

Factor analysis program INDICES for prediction of the number of components in spectroscopic data

Milan MELOUN

*Department of Analytical Chemistry, Faculty of Chemical Technology,
University Pardubice, CZ-532 10 Pardubice, Czech Republic,
milan.meloun@upce.cz*

Determining the number of chemical components in a mixture is the first important step for further qualitative and quantitative analysis in all forms of spectral data treatment. Accuracy of 13 various statistical indices methods for estimation of the number of components that contribute to spectra was critically tested on simulated and on experimental data sets using algorithm INDICES in S-Plus software. All indices methods are classified into two categories, the precise methods based upon a knowledge of the instrumental error of the absorbance data, $s_{\text{inst}}(A)$, and the approximate methods requiring no such knowledge. Most indices always predict the correct number of components even a presence of the minor one when the signal-to-error ratio SE_R is higher than 10 but in case of RESO and IND higher than 6. On base of SE_R the detection limit of every index method is estimated. Two indices, RESO and IND, correctly predict a minor component in a mixture even its relative concentration is about 0.5 - 1% and solve an ill-defined problem with severe collinearity in spectra. Wernimont-Kankare procedure performs reliable determination of the instrumental standard deviation of spectrophotometer used. In case of real experimental data the RESO, IND and indices methods based on knowledge of instrumental error should be preferred. To investigate all statistical properties of absorbance data matrix which were designed to be quite similar to real experimental data and cover some typical situations of analytical practice, several data sets of absorption spectra were simulated for a three-components system in mixture: potassium bichromate, cobalt(II) sulphate and copper(II) sulphate, a mixture abbreviated {Cr-Co-Cu}. An absorbance matrix was created by multiplying absorptivity spectra of three components by their simulated concentration profiles to reach resulting absorbance. Each matrix data set contains digitized spectra consisted of digitized wavelengths. Random noise was added to the spectra by generating random numbers with a Gaussian distribution with mean 0 and standard deviation equal to the pre-selected noise level, $s_{\text{inst}}(A)$, to reach an optioned SE_R value.

Acknowledgments:

The financial support of the Grant Agency IGA MZ ČR (No NS9831-4/2008) and of the Czech Ministry of Education (Grant No MSM0021627502) is gratefully acknowledged.

References:

- [1] Meloun M., Čapek J., Mikšík P., Brereton R. G., Critical comparison of methods predicting the number of components in spectroscopic data *Anal. Chim. Acta* **2000**, Vol. 423, 51 - 68.

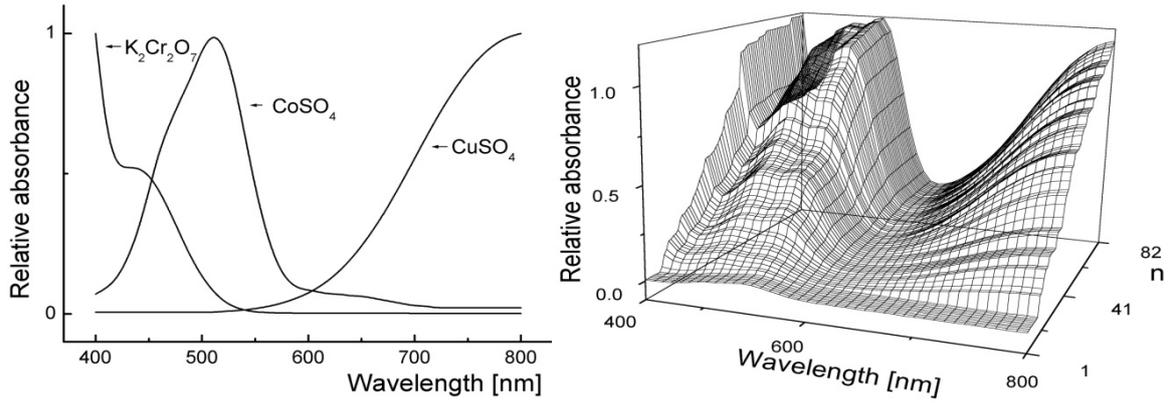


Fig. 1: (a) Spectra of relative absorbance for three components in a simulated three-components system in mixture, potassium bichromate - cobalt(II) sulphate - copper(II) sulphate, with $r = 3$, $n = 82$, $m = 41$ and $SER = 1570$, S-Plus, (b) 3D-relative absorption spectra forms the input of INDICES.

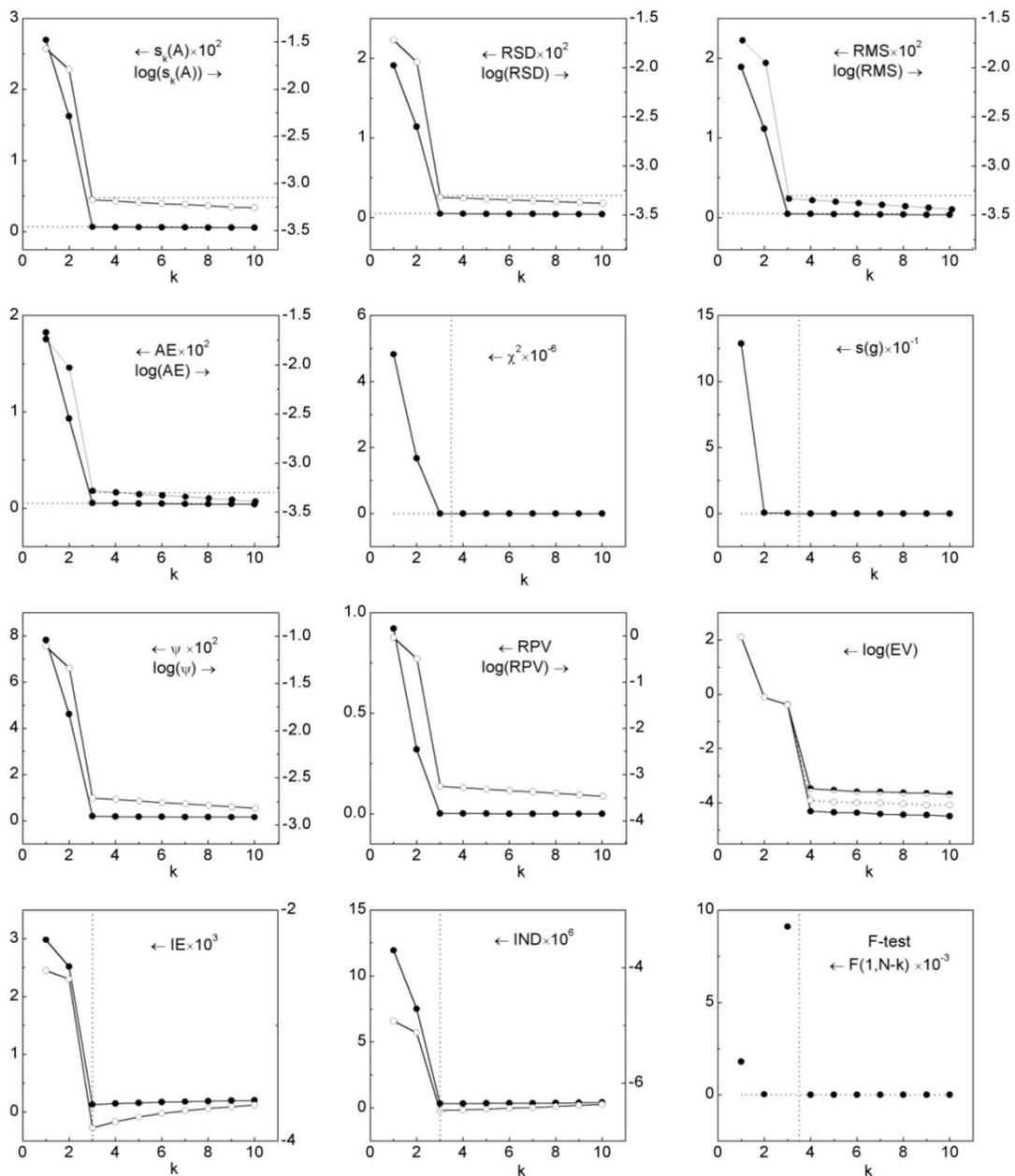


Fig. 2 The output of INDICES program: the indices (full circles) and logarithm of the indices (empty circles) of 13 methods as a function of the number of principal components k for a data of Fig. 1.