\bar{e}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (2), psi (1), RPV (2), log(g) (3), IE (-), IND (5), F (3), Bart (14)
B2	$CoSO_4 \cdot 7H_2O - CuSO_4 \cdot 5H_2O$	Experimentální, 0.0007, 0.00047	Kank (2), RE (2), XE (2), $\bar{e}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (1), psi (1), RPV (2), log(g) (2), IE (-), IND (8), F (2), Bart (11)
B3	$K_2Cr_2O_7 - CuSO_4 \cdot 5H_2O$	Experimentální, 0.0007, 0.00047	Kank (2), RE (2), XE (2), $\bar{e}$ (2), $\chi^2$ (2), $\sigma_g$ (2), g (1), psi (1), RPV (2), log(g) (2), IE (-), IND (7), F (2), Bart (10)
B4	$K_2Cr_2O_7 - CoSO_4 \cdot 7H_2O - CuSO_4 \cdot 5H_2O$	Experimentální, 0.0007, 0.00023	Kank (3), RE (3), XE (3), $\bar{e}$ (3), $\chi^2$ (3), $\sigma_g$ (2), g (2), psi (2), RPV (2), log(g) (3), IE (-), IND (3), F (3), Bart (3)

**Tabulka 4 Spolehlivost testovaných metod: 1 odpovídá správně a 0 nesprávně určenému počtu světlo-absorbujících částic, (Kank Kankareho metoda, RE metoda reálné chyby, XE metoda extrahované chyby,  $\bar{e}$  metoda průměrné chyby,  $\chi^2$  metoda  $\chi^2$  kritéria,  $\sigma_g$  metoda směrodatných odchylek vlastních čísel, g metoda průměru vlastních čísel, psi metoda Exnerovy funkce, RPV metoda reziduální variance, log(g) metoda logaritmu vlastních čísel, IE metoda zabudované chyby, IND metoda faktorové indikátorové funkce, F Malinowského F-test, Bart Bartlettův test izotropie)**

Soub.	Kank	RE	XE	$\bar{e}$	$\chi^2$	$\sigma_g$	g	psi	RPV	log(g)	IE	IND	F	Bart
A1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A3	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A4	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A5	1	1	1	1	0	1	1	1	1	1	0	1	1	0
A6	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A7	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A8	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A9	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A10	1	1	1	1	1	1	1	1	1	1	0	1	1	1
A11	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A12	1	1	1	1	1	1	1	1	1	1	0	1	1	0
A13	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A14	1	1	1	1	0	1	1	1	1	1	0	1	1	0
A15	1	1	1	1	1	1	1	1	1	1	1	1	1	0
A16	1	1	1	1	1	1	0	0	0	1	0	0	0	0
B1	0	1	1	1	1	1	1	0	1	0	0	0	0	0
B2	1	1	1	1	1	1	0	0	1	1	0	0	1	0
B3	1	1	1	1	1	1	0	0	1	1	0	0	1	0
B4	1	1	1	1	1	0	0	0	0	1	0	1	1	1
Spr.	19	20	20	20	18	19	16	15	18	19	11	16	18	3
Úspěš. [%]	95	100	100	100	90	95	80	75	90	95	55	80	90	15
Skup.	I	I	I	I	II	I	III	III	II	I	III	III	II	III

Tabulka 5 Souhlas testovaných metod při vyšetřování rovnovah sulfoazoxinů: 1 odpovídá shodně a 0 neshodně určenému počtu světlo-absorbujících částic, shoda znamená souhlas s většinou metod, (Kank Kankareho metoda, RE metoda reálné chyby, XE metoda extrahované chyby,  $\bar{e}$  metoda průměrné chyby,  $\chi^2$  metoda  $\chi^2$  kriteria,  $\sigma_g$  metoda směrodatných odchylek vlastních čísel, g metoda průměru vlastních čísel, psi metoda Exnerovy funkce, RPV metoda reziduální variance, log(g) metoda logaritmů vlastních čísel, IE metoda zabudované chyby, IND metoda faktorové indikátorové funkce, F Malinowského F-test, Bart Bartlettův test izotropie)

Soub.	Kank	RE	XE	$\bar{e}$	$\chi^2$	$\sigma_g$	g	psi	RPV	log(g)	IE	IND	F	Bart
C1	1	1	1	1	1	0	0	0	1	1	0	0	1	0
C2	1	1	1	1	1	0	0	0	1	1	0	0	1	0
C3	1	1	1	1	1	0	0	0	0	1	0	0	0	0
C4	1	1	1	1	1	1	0	0	1	1	0	0	0	0
C5	1	1	1	1	0	0	0	0	1	1	0	0	0	0
C6	1	1	1	1	1	0	0	0	1	1	0	0	1	0
D1	1	1	1	1	1	0	0	0	0	1	0	0	1	0
D2	1	1	1	0	1	1	0	0	1	1	0	0	0	0
D3	0	1	1	1	1	0	0	0	1	1	0	0	1	0
D4	1	1	1	1	1	0	0	0	1	1	0	0	0	0
D5	1	1	1	1	1	0	0	0	1	1	0	0	1	0
D6	1	1	1	1	1	0	0	0	1	1	0	0	1	0
D7	1	1	1	1	1	1	0	0	1	1	0	1	1	1
D8	1	1	1	1	1	1	0	0	1	1	0	0	0	0
D9	1	1	1	1	1	0	1	1	0	1	0	0	1	0
D10	1	1	1	1	1	1	0	0	1	1	0	0	1	0
D11	1	1	1	1	1	0	0	0	1	1	0	1	1	0
D12	1	1	1	1	1	0	0	0	1	1	0	0	0	0
D13	1	1	1	1	1	1	1	1	1	1	0	0	0	0
D14	1	1	1	1	1	0	0	0	1	1	0	1	1	0
D15	0	1	1	0	1	1	1	1	0	1	0	0	0	0
D16	1	1	1	1	1	1	1	1	1	1	0	0	1	0
Souhl.	20	22	22	20	21	8	4	4	18	22	0	3	13	1
Úspěš. [%]	91	100	100	91	95	36	18	18	82	100	0	14	59	5
Skup.	I	I	I	I	I	III	III	III	II	I	III	III	III	III