Název práce:

Faktorová analýza časových řad jako nástroj pro analýzu spekter z časově rozlišené laserem indukované spektroskopie

Autor: Jakub Višňák

Obor: Aplikovaná jaderná chemie

Druh práce: Diplomová práce

Vedoucí práce: Mgr. A. Vetešník, Ph.D.

ČVUT v Praze, FJFI, Katedra jaderné chemie

Konzultant: Doc. RNDr. Jiří Bok, CSc.,

Fyzikální ústav UK

Abstrakt: Práce se zabývá podrobným popisem některých metod analýzy dvou- a tříindexových datových polí pocházejících z (časově rozlišených) měření laserem indukované fluorscenční spektroskopie (TRLFS) (postačujícím pro snadnou konstrukci programu danou metodu vykonávajícího). V praktické části je pak nejkomplexnější a nejrobustnější z nich – FAČRSV validována na uměle vygenerovaných datech a následně jsou všechny v teoretické části uvedené metody aplikovány na experimentálně získaná data pocházející z měření několika TRLFS kinetických sérií série roztoků pro systém UO_2^{2+} - SO_4^{2-} - ClO_4^- - Na^+ - H_3O^+ (iontová síla I = 0,3, pH = 2, c(U(VI)) = 10^{-5} M) lišících se celkovou molární koncentrací S(VI). Získané parametry systému/jeho TRLFS spekter (speciační konstanty β_1, β_2, fluorescenční doby žiovota, fluorescenční spektra jednotlivých fluorescenčně aktivních a nezanedbatelně se v roztocích vyskytujících species – UO_2^{2+}, UO_2SO_4 a $UO_2(SO_4)_2^{2-}$) jsou analyzovány pro každou metodu zvlášť a následně srovnány v kapitole „Shrnutí“ navzájem pro různé metody.

Metody spojuje použití faktorové analýzy. V případě dvouindexových datových polí „vlnová délka-čas“ a „vlnová délka - koncentrace S(VI)“ se jedná o singulární rozklad matice, v případě tříindexového datového pole „vlnová délka – čas – koncentrace S(VI)“ o rozklad známý jako „Paralelní faktorová analýza“ (PARAFAC-3, nebo s předúpravou dat normalizací – NPARAFAC-3). Získané elementy matic popisujících závislost zastoupení spekter jednotlivých komponent signálu (v případě SVD označovaných jako subspektra) na zbývajících indexech (čas a/nebo koncentrace S(VI)) jsou následně fitovány pomocí metody Maximální věrohodnosti („Maximum Likelihood“) na vhodný matematický model průběhu signálu.

V případě dvouindexových datových polí (tj. datových matic) „vlnová délka-čas“ se jedná o metodu „Faktorové analýzy časových řad“ (FAČR, [1] [4]-[6]) a jako vhodný model časového průběhu v tomto případě uvažuji monoexponenciální pokles pro každou složku.

V případě dvouindexových datových polí „vlnová délka- koncentrace S(VI)“ se jedná o metodu „Faktorové analýzy spojené se speciačním výpočtem“ (FASV) [1], [4], [6], jako vhodný model průběhu signálu uvažuji zjednodušený speciační model komplexace UO_2^{2+} ionty SO_4^{2-}.

Konečně, v případě tříindexových polí „vlnová délka – čas – koncentrace S(VI)“ existuje více alternativ jak spektra analyzovat – FASV „dlouhých sloupců“, FAČRSV – Faktorová analýza časových řad spojená se speciačním výpočtem a (N)PARAFAC-3.

Klíčová slova: TRLFS, sulfáto- komplexy uranylu(2+), nelineární regrese, χ^2, faktorová analýza, analýza hlavních komponent, metoda maximální věrohodnosti, SVD dekompozice, Paralelní faktorová analýza (PARAFAC), speciační výpočty, chemometrie, SIT rovnice

Title:

The factor analysis of time series as a toll for the analysis of spectra from the time-resolved laser induced fluorescence spectroscopy

Author: Jakub Višňák

Abstract: Presented thesis is concerned with the description of several methods for the analysis of two- and there-way data arrays from the TRLFS measurements (sufficiently detailed for a straightforward computer implementation). In the experimental part, the most complex and robust one – „The factor analysis of the time series coupled with the speciation model“ (FACRSV) is validated on the simulated spectra. Also, all methods are then applied on the experimental data from the TRLFS kinetic series of the solution series for the system UO_2^{2+} - SO_4^{2-} - ClO_4^- - Na^+ - H_3O^+ (ionic strength I = 0,3, pH = 2, c(U(VI)) = 10^{-5} M) with various S(VI) molar concentration. Obtained parameters of the system and its TRLFS spectra (speciation constants β_1, β_2, fluorescence life-times, fluorescence spectra of the different fluorescence active and nonnegligibly present species - UO_2^{2+}, UO_2SO_4 a $UO_2(SO_4)_2^{2-}$) are analýzed for every method and then compared across the methods.

All presented methods are based on the factor analysis. In the case of two-way data arrays „wave length-time“ and „wave length - S(VI) concentration“ , the method leads to the singular value decomposition, in the case of the three-way data array „wave length – time – S(VI) concentration“, the method leads to the Parallel factor analysis (PARAFAC-3, with preprocessing by normalization – NPARAFAC-3). Obtained loadings, which describes the dependence of the signál components (in the case of the SVD called „subspectra“ and having orthonormal property) on the time and/or concentration indices, are then fitted by the Maximum-Likelihood method on a suitable mathematic model of the signál.

In the case of the two-way data arrays „wave length-time“, the method led us to the „Factor analysis of the time series“ method (FACR) [1], [4]-[6] which uses for each specie the monoexponential decay as a suitable model of the time dependence of fluorescence intensity.

In the case of the two-way data arrays “wave length-S(VI) concentration”, the method led us to the “Factor analysis coupled with speciation model” (FASV) [1], [4], [6] as a suitable model of signál dependence, the most simplified speciation model of the complexation of UO_2^{2+} by SO_4^{2-} was used.

Finally, in the case of the three-way data arrays “wave length – time – concentration S(VI)” there are several methods how spectra can be analysed - “Factor analysis of time-resolved spectra coupled with speciation model” (FASVD), “Factor analysis of time series coupled with speciation model” (FACRSV) and Parallel factor analysis (PARAFAC-3 or NPARAFAC-3).

Key words: TRLFS, sulfate- complexes of uranyl(2+), nonlinear regression, χ^2, factor analysis, Maximum-likelihood method, principal component analysis, SVD decomposition, Parallel factor analysis (PARAFAC), speciation calculations, chemometry, SIT equation