

# Počet světlo-absorbujících částic v rovnovážné směsi faktorovou analýzou spekter

Ing. Zuzana Ferenčíková a prof. RNDr. Milan Meloun. DrSc., *Katedra analytické chemie, Univerzita Pardubice, 532 10 Pardubice*

*[zuzana.ferencikova@student.upce.cz](mailto:zuzana.ferencikova@student.upce.cz), [milan.meloun@upce.cz](mailto:milan.meloun@upce.cz)*

## *Souhrn:*

Při studiu protonačních rovnováh pomocí spektrofotometrické titrace při více vlnových délkách je vždy důležité vědět, kolik částic se nachází ve studovaném roztoku. K odhalení počtu světlo-absorbujících částic je výhodné využít metod faktorové analýzy. K tomuto účelu byl využit program S-PLUS, ve kterém bylo naprogramováno 12 metod faktorové analýzy, kterými je odhadován počet barevných částic v roztoku. Bez znalosti počtu světlo-absorbujících částic v roztoku nelze správně navrhnout chemický model, na jehož základě se dále nelineární regresí vyčíslují odhady protonačních konstant  $pK_i$ .

## 1. Úvod

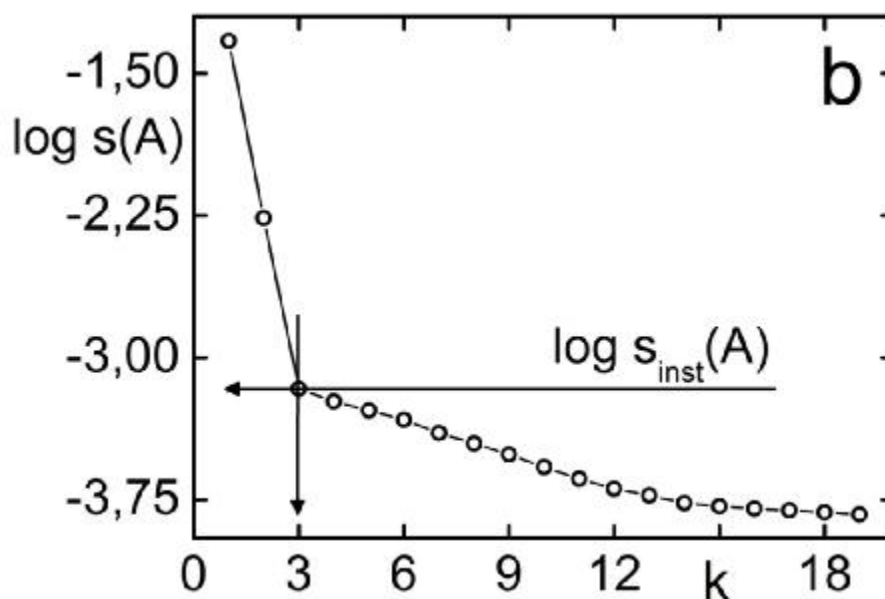
Faktorová analýza patří mezi metody pro snížení dimenzionality čili k redukci počtu původních znaků. Hlavními cíly faktorové analýzy jsou sumarizace dat a jejich redukce. V sumarizaci dat využívá tato metoda faktorů tak, aby data vysvětlila a usnadnila jejich pochopení daleko menším počtem latentních proměnných, než je počet původních znaků. Redukce dat je dosaženo vyčíslením skóre pro každý faktor a následnou náhradou původních znaků novými latentními proměnnými čili faktory.

Ve spektrofotometrii je datovou maticí absorbanční matice  $\mathbf{A}$ , jejíž sloupce představují měřené vlastnosti (pH) a řádky pak vlnové délky. Zdrojovou matici  $\mathbf{A}$  o rozměrech  $n \times m$  lze pak vyjádřit ve tvaru  $\mathbf{A}=\mathbf{L}\mathbf{F}+\mathbf{E}$  (obr. 1), kde  $\mathbf{L}$  s rozměry  $n \times k$  představuje matici faktorového skóre,  $\mathbf{F}$  o rozměrech  $k \times m$  představuje matici faktorových zátěží a  $\mathbf{E}$  o rozměrech  $n \times m$  představuje matici chybových faktorů. V našem případě je  $n$  počet spekter,  $m$  počet vlnových délek,  $k$  index testované komponenty a  $r$  skutečný počet světlo-absorbujících částic ve směsi.

$$\begin{array}{|c|} \hline \text{Zdrojová matice} \\ \hline \mathbf{A} \\ \hline \end{array} = \begin{array}{|c|} \hline \text{Matice skóre} \\ \hline \mathbf{L} \\ \hline \end{array} \times \begin{array}{|c|} \hline \text{Matice zátěží} \\ \hline \mathbf{F} \\ \hline \end{array} + \begin{array}{|c|} \hline \text{Matice chyb} \\ \hline \mathbf{E} \\ \hline \end{array}$$

Obr. 1 Rozklad zdrojové matice

O prvním grafickém výstupu faktorové analýzy lze obecně hovořit jako o tzv. „Cattelově grafu úpatí vlastních čísel“, ve kterém se vynášejí vypočtená funkce  $f(k)$  proti počtu komponent  $k$ . Hlavním kritériem určení počtu komponent  $k$  je u většiny metod zlom na křivce  $f(k)$  vs.  $k$  (Obr. 2). Cattelův graf úpatí je výhodnější vynášet jako logaritmus hodnot funkce  $\log(f(k))$  pro dosažení lepšího rozlišení zlomu křivky, což mnohdy napomůže k odhalení i minoritních složek ve směsi.



Obr. 2 Graf úpatí

## 2. Metody faktorové analýzy

### 2.1. Precizní metody

Počet komponent ve směsi se u precizních metod určuje na základě porovnání zbytkové reziduálové chyby  $s_k(\mathbf{A})$  pro  $k$ -tou komponentu s instrumentální chybou přístroje  $s_{inst}(\mathbf{A})$ . Do grafu je pak vynášena směrodatná odchylka absorbance  $s_k(\mathbf{A})$  vůči počtu komponent  $k$ . Řešením je pak takové číslo  $k$ , u kterého dochází na křivce k prudkému zlomu. Do metod precizních řadíme Kankareho reziduálovou směrodatnou odchylku absorbance  $s_k(\mathbf{A})$ , reziduálovou směrodatnou odchylku  $RSD$ , odmocninu z průměrné chyby  $RMS$ , kritérium průměrné absorbance  $AE$ , Bartlettův  $\chi^2$  test a směrodatnou odchylku vlastních čísel  $s(g_k)$ . Kankareho metoda a Bartlettův  $\chi^2$  test patří mezi metody nejspolehlivější.

#### 2.1.1. Kankareho reziduálová směrodatná odchylka absorbance $s_k(\mathbf{A})$

Výpočet vychází z druhého momentu  $\mathbf{M}$  analyzované absorbanční matice  $\mathbf{A}$  dle

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

kde  $n$  je počet vlnových délek. Použitím vlastních čísel  $r_a$  matice  $\mathbf{M}$  se určí zbytková směrodatná odchylka absorbance  $s_k(\mathbf{A})$  vztahem

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

kde  $\text{tr}(\mathbf{M})$  je stopa matice  $\mathbf{M}$  a  $k$  je počet latentních proměnných, pro který je vyčíslen pro výslednou  $s_k(\mathbf{A})$  a představuje zde počet světlo-absorbujících komponent.

#### 2.1.2. Bartlettův $\chi^2$ test

Pro data, jejichž směrodatná odchylka se mění od bodu k bodu navrhl Bartlett použít metodu  $\chi^2$ -kritéria. Jedná se o určení první matice  $\mathbf{A}^{\text{pred}}$ , která reprodukuje matici  $\mathbf{A}$  v rámci

experimentální chyby. Nevýhodou je nutnost znát směrodatnou odchylku jednotlivě u každého bodu. Výpočet probíhá podle rovnice

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

kde  $\sigma_{ij}$  je směrodatná odchylka měření asociovaná s každým prvkem absorbanční matice, označeným  $A_{ij}$ . Hodnoty  $\chi^2$  pro  $k = 1, 2, \dots$  se porovnávají s kritickou hodnotou  $\chi^2_{\text{krit}} = (n - k)(m - k)$ . Počet světlo absorbujících částic je pak roven takovému  $k$ , pro nějž je hodnota  $\chi^2$  menší než výše uvedená kritická hodnota.

## 2.2. Aproximativní metody

Pokud není známa směrodatná odchylka přístroje  $s_{inst}(\mathbf{A})$ , využijeme metod aproximativních. Mezi tyto metody řadíme logaritmus vlastních čísel  $\log g_a$ , Exnerovu funkci  $\psi$ , graf úpatí  $RPV(k)$ , vloženou chybu  $IE$ , faktorovou indikátorovou funkci  $IND(k)$  a metodu  $RESO$ . Ve většině případů se jedná o funkce empirické.

### 2.2.1. Graf úpatí

Tato metoda určuje počet komponent za předpokladu, že rozptyl reziduí se stabilizuje na menším počtu dimenzí, které obsahují pouze náhodnou chybu. Rozptyl reziduí je definován vztahem

$$RV(k) = \frac{\sum_{i=1}^n \sum_{j=1}^m (A_{ij} - \hat{A}_{ij})^2}{nm},$$

který je roven čtverci odmocnině z průměrné chyby  $RMS(k)$ . Procentoreziduálového rozptylu se vypočte podle vztahu

$$RPV(k) = 100 \left( \frac{\sum_{i=1}^n \sum_{j=1}^m (A_{ij} - \hat{A}_{ij})^2}{\sum_{i=1}^n \sum_{j=1}^m A_{ij}^2} \right) = 100 \left( \frac{\sum_{a=k+1}^m s_a}{\sum_{a=1}^m s_a} \right),$$

který může být vyjádřen i

pomocí vlastních čísel.

*Testování:* Do grafu se vynáší procenta reziduálového rozptylu  $RPV(k)$  proti indexu  $k$ . Počet světlo-absorbujících částic  $r$  je indikován takovou hodnotou  $k$ , u které dochází ke zlomu na křivce.

### 2.2.2. Faktorová indikátorová funkce

Jde o empirickou funkci, která je dána vztahem

$$IND(k) = \sqrt{\frac{\sum_{j=k+1}^q g_j}{n(m-k)}} = \frac{RSD(k)}{(q-k)^2}.$$

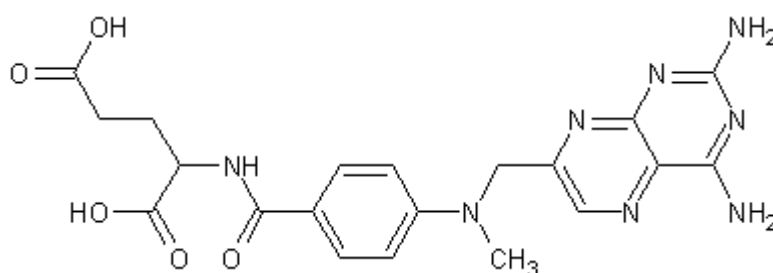
*Testování:* Za počet světlo-absorbujících částic  $r$  je považován bod  $k$  na křivce, kdy je dosaženo minima.

### 2.3. Derivační metody

Pokud je počet komponent ve směsi větší než tři, může se stát, že specifická kritéria nemusí vypovídat zcela jasně. Proto je výhodnější použít metod derivací pro zřetelnější odhalení zlomu na křivce. Do derivačních metod řadíme metodu druhé derivace  $SD$ , metodu třetí derivace  $TD$  a konečně metodu poměru derivací  $ROD$ . U metody druhé derivace a metody poměru derivací je počet světlo- absorbujičích částic roven takovému  $k$ , pro nějž křivka funkce  $SD(k)$ , resp. křivka funkce  $ROD(k)$  vykazuje první maximum, u metody třetí derivace je pak tento počet určen takovým  $k$ , pro nějž křivka funkce  $TD(k)$  vykazuje první minimum. Za první minimum či maximum ovšem nebereme  $k = 2$ .

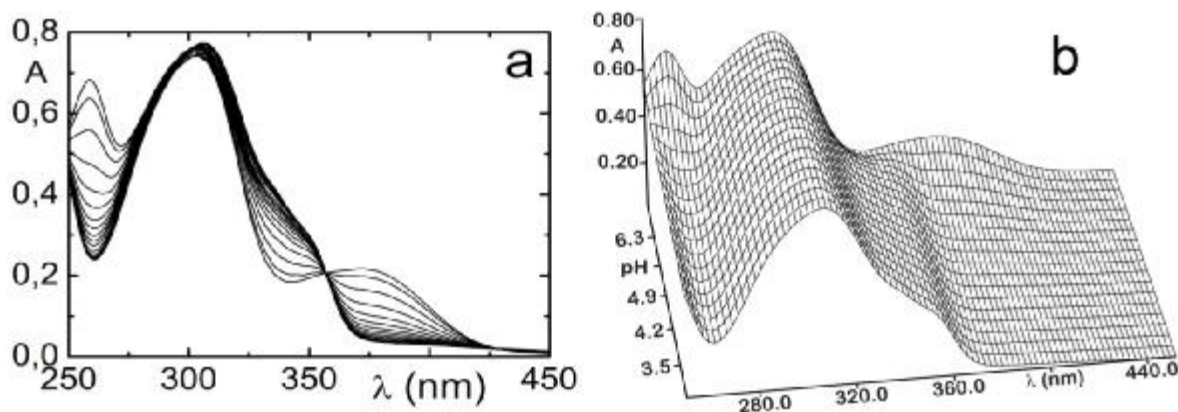
### 3. Stanovení počtu světlo- absorbujičích částic v rovnovážné směsi

Využití metod faktorové analýzy (FA) pro určení počtu světlo- absorbujičích částic ve směsi bude demonstrován na protonačních rovnováhách léčiva methotrexát (Obr. 3).



Obr. 3 Methotrexát

Toto léčivo vykazuje tři velmi blízké hodnoty protonačních konstant  $pK_i$  jak plyne z distribučního diagramu (Obr. 9b). O blízkých protonačních konstantách mluvíme tehdy, pokud  $|pK_{i+1} - pK_i| \leq 3$ . Proto byl kladen velký důraz na pečlivé provedení experimentu a dostatečně přesná data. Spektra byla měřena na spektrometru GBC Cintra 40 s instrumentální chybou  $s_{inst}(A) = 0.35$  mAU, což zaručuje dostatečnou přesnost naměřených dat.

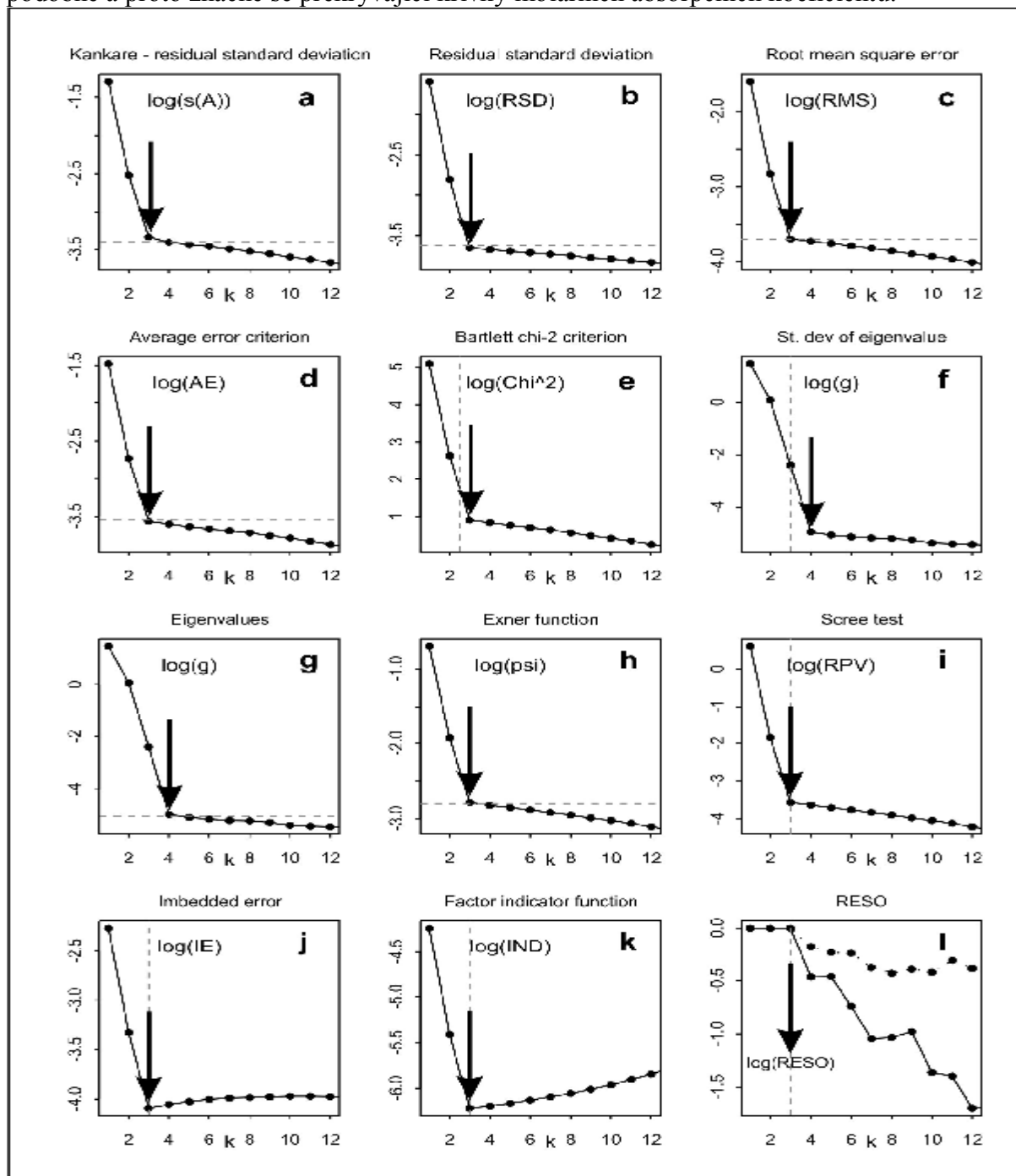


Obr.4 (a) 2D-graf naměřených spekter, (b) 3D-graf naměřených spekter

Naměřená spektra byla v programu S-PLUS testována algoritmem INDICES na počet světlo-absorbujících částic. Algoritmus INDICES je tvořen 12 zmíněnými metodami faktorové analýzy. Obr. 5 přináší grafy logaritmu metod faktorové analýzy a řešení jsou vždy označena šipkou. Těchto 12 metod odhalilo, že protonační rovnováhy methotrexátu se účastní 3 světlo-absorbující částice. Ve složitějších případech je výhodnější užít derivačních metod zobrazení Cattelova grafu úpatí. Výsledky obou derivačních metod jsou na obr. 6 a 7, poměr

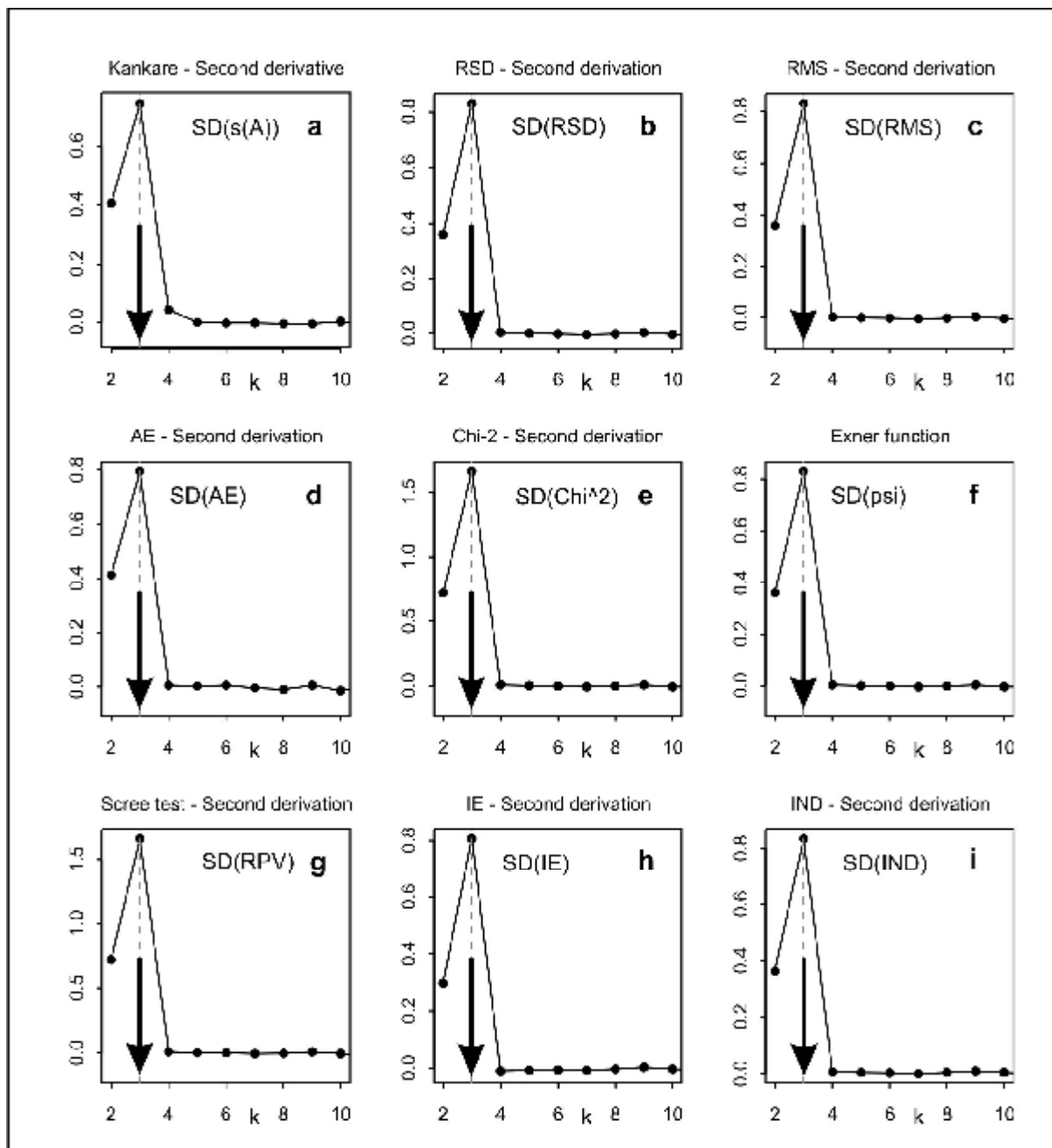
derivací na obr. 8. Nalezený počet částic je opět označen šipkou. Je zřejmé, že druhá derivace  $SD(k)$  vykazuje maximum u  $k = 3$ , třetí derivace  $TD(k)$  vykazuje minimum opět u  $k = 3$  stejně jako poměr derivací  $ROD(k)$  u  $k = 3$ .

Z výstupů programu S-PLUS lze říci, že počet světlo-absorbujících částic, účastnících se protonačních rovnováh methotrexátu, je roven  $r = 3$ . Pokud však porovnáme výsledky algoritmu INDICES s výsledky nelineární regrese, konkrétně s grafem molárních absorpčních koeficientů jednotlivých částic (obr. 9) zjistíme, že počet světlo-absorbujících částic ve směsi je roven 4. Nesrovnalost je způsobena tím, že částice  $LH_2$  a  $LH_3$  vykazují barevně velmi podobné a proto značně se překrývající křivky molárních absorpčních koeficientů.

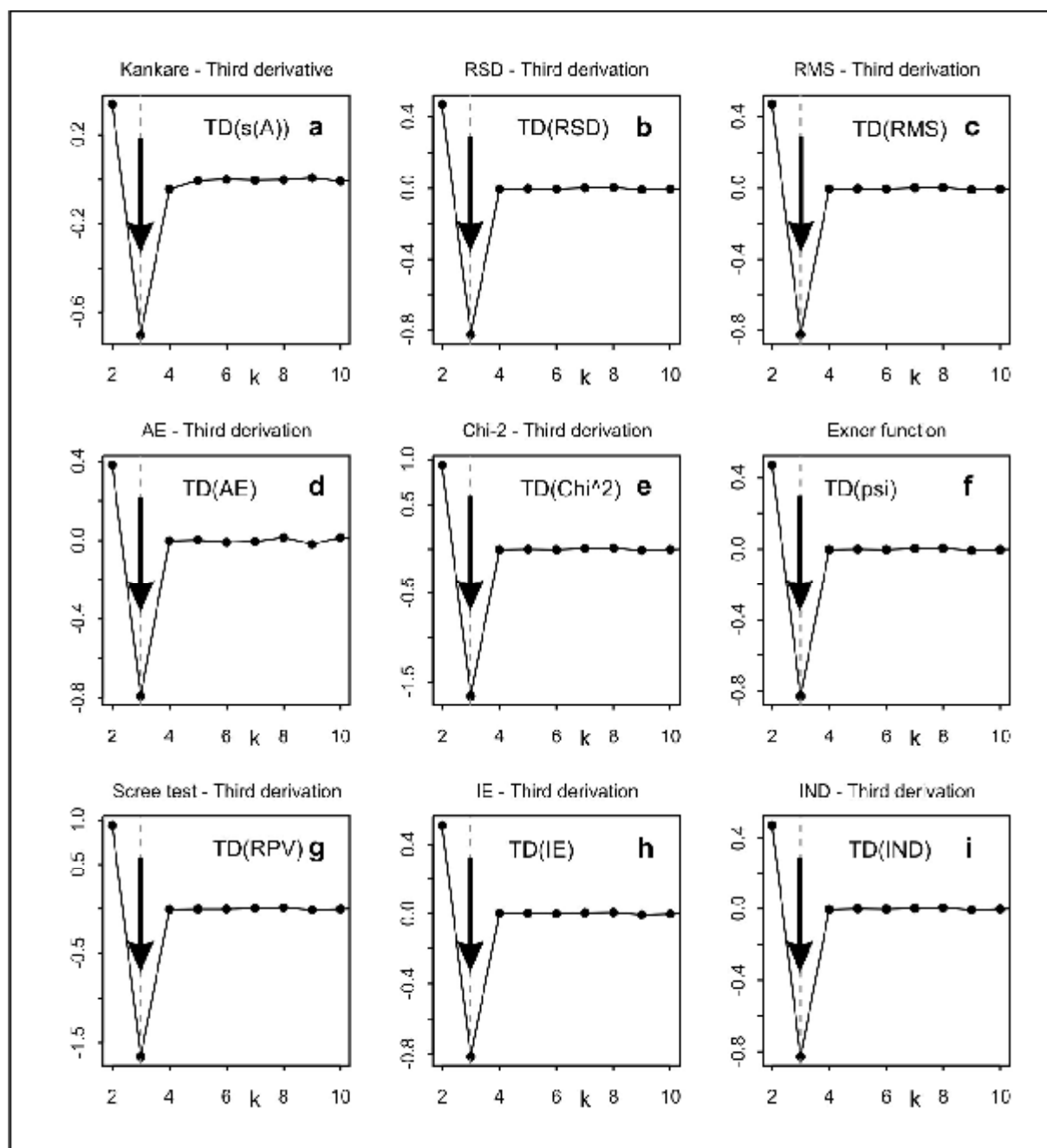


Obr. 5 Grafy závislosti logaritmů metod faktorové analýzy na počtu komponent. (a) Kankareho metoda, (b) metoda reálné chyby, (c) metoda extrahované chyby, (d) metoda

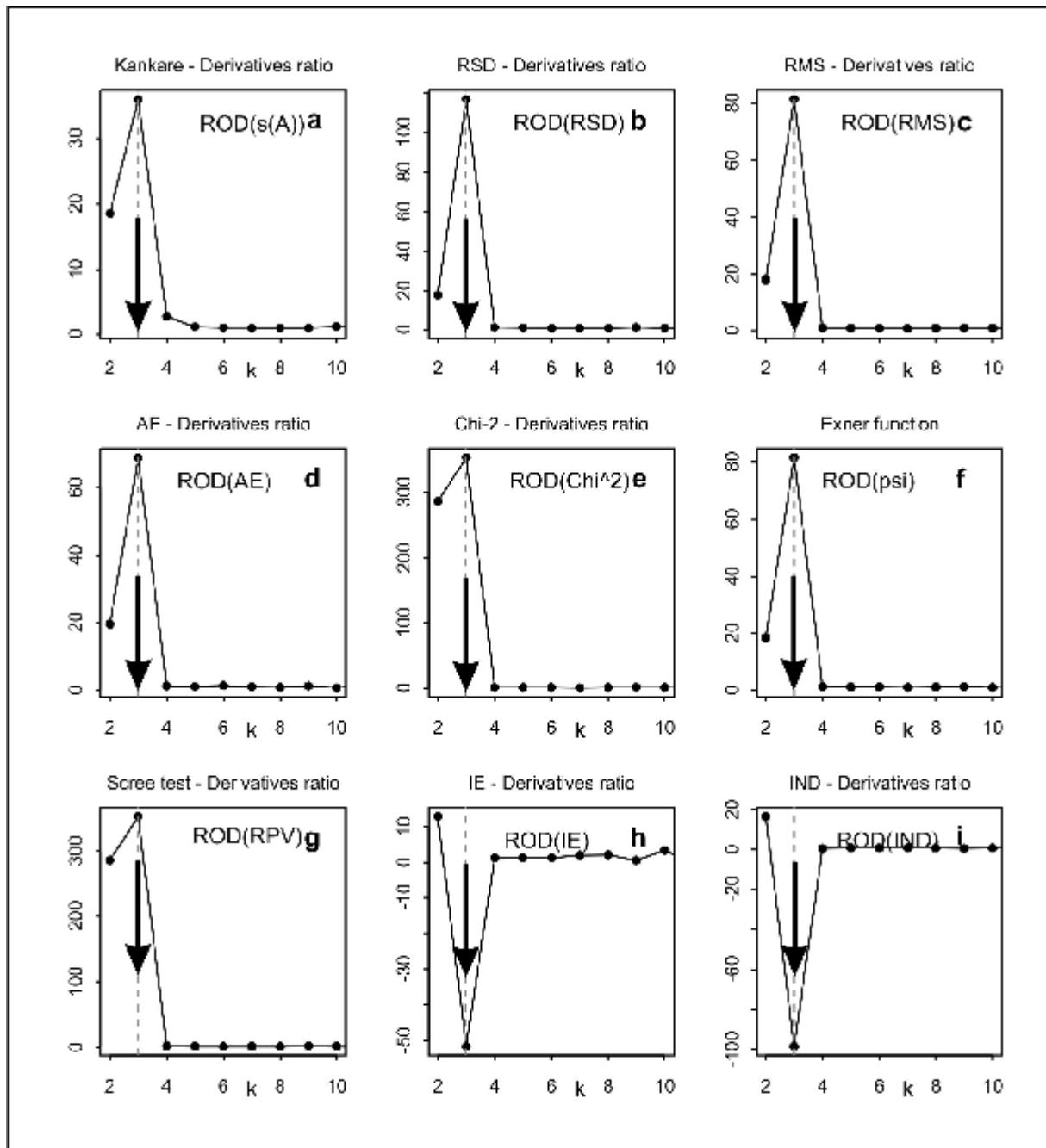
průměrné chyby, (e) Bartlettova metoda  $\chi^2$ -kritéria, (f) metoda odhadu směrodatné odchylky vlastních čísel, (g) metoda průměru vlastních čísel, (h) metoda Exnerovy funkce, (i) metoda reziduální variace, (j) metoda zabudované chyby, (k) metoda faktorové indikátorové funkce, (l) RESO metoda.



Obr. 6 Grafy druhých derivací metod faktorové analýzy: (a) Kankareho metoda, (b) metoda reálné chyby, (c) metoda extrahované chyby, (d) metoda průměrné chyby, (e) Bartlettova metoda  $\chi^2$ -kritéria, (f) metoda Exnerovy funkce, (g) metoda reziduální variace, (h) metoda zabudované chyby, (i) metoda faktorové indikátorové funkce.

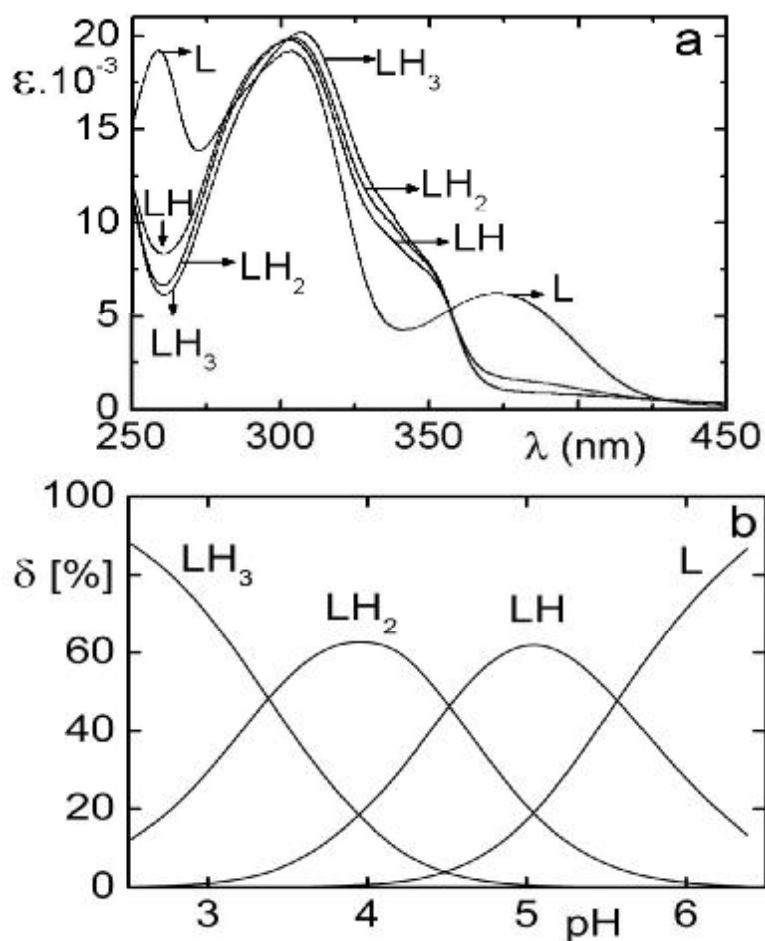


Obr. 7 Grafy devíti metod třetích derivací (TD) modifikací Cattelova indexového grafu vlastních čísel (po řádcích): (a) Kankareho metoda, (b) metoda reálné chyby, (c) metoda extrahované chyby, (d) metoda průměrné chyby, (e) Bartlettova metoda  $\chi^2$ -kritéria, (f) metoda Exnerovy funkce, (g) metoda reziduální variace, (h) metoda zabudované chyby, (i) metoda faktorové indikátorové funkce, (S-Plus).



Obr. 8 Grafy devíti metod poměru derivací (ROD) modifikací Cattelova indexového grafu vlastních čísel (po řádcích): (a) Kankareho metoda, (b) metoda reálné chyby, (c) metoda extrahované chyby, (d) metoda průměrné chyby, (e) Bartlettova metoda  $\chi^2$ -kritéria, (f) metoda Exnerovy funkce, (g) metoda reziduální variace, (h) metoda zabudované chyby, (i) metoda faktorové indikátorové funkce, (S-Plus).





Obr. 9 (a) Graf molárních absorpčních koeficientů pro částice L, LH, LH<sub>2</sub> a LH<sub>3</sub> methotrexátu, (b) distribuční diagram relativního zastoupení rozličně protonovaných částic methotrexátu v závislosti na pH.

#### Poděkování:

Práce vznikla za podpory vědeckých záměrů Ministerstva školství, kultury a mládeže č. MSM0021627502.

#### Literatura:

1. Elbergali A. K., Nygren J. a Kubista M., *An automated procedure to predict the number of components in spectroscopic data*, Anal. Chim. Acta, 1999
2. Meloun M., Čapek J., Mikšík P., Brereton R. G., *Critical comparison of methods predicting number of components in spectroscopic data*, Anal. Chim. Acta, 2000
3. Meloun M., Militký J., Hill M., *Počítačová analýza vícerozměrných dat v příkladech*, Academia, 2005, ISBN80-200-1335-0
4. Meloun M., Ferenčíková Z., Vrána A., *The thermodynamic dissociation constants of methotrexate by the nonlinear regression and factor analysis of multiwavelength spectrophotometric pH-titration data*, Cent. Eur. J. Chem., 2010