Talanta 117 (2013) 209–220

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Analysis of dilute aqueous multifluorophoric mixtures using excitation–emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy: A comparative evaluation

Keshav Kumar, Ashok Kumar Mishra *

Department of Chemistry, Indian Institute of Technology-Madras, Chennai-600036, India

ARTICLE INFO

Article history:
Received 20 July 2013
Received in revised form
31 August 2013
Accepted 2 September 2013
Available online 11 September 2013

Keywords:
EEMF
TSF
Rayleigh scattering
Raman scattering
Trilinear
Bilinear
PARAFAC
N-PLS
MCR-ALS

ABSTRACT

Excitation–emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy are two conceptually different fluorescence techniques that have been used to map the fluorescence responses of the fluorophores present in a multifluorophoric mixture. EEMF was introduced four decades back and most of the fluorimeters have the suitable computer program which allows the acquisition EEMF spectra. Recently introduced TSF spectroscopy has been shown to possess good application potential in analytical fluorimetry and has started attracting the attention of analytical chemists. TSF data structure, however, is intrinsically different from EEMF data structure and a better understanding of TSF data structure is crucial to utilising its application potential. In the present work, a comprehensive comparative study between EEMF and TSF spectroscopic data set was performed by taking aqueous mixtures containing low concentrations of benzo[a]pyrene, chrysene, and pyrene as test case. The EEMF and TSF data structures were clearly explained by taking pyrene as an example. The effects of Rayleigh and Raman scattering on the quality of EEMF and TSF data sets were studied. EEMF and TSF data sets of dilute aqueous mixtures of benzo[a]pyrene, chrysene, and pyrene were subjected to three chemometric techniques PARAFAC, N-PLS, and MCR-ALS analysis. TSF data set in particular was found to be highly attuned to MCR-ALS analysis. Obtained results of chemometric analyses on EEMF and TSF data sets show that TSF data of dilute aqueous mixtures provides more accurate spectral and concentration information than EEMF data sets. Therefore, TSF spectroscopy could be considered as an alternate to the EEMF for the analyses of dilute multifluorophoric mixtures.

1. Introduction

Excitation–emission matrix fluorescence (EEMF), introduced by Christian and co-worker [1–5], and total synchronous fluorescence (TSF), introduced by Patra and Mishra [6], are the two conceptually different fluorescence techniques which have been used for the analysis of multifluorophoric systems [6–28]. Both EEMF and TSF contain the fluorescence signatures of all the fluorophores of a multifluorophoric system in a single three dimensional spectrum. However, spectral characteristics of a fluorophore in EEMF and TSF modes are different from each other; it is due to the difference in the data acquisition process in EEMF and TSF spectroscopy.

Christian and co-worker introduced computerised method for displaying EEMF spectrum in a single three dimensional graphical format [1–5]. Rho and Stuart introduced a non-computerised method for plotting EEMF contour plot [29]. An EEMF spectrum essentially consists of number of excitation and emission spectra of the fluorophores collected at various emission and excitation wavelengths, respectively. An emission spectrum is acquired by keeping the excitation monochromator at a fixed wavelength and scanning the emission monochromator. Similarly, excitation spectrum is acquired by scanning the excitation monochromator and keeping the emission monochromator at a fixed wavelength. On the other hand, a TSF spectrum consists of number of synchronous fluorescence (SF) [30,31] spectra of fluorophores collected over a range of excitation wavelength at various wavelength offsets ($\Delta\lambda$s) [6]. A SF spectrum is collected by simultaneously varying the excitation and emission monochromator with a constant $\Delta\lambda$ over a given excitation wavelength range [30,31]. EEMF spectrum is generally created by plotting the emission wavelength, excitation wavelength, and fluorescence intensity along *X*, *Y*, and *Z*-axes, respectively [5,29]. TSF spectrum is created by plotting the excitation wavelength along *X*-axis, $\Delta\lambda$ along *Y*-axis, and SF intensity along *Z*-axis [6,32]. Provided, Beer–Lambert law [33] is obeyed i.e. primary and secondary inner filter effects [34] are absent and spectral intensities are additive, then fluorescence intensities of a fluorophore in

* Corresponding author. Tel.: +91 44 22574207; fax: +91 44 22574202.
E-mail address: mishra@iitm.ac.in (A.K. Mishra).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.09.002

EEMF and TSF are given by Eqs. (1) and (2), respectively [35],

$$I_{EEMF}(c, \lambda_{ex}, \lambda_{em}) = KcdE_X(\lambda_{ex})E_M(\lambda_{em}) \quad (1)$$

$$I_{TSF}(c, \lambda_{ex}, \Delta\lambda) = KcdE_X(\lambda_{ex})E_M(\lambda_{ex} + \Delta\lambda) \quad (2)$$

where I_{EEMF} and I_{TSF} are the fluorescence intensity in EEMF and TSF mode, respectively, K is a constant accounting for instrumental factor, c is concentration of the fluorophore, d is the path length, E_X and E_M are the excitation and emission profile, respectively, λ_{ex} and λ_{em} are the excitation and emission wavelength, respectively, and $\Delta\lambda = \lambda_{em} - \lambda_{ex}$.

In EEMF, each fluorophore of a multifluorophoric mixture has one excitation and one emission spectral profile and shape of which is invariant to the changes in the λ_{em} and λ_{ex}, respectively. Thus, in Eq. (1), $E_X(\lambda_{ex})$ for the spectral profile along the emission mode (i.e. emission spectrum) and $E_M(\lambda_{em})$ for the spectral profile along the excitation mode (i.e. excitation spectrum) acts as intensity scaling factor. In TSF spectrum, both shape and intensity of the spectral profile (i.e. SF signal) of a fluorophore changes with $\Delta\lambda$; and Eq. (2) clearly shows that this variation in the shape arises because each SF spectral profile in TSF is a consequence of the product of two functions $E_X(\lambda_{ex})$ and $E_M(\lambda_{ex}+\Delta\lambda)$ which changes continuously and simultaneously over a given range of λ_{ex}. Product of two simultaneously changing functions also make spectral features in TSF mode much more resolved than in EEMF mode. As it was discussed above, shape of the SF spectral profile of a fluorophore changes with $\Delta\lambda$, thus, TSF profile are more distinctive than EEMF profile.

Both EEMF and TSF provide three way data arrays of dimensions, sample × excitation × emission, and sample × excitation × $\Delta\lambda$, respectively. EEMF and TSF based three way arrays are shown graphically in Fig. 1; first mode (sample mode) contains the relative concentration profile of the components, second (excitation) and third (emission or $\Delta\lambda$) modes contain the spectral profiles. Three way spectral data sets are classified as trilinear, provided, for each fluorophores of the mixture shape of the profile along one spectral mode is invariant to the changes in other spectral mode [35–38]. Thus, EEMF based three way array has the trilinear structure and TSF based three way array lack the trilinearity. It should be noted that the trilinear structure of the data set holds only when the Beer–Lambert law is valid.

In the last two decades, chemometrics has evolved as a potential tool for the fast and efficient analyses of the large volume of data sets obtained from various spectroscopic and chromatographic

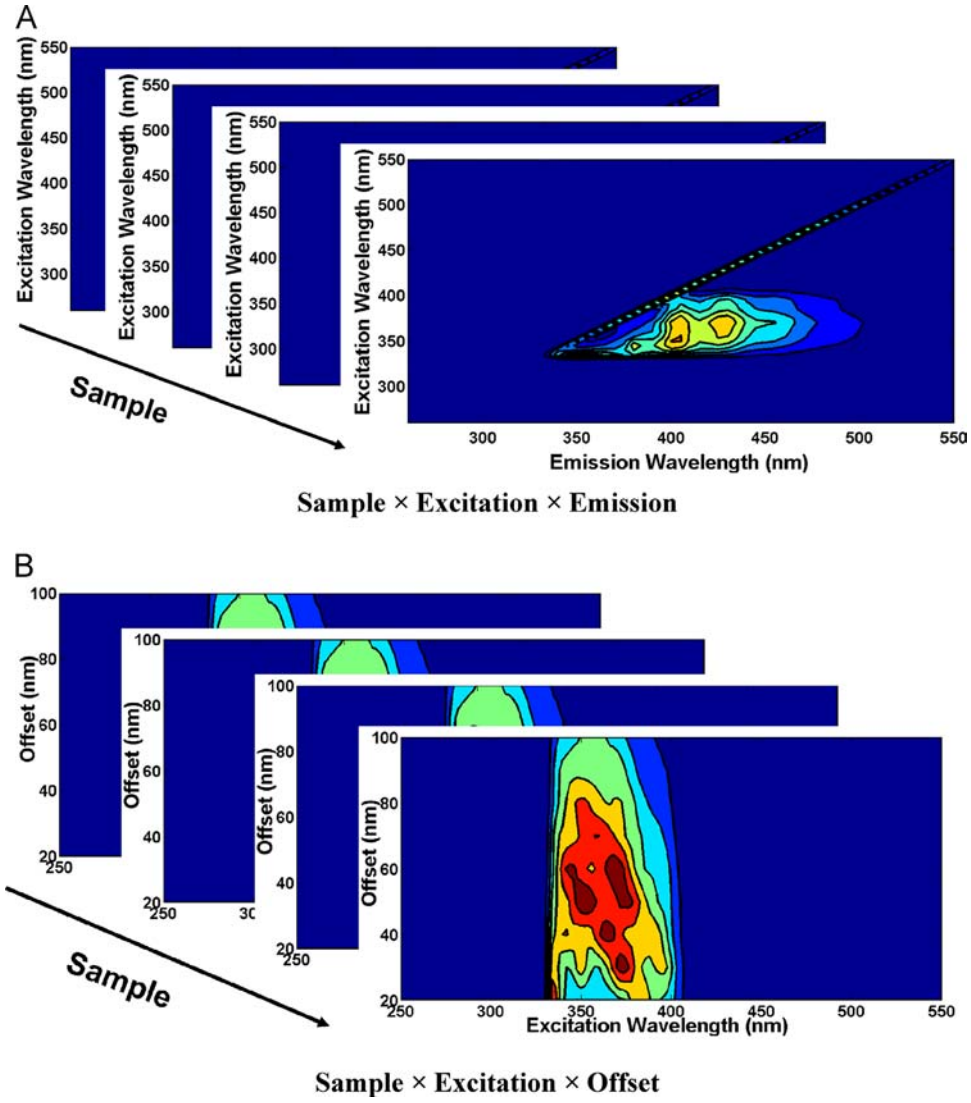

Fig. 1. Graphical representation of [A] EEMF and [B] TSF based three way arrays of dimension sample × excitation × emission and sample × excitation × offset, respectively.

techniques [15–20,39–47]. Chemometric techniques involve simultaneous analysis of more than one variable at a time; thus, they are also known as multivariate analysis techniques. Essentially, these techniques involve mathematical deconvolution of the data sets and generate the contribution (i.e. concentration) and profiles along other modes (i.e. spectral or elution profiles etc.) for each component of the mixture [48–50]. Principal component analysis (PCA) [48–51], parallel factor (PARAFAC) analysis [35,38,49,50,52–54], N-way partial least square (N-PLS) analysis [49,50,55], and multivariate curve-resolution alternating least square (MCR-ALS) analysis [37,49,56–59] are some of the commonly used chemometric techniques for the analyses of multi way-arrays. PARAFAC is used for the analysis of three way data sets having trilinear structure [35–38,50]. N-PLS is another technique which used for the analyses of three way arrays [49,50,55]; N-PLS is known to work well even if there are slight deviations from the trilinearity in the data sets [60,61]. PCA and MCR-ALS methods are used for the analysis of two dimensional bilinear data sets (i.e. every component has one spectral profile and spectral intensity is directly proportional to the concentration) [48–51,56–59]. The two dimensional data sets from three or higher way arrays are obtained by the data unfolding procedure [50].

The existences of bilinear and trilinear structures in EEMF data sets have successfully attuned EEMF to the chemometric techniques [35–38,50]. Combinations of EEMF and chemometric techniques have been used for the analyses of environmental [19,62–65], pharmaceutical [66–68], and petroleum [39,69,70] samples. On the other hand, chemometric analyses of TSF data sets or TSF itself have not become as popular as EEMF. It is mainly due to three reasons (i) very recent introduction, (ii) most of the present fluorimeter do not have the proper computer programs which would enable the direct acquisition of fluorescence data in TSF mode, and (iii) structure of TSF data set is not clearly understood.

As it was discussed above, TSF based three way arrays, sample × excitation × $\Delta\lambda$, do not have the trilinear structure; thus, trilinear decomposition of TSF data set is not possible. However, in our recent work it was shown that trilinear decomposition of the TSF data sets can be achieved, provided, TSF data is represented in EEMF like layout i.e. sample × excitation × (excitation + $\Delta\lambda$) [35]. In another recent work, we have shown that by unfolding the TSF three way arrays along the first mode (i.e. sample) a bilinear data set of dimension sample × (excitation × $\Delta\lambda$) can be generated [37]. The unfolded-TSF data set can be easily analysed with bilinear models such as PCA and MCR-ALS analysis [37]. Thus, by suitably arraigning the TSF data sets it is possible to apply various chemometric techniques which are typically used for the analysis of EEMF data sets.

The present work is intended to compare the quality of the EEMF and TSF spectroscopic data sets in qualitative and quantitative manner. To the best of our knowledge this is the first work where a comprehensive comparative study between EEMF and TSF spectroscopy was carried out. For qualitative comparison, spectral characteristics of the real EEMF and TSF data sets collected for dilute aqueous multifluorophoric mixtures consisting of benzo[a] pyrene, chrysene, and pyrene were analysed. For quantitative comparison the EEMF and TSF data set of dilute multifluorophoric mixtures consisting of these three fluorophores were subjected to three chemometric techniques PARAFAC, N-PLS, and MCR-ALS analysis. These three chemometric methods have different properties and they model the data in different ways. Selected three fluorophores have significant spectral overlap which would make analysis nontrivial. In addition, these three fluorophores contain vibrational fine structures in their fluorescence spectra which would also make curve-resolution relatively difficult. In the case of mixtures containing all the known fluorophores, we have a priori information about the number of components required to perform the chemometric analyses on EEMF and TSF data sets. We also have true EEMF and TSF spectral and concentration profile of each fluorophores, thus, with respect to every component of the mixture it is possible to compare the outcomes of chemometric analyses on TSF and EEMF data sets with their actual values. It would also allow us to evaluate the ability of different chemometric methods to extract information (i.e. concentration and spectral information of each fluorophore) from EEMF and TSF data sets of dilute multifluorophoric mixtures.

2. Materials and methods

2.1. Chemicals and sample preparation

Three fluorophores benzo[a]pyrene (BaP), chrysene (CY), and pyrene (PY) were purchased from Sigma-Aldrich. BaP, CY, and PY have 97%, 98%, and 98% purity, respectively and were used as such without any further purification. All the three components were found to have no fluorescent impurities. Separate stock solutions were prepared for BaP, CY, and PY in analytical grade acetone. Further dilutions were also done with the acetone. Different amounts of BaP, CY, and PY were taken from the respective dilute stock solutions in 29 different vials of 10 ml volume. Acetone was removed completely and 5 ml of triple distilled water was added to the residue to make the working solutions. Amounts (ng/ml) of BaP, CY, and PY in 29 samples are summarised in Table 1. Out of these 29 samples, first 23 samples were used to make the calibration set and last 6 samples were used to make the validation set. Deliberately, in many samples of the calibration and validation set, concentration of a particular fluorophore was kept constant and the concentrations of other two fluorophores were varied. Such variations in the concentration are expected to make

Table 1
Amounts of BaP, CY, and PY in the BaP–CY–PY mixtures; sample numbered 1–23 were used for calibration set and sample numbered 24–29 were used for validation set.

Sample	BaP (ng/ml)	CY (ng/ml)	PY (ng/ml)
1	3.16	0.00	3.77
2	3.16	1.69	1.88
3	2.08	1.69	3.77
4	2.08	3.39	1.88
5	3.16	1.02	2.64
6	1.25	2.37	3.77
7	2.91	3.39	1.13
8	0.00	0.00	7.54
9	1.04	0.85	5.65
10	1.04	4.08	0.94
11	4.98	0.42	0.47
12	0.52	0.42	6.60
13	5.32	0.00	0.00
14	3.16	3.39	3.77
15	0.00	3.39	3.77
16	3.16	3.39	0.00
17	2.08	3.39	3.77
18	4.24	0.85	0.94
19	0.52	4.93	0.47
20	2.08	1.69	1.88
21	0.00	5.78	0.00
22	3.16	3.39	1.88
23	3.16	1.69	3.77
24	4.16	4.07	2.22
25	4.99	2.03	3.70
26	2.47	3.39	4.44
27	2.91	3.39	2.96
28	4.16	2.71	2.59
29	3.33	2.37	3.70

the chemometric analyses nontrivial, particularly when the fluorophores have significant spectral overlap. It would actually allow us to test the quality of EEMF and TSF data sets. It is to be noted that concentrations of the fluorophores were kept low (of the order of ng/ml) so that primary and secondary inner filter effect [34] or any other energy degradation process which may cause the non-linearity between the fluorescence intensity and the concentration of the fluorophore could be avoided.

2.2. Instrument and data acquisition

Fluorescence experiments were carried out using Fluoromax-4 (Horiba Jobin Yvon) spectrofluorometer having Xenon lamp of 150 W as excitation source. Various instrumental parameters used for the EEMF and TSF data acquisition for the BaP–CY–PY mixtures are summarised in Table 2.

2.3. Selection of optimum numbers of factors

Optimum numbers of factors required for the PARAFAC, MCR-ALS, and N-PLS analyses were selected by subjecting the data sets to the principal component analysis (PCA) where internal validation was carried out using leave-one-out cross validation (LOOCV) approach [37,49]. Number of factors giving minimum root mean square error of cross validation (RMSECV) [48,49] values in the PCA was taken as the rank of the data set. In an ideal case, mathematical rank should be equal to the chemical rank i.e. number of fluorophores present in the multifluorophoric mixtures. In addition to PCA, core consistency values [17,35,50,71] were also analysed for finding the optimum number of factors required to perform the PARAFAC analysis. The core consistency values for the PARAFAC model with different factors are calculated using Eq. (3)

$$\text{Core consistency} = \left(1 - \frac{\sum_{d=1}^{F}\sum_{e=1}^{F}\sum_{f=1}^{F}(g_{\text{def}} - t_{\text{def}})^2}{F}\right) \times 100 \quad (3)$$

where g_{def} and t_{def} are the elements of G (calculated) and T (theoretical) core arrays, respectively, and F is the number of factors. T contains one along all the superdiagonal positions and zero at off superdiagonal positions. In ideal case G should resemble T i.e. core consistency is equal to 100%. Core consistency value less than 70% indicates over fitting the data [17,35].

2.4. Software used

PARAFAC, N-PLS, and MCR-ALS analyses were carried out using PLS_Toolbox 5.0.3 written in MATLAB language. Data were plotted using Origin 8 and MATLAB (R2008b) softwares. The codes (i) for creating the three way arrays for TSF and EEMF data, (ii) for representing the TSF in EEMF layout, and (iii) for extracting the TSF data from EEMF layouts were written in MATLAB language. These codes are given in Supplementary material.

Table 2
Instrumental parameter for EEMF and TSF spectroscopy of BaP–CY–PY mixtures.

Instrumental parameter	
Excitation/emission band-pass	7/7 nm
Excitation wavelength for EEMF	250–500 nm, step size of 10 nm
Emission wavelength for EEMF	250–550 nm, step size of 2 nm
Excitation wavelength for TSF	250–550 nm, step size of 2 nm
Offsets ($\Delta\lambda$) for TSF	20–100 nm, step size of 10 nm
Integration time	0.01 (s)

3. Results and discussion

3.1. Qualitative comparison of EEMF and TSF spectroscopy characteristic of BaP–CY–PY mixture

EEMF spectra of dilute aqueous solutions of BaP, CY, and PY are shown in Fig. 2(A); it mainly contains diagonal bands of first order Rayleigh scattering. First and second order Rayleigh scattering appears at λ_{ex} equal to λ_{em} and $2 \times \lambda_{em}$, respectively. Fluorescence responses of none of the three fluorophores were observed in the EEMF spectrum. Fluorescence of BaP, CY, and PY were masked due to the highly intense Rayleigh scattering lines. Thus, at the instrumental level it is difficult to observe the fluorescence of the dilute mixtures of fluorophores in EEMF mode. It is to be noted that the selected three fluorophores belong to the class of molecules having highly rigid molecular framework. The rigidity

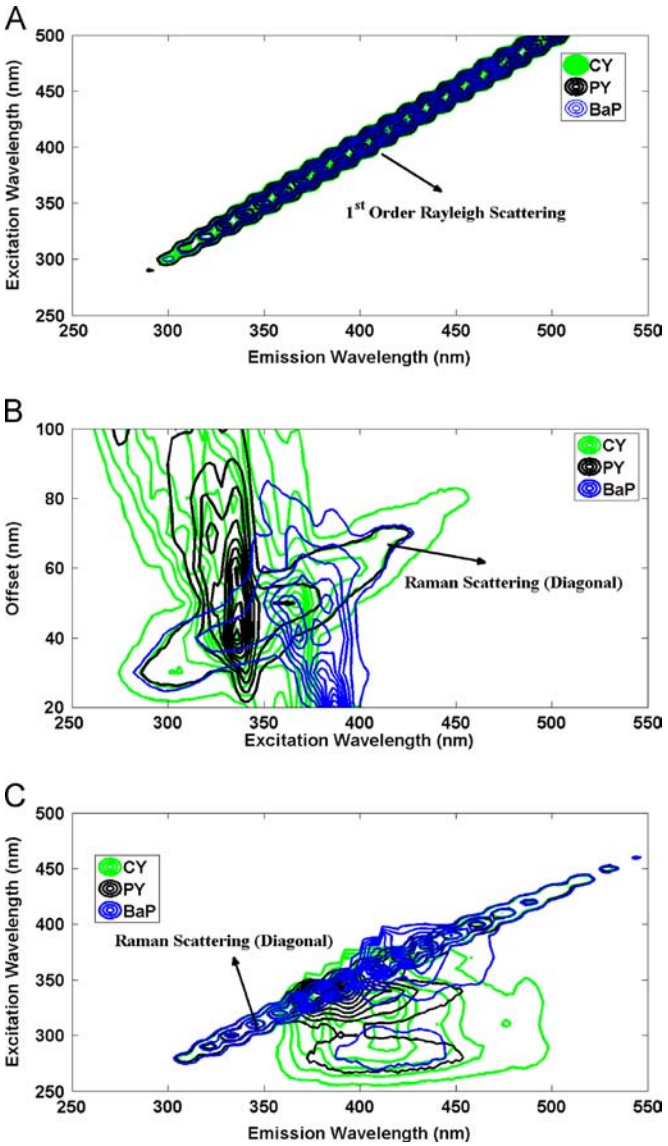

Fig. 2. [A] EEMF spectra, [B] TSF spectra, and [C] Rayleigh removed EEMF spectra of BaP, CY, and PY.

in the molecular framework makes these molecules highly fluorescent even at low concentrations compare to other fluorophores such as amino acids, humic acids, fulvic acids, and porphyrins etc. found in water or other environmental samples. In order to see the fluorescence signal of these fluorophores BaP, CY, and PY, it is essential that Rayleigh scattering is eliminated. It is usually done after collecting the data by any of these three methods (i) subtracting the EEMF of blank [72–74], (ii) filling the zero or missing values in the EEMF regions where $\lambda_{em} \leq \lambda_{ex}$ and $\lambda_{em} \geq 2 \times \lambda_{ex}$, [38,75–78] and (iii) interpolation method [79]. All these methods have their own limitations in eliminating the Rayleigh scattering profile from the EEMF spectrum. For example, availability of a blank is often a problem with the real life sample. Using zero or NaN or interpolation method also do not work successfully in all the cases, because in many systems the scattering lines do appear very near or on top of the fluorescence signal of some of the fluorophores of the multifluorophoric mixture. Thus, none of the available methods can be used to eliminate the Rayleigh scattering completely from the samples obtained from different sources. In other words, there is no single foolproof remedy for the Rayleigh scattering.

On the other hand, wavelength offset ($\Delta\lambda$) a purely instrumental parameter allows removal of both first and second order Rayleigh scattering signals from the TSF spectrum right at the data acquisition stage. It is done by collecting the fluorescence data with suitable $\Delta\lambda$s so that λ_{em} is not equal to λ_{ex} and $2 \times \lambda_{ex}$. TSF spectra of the same dilute aqueous solutions of BaP, CY, and PY are shown in Fig. 2(B), where fluorescence of the fluorophores can be easily observed. It shows by collecting the fluorescence data in TSF way, it is possible to observe the fluorescence of the fluorophores even at low concentrations (ng ml^{-1}) right at the instrumental level. No further steps are needed to eliminate the Rayleigh scattering.

Raman scattering, another redundant signal which do not contain any fluorescence information, appear diagonally in both EEMF and TSF which can also be seen in the TSF and EEMF (Rayleigh eliminated) contour plots of the BaP, CY, and PY, shown in Fig. 2(B) and (C), respectively. Raman scattering lines are often deeply buried in the fluorescence regions. However, the extent to which the Raman scattering signals are mixed with the fluorescence can differ in EEMF and TSF spectroscopy. For example, it can be seen from Fig. 2(B) that Raman scattering due to solvent (water) and fluorescence signals of the BaP are well separated in TSF spectra; whereas, in EEMF spectra (Fig. 2(C)) there is an extensive overlap between the Raman signals of the solvent and the fluorescence signals of the BaP. Elimination or minimisation of Raman scattering from the fluorescence data is not easy at the instrumental level in any of the two techniques. The only method which can be easily and efficiently employed to minimise the Raman scattering is the blank subtraction. Though, it is important that some alternate methods (e.g. mathematical ways) should be discovered to handle the Raman scattering in both EEMF and TSF spectroscopy.

Another point which needs to be highlighted here is that TSF data points are relatively more diverse and rich in capturing the fluorescence information. It is so because in TSF mode both excitation and emissions wavelength changes simultaneously for a given SF spectrum at a particular $\Delta\lambda$; thus, each data point contain both excitation and emission information of a fluorophore. Hence, even with small volume of TSF data sets it is possible to have complete fluorescence information of a fluorophore along emission as well as excitation modes. Whereas for the EEMF data points, only one wavelength either excitation or emission changes continuously while the other wavelength remains unchanged depending on the mode along which data are analysed. Therefore, in order to have complete fluorescence information, it is necessary that one should have sufficient data points along both the modes.

Above discussion illustrate that TSF mode makes possible even for a "novice" to see the fluorescence response of the fluorophores even at low concentrations at the data acquisition stage itself without involving any data processing step.

Structure of the EEMF and TSF data sets were briefly discussed in the introduction part of the current work. However, in the context of the present study it is necessary that the structures of the EEMF and TSF data sets are clearly explained taking a real example. For this purpose, emission and excitation spectra of PY (aqueous solution) at various excitation and emission wavelengths, respectively, are shown in Fig. 3(A and B). It can be observed that shapes of the emission and excitation spectral profiles are invariant to excitation and emission wavelengths, respectively. The independence of the two spectral modes gives EEMF based three way arrays trilinear property. On the other hand, Fig. 3(C) shows that shape of the SF profile along a given excitation wavelength range changes with the $\Delta\lambda$s; the lack of independence between the two spectral modes i.e. excitation and $\Delta\lambda$s, make the TSF data non-trilinear in nature.

A bilinear unfolded-EEMF data set can be obtained by unfolding the EEMF three way arrays (i.e. sample × excitation × emission) along the sample mode. Similarly, TSF based three way array sample × excitation × $\Delta\lambda$ can be unfolded along the sample mode to generate the bilinear data set of dimension sample × (excitation × $\Delta\lambda$). It is easy to realise that the unfolded-EEMF and unfolded-TSF data sets have the bilinear character i.e. spectral profile of each fluorophores remains the same in every sample; however, they are weighted to different extent by the concentration. It indicates a three way array such as TSF may lack trilinearity but it can still have the linearity at the lower dimensions, provided it is suitably arranged.

In the present work, number of variables in the EEMF and TSF were set to missing i.e. NaN. These missing values were further handled using expectation maximisation (EM) technique [78,80,81].

3.2. Chemometric analyses on EEMF and TSF data of BaP–CY–PY mixtures

EEMF and TSF data sets of dilute aqueous mixtures of BaP–CY–PY were subjected to the PARAFAC, N-PLS, and MCR-ALS analyses and their outcomes were analysed to compare the quality of the EEMF and TSF data sets; conversely, ability of PARAFAC, N-PLS and MCR-ALS techniques to extract information from EEMF and TSF data sets were evaluated.

Before subjecting the data sets to the chemometric methods blank EEMF and blank TSF data were subtracted from EEMF and TSF data sets, respectively from all the samples to minimise the scattering (i.e. non-trilinear part) and any other noise present in the data sets. Blank subtracted EEMF and TSF data sets of dimensions were unfolded along the first mode i.e. sample mode to obtain two way arrays of dimension 23×3926 and 23×1359, respectively. The data were mean centred and subjected to PCA for finding the optimum number of factors required to perform the chemometric analyses. RMSECV versus number of factors were plotted for PCA models of EEMF data sets and are shown in Fig. 4(A). RMSECV value was found to be minimum for one component PCA model. It can be seen, there is mismatch between the actual and predicted rank of EEMF data sets. However, we have a priori knowledge of the BaP–CY–PY system, and it is known that there are three fluorophores in the mixture. Thus, PARAFAC, N-PLS, and MCR-ALS analyses of EEMF data sets were performed with three factors. On the other hand, PCA result of TSF data sets, given in Fig. 4(B), shows RMSECV is minimum for a three component PCA model. Mathematical rank correlates well with the chemical rank of the system; therefore, there is no ambiguity in the selection of number of components. All the subsequent chemometric analyses of TSF data were performed

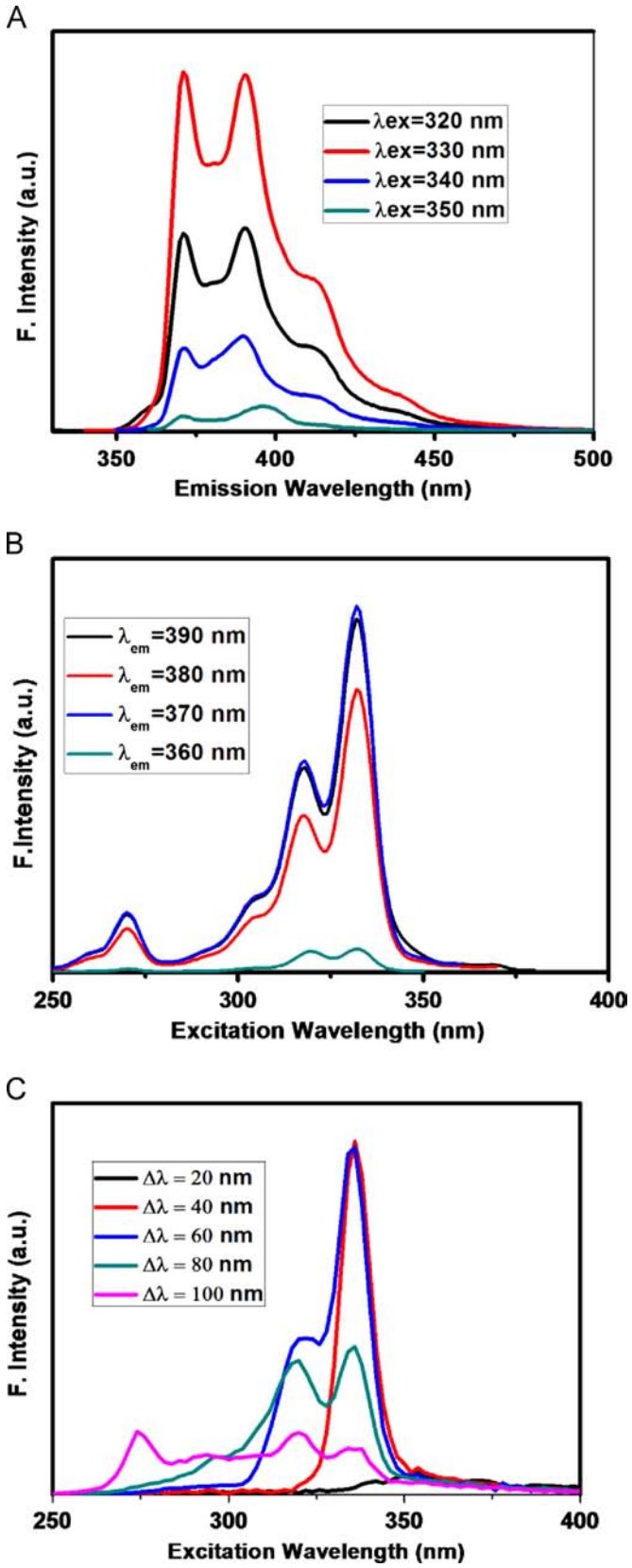

Fig. 3. [A] Emission, [B] excitation and [C] SF spectra of pyrene collected at various excitation, emission, and offsets, respectively.

with the three factors. However, results of core consistency analysis were also studied while selecting the factors for the PARAFAC analyses of TSF and EEMF data sets.

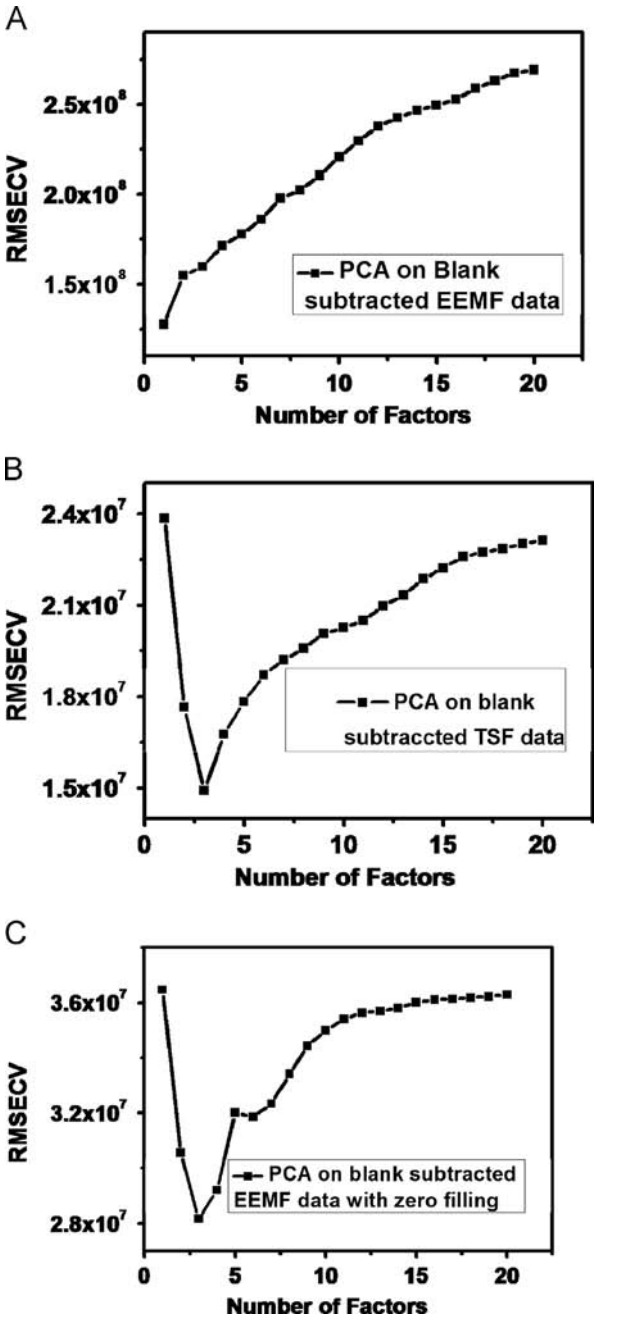

Fig. 4. [A] PCA on blank subtracted EEMF data set, [B] PCA on blank subtracted TSF data set, and [C] PCA on blank subtracted EEMF data with zeros in the wavelength range of $\lambda_{em} \leq (\lambda_{ex}+20)$ and $\lambda_{em} \geq (2 \times \lambda_{ex}-20)$.

3.2.1. *PARAFAC analysis of EEMF and TSF data sets of BaP–CY–PY mixtures*

3.2.1.1. PARAFAC analysis of EEMF data set. Blank subtracted EEMF data set of dimension $23 \times 26 \times 151$(sample × excitation × emission) was subjected to the PARAFAC analysis with the non-negative constraints on all the three modes. Non-negative constraints were used because neither spectral intensity nor concentrations of the fluorophores can be negative, moreover such constraints helps in arriving at the unique solution. Core consistency values of the PARAFAC models of 1–5 factors were found to be $\geq$90%. With addition of an extra factor core consistency value of PARAFAC model dropped to zero. Since there are only three fluorophores in the mixtures, PARAFAC analyses of EEMF data were performed with three factors i.e. one for each fluorophore. It is to be noted that

neither PCA nor core consistency analysis gave the correct number of factor required for the PARAFAC analysis. Nevertheless, three factor PARAFAC model was made which could explain only 33.63% variance of the spectral data set X. Different pre-processing techniques were also used along various modes of the data set to see if the model can capture more information. However, 33.66% was the highest variance which PARAFAC model could explain. Contribution of first (F1), second (F2), and third (F3) factor towards explaining the variance of the data were 12.21%, 10.99%, and 10.46%. The spectral loading vectors along second and third modes were multiplied together for all the three factors to generate their EEMF contour plots which are shown in Fig. 5. As it can be seen there is no resemblance between the resolved EEMF profiles of F1, F2, and F3 with BaP or CY or PY. In order to see if the relative concentration profiles along the first mode can be used for the quantification purpose, the score matrix (i.e. concentration matrix) containing the relative concentration profile of the three factors were taken as the independent data set and the actual concentrations of the BaP, CY, and PY were taken as the dependent data set for the principal component regression (PCR) analysis. A full rank PCR model of three factors was made. Linear regression equations for the actual and predicted concentrations of BaP, CY, and PY of the PCR model are given in Table 3. The slope of the regression equation for BaP, CY, and PY are very small, it shows that there are significant errors in the predictions of the concentration of the fluorophores. In an ideal case slope should be close to one for each component. The PARAFAC outcomes along all the three modes are different from the actual profiles i.e. true concentration profiles and EEMF spectra.

Above results show that PARAFAC could not model the EEMF data appropriately even though EEMF data of each sample were subtracted with the blank EEMF to minimise the scattering effects or in mathematical language elimination or minimisation of the non-trilinear part of the EEMF data sets.

In order to further minimise the non-trilinear part of the blank subtracted EEMF data sets; the variables in the wavelength range where $\lambda_{em} \leq (\lambda_{ex}+20)$ and $\lambda_{em} \geq (2\times\lambda_{ex}-20)$ were set to missing i.e. NaN. The extra 20 nm on the right side of first order Rayleigh scattering line and left side of second order Rayleigh scattering lines were also kept to missing so that scattering could be completely eliminated. However, it may cause the loss of fluorescence information. EEMF data sets of dimension $23\times26\times151$ with missing values were further subjected to PARAFAC analysis with non-negative constraints. Core consistency values of the PARAFAC models of 1–7 factors were calculated and were found to be greater than 85%. However, PARAFAC model of three factors having core consistency of 100% were preferred knowing that there three fluorophores are present in the mixtures. It is also to be noted that PCA of unfolded-EEMF data sets with missing values did not converge with any number of factors. PARAFAC model of three factors explained 69.79% variance of the data set. Thus, by using NaN in EEMF data sets a more informative PARAFAC model was obtained. Three factors F1, F2, and F3 explained 43.12%, 13.69%, and 12.98% variance of the total captured variance of the data sets, respectively. The loading vectors along the second and third modes were multiplied together and the mathematically resolved EEMF profiles of F1, F2, and F3 were generated; the obtained profiles are shown in Fig. 6. EEMF profiles of F1, F2, and F3 were found to match with CY, BaP, and PY, respectively. However, the resolved EEMF spectra of BaP and PY were not found to be very pure. They had quite a few new features (artifacts) in their resolved spectra which were not present in their actual EEMF spectral profile. In addition to this, regions of $\lambda_{em} \leq (\lambda_{ex}+20)$ in the EEMF of F2 (i.e. BaP) were also found to contain the pattern of missing values. These artifacts (undesirable features) have appeared probably due to the wrong approximation for the missing values in the wavelength range of $\lambda_{em} \leq (\lambda_{ex}+20)$ and $\lambda_{em} \geq (2\times\lambda_{ex}-20)$ in EEMF data sets. Some errors might be due to the blank subtraction used to minimise the scattering or noise from the EEMF data sets. It should be noted that the blank subtraction generates number of variables which might have negative values, and PARAFAC was performed with non-negative constraints on all the three modes, therefore such constraints may also lead to lack of fit to the data sets. The artifacts in mathematically resolved spectral profile of unknown samples may bias the results of qualitative and quantitative analysis.

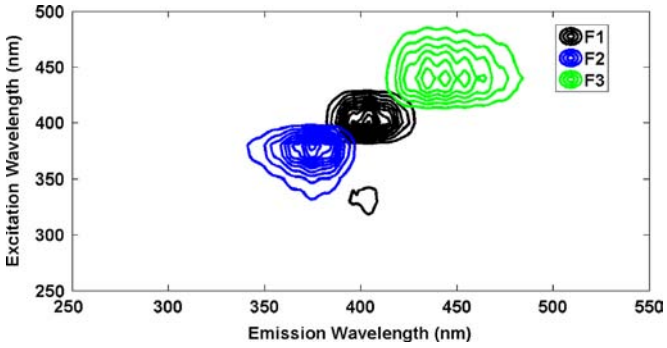

Fig. 5. EEMF spectra of F1–F3 retrieved from the PARAFAC analysis on the blank subtracted EEMF data of BaP–CY–PY mixtures.

The score values of F1–F3 were subjected to PCR analysis with the actual concentrations of BaP, CY, and PY, and a full rank PCR model was created. The regression equations are summarised in Table 3. It was found that slope of the regression equations has improved significantly for CY and PY, whereas it is still low (~0.7) for the BaP. The obtained PARAFAC model was tested by using the EEMF data of the six samples of validation set. Dimension of EEMF data set of the validation set is $6\times151\times251$. Prior to the PARAFAC analysis, non-trilinear part from these samples were minimised by blank subtraction and assigning NaN to the variables. The obtained score values of the F1, F2, and F3 of the validation set were subjected to the PCR model. Predicted amounts of the CY, BaP, and PY are summarised in Table 4. Root mean square error of prediction (RMSEP) [48,49] values were also calculated and summarised in Table 4. By analysing the slope and RMSEP values, it was found

Table 3
Regression equation "Y=slope $\times X$+intercept" relating the actual and predicted amounts of BaP, CY, and PY in calibration set; Y is the predicted amount and X is the actual amount.

Analysis	BaP	CY	PY
PARAFAC analysis of blank subtracted EEMF data	$Y=-0.003\times X+2.560$	$Y=0.389\times X+1.540$	$Y=0.025\times X+2.460$
PAAFAC analysis of blank subtracted EEMF data with missing values (NaN)	$Y=0.712\times X+0.555$	$Y=0.914\times X+0.230$	$Y=0.877\times X+0.261$
PARAFAC analysis of blank subtracted TSF data with missing values (NaN)	$Y=0.958\times X+0.080$	$Y=0.894\times X+0.262$	$Y=0.863\times X+0.318$
N-PLS analysis of blank subtracted EEMF data with missing values (NaN)	$Y=0.779\times X+0.409$	$Y=0.923\times X+0.201$	$Y=0.883\times X+0.189$
N-PLS analysis of blank subtracted TSF data with missing values (NaN)	$Y=0.958\times X+0.087$	$Y=0.922\times X+0.181$	$Y=0.877\times X+0.297$
MCR-ALS analysis of blank subtracted unfolded-EEMF data with zeros	$Y=0.723\times X+0.534$	$Y=0.903\times X+0.260$	$Y=0.877\times X+0.205$
MCR-ALS analysis of blank subtracted unfolded-TSF data	$Y=0.960\times X+0.082$	$Y=0.928\times X+0.162$	$Y=0.872\times X+0.162$

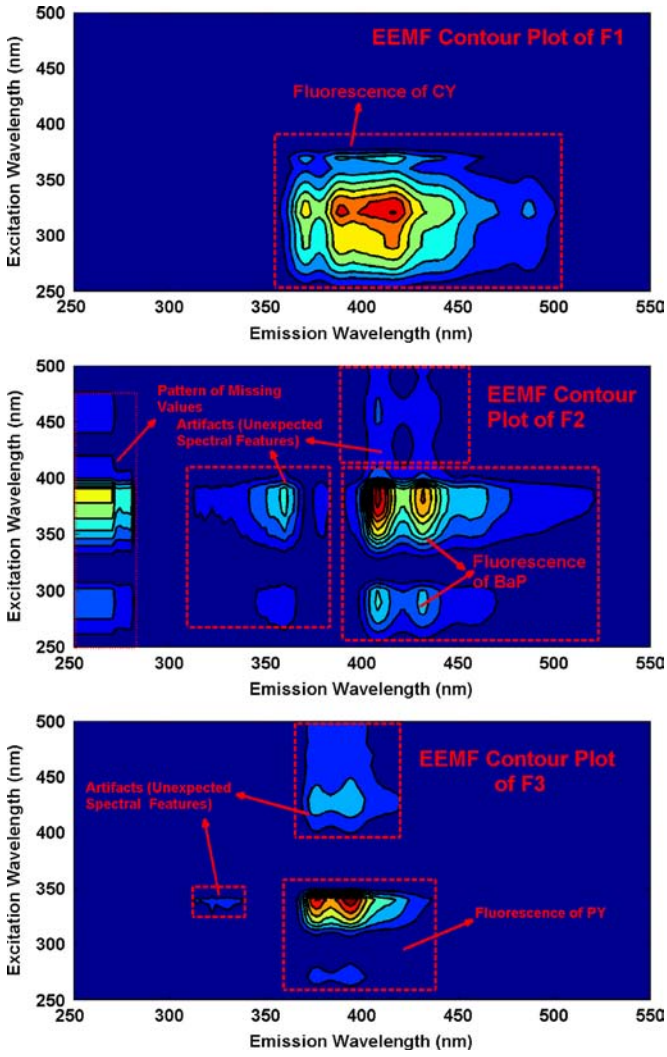

Fig. 6. EEMF spectra of F1(CY), F2(BaP), and F3(PY) retrieved from the PARAFAC analysis on blank subtracted EEMF data with missing values(NaN) in the wavelength range of $\lambda_{em} \leq (\lambda_{ex}+20)$ and $\lambda_{em} \geq (2\times\lambda_{ex}-20)$.

that model is most robust in predicting the amounts of CY. The slope of BaP was found to be small (~0.7) which indicates a poor calibration model; interestingly, its amounts were predicted more accurately than PY in the validation set. It is also evident from lower RMSEP value of BaP.

In the next part of the work, TSF data collected for the same set of the BaP–CY–PY samples were subjected to the PARAFAC analysis and their outcomes were analysed.

3.2.1.2. PARAFAC analysis of TSF data set. TSF based three way arrays do not have trilinear structure; therefore, before subjecting them to PARAFAC analysis, they must be represented in EEMF like layout which would ensure the required trilinearity [30]. Thus, blank subtracted TSF data sets of dimension $23\times151\times9$ (sample × excitation × $\Delta\lambda$) were presented in EEMF layout of dimension $23\times151\times251$ (i.e. sample × excitation × (excitation + $\Delta\lambda$) = $23\times151\times(151+100)$, 100 is the highest $\Delta\lambda$ with which TSF data were collected). The representation of TSF data in EEMF layout generates number of variables in the EEMF layout which do not contain any experimentally acquired information. Such variables were handled by setting them as NaN. These missing values were further handled by EM technique. TSF data in EEMF layout with NaN were subjected to PARAFAC analysis with non-negative constraints. The core consistency values of PARAFAC models with different factor were calculated. Core consistency value for 1–3 factor PARAFAC model were 100%. With addition of extra factor core consistency value dropped to zero. Hence, three factors were found to be sufficient for the PARAFAC analysis. It can be seen there is no ambiguity in the results of PCA and core consistency analysis for the selection of number of factors required for the PARAFAC analysis. Three factor PARAFAC model explained ~80% variance of the data set; F1, F2, and F3 individually explained 36.03%, 23.74%, and 19.60% variances, respectively. Pure EEMF layouts which contain the pure TSF data of each factor were created by multiplying the loading vectors along "excitation" and "excitation + $\Delta\lambda$" modes. From EEMF layout, by collecting the data diagonally at different offsets between the two modes, pure SF profiles at different $\Delta\lambda$s were obtained. These, SF profiles were further arranged as a function of increasing offsets and pure TSF spectra were created for each of the three factors, and are shown in Fig. 7. TSF profiles of F1, F2, and F3 were found to resemble with CY, PY, and BaP, respectively. Resolved TSF profiles of all the three fluorophores are much pure than their resolved EEMF profiles. It is due to the fact that PARAFAC analysis was performed on the TSF data in EEMF like layout and from there SF spectra at offsets of interest were extracted to create the TSF profiles of the fluorophores. Thus, the chances of the artifact which might appear in the PARAFAC resolved TSF spectra because of the wrong approximation of missing values in the EEMF layout are minimised. However, resolved TSF profile of CY is still slightly distorted compared to its experimentally acquired profile and it could be due to (i) subtraction of blank used for the minimisation of the Raman scattering and (ii) the non-negative constraints imposed on the PARAFAC analysis. Nevertheless, shape of SF profile varies as a function of $\Delta\lambda$ which makes TSF profile more distinctive for a fluorophore than EEMF profile.

The score values of F1–F3 and the actual amounts of the CY, BaP, and PY were subjected the PCR analysis to create a full rank PCR model. Regression equation relating the actual and PCR predicted amounts of the three fluorophores are summarised in Table 3. Slope of the regression equations for BaP, CY, and PY were found to be well within the acceptable limits. The obtained PCR model has the highest slope of 0.96 for the BaP, whereas in EEMF based PARAFAC model BaP had lowest slope of ~0.7. For CY and PY the slopes are close 0.91 and 0.84, respectively. Thus, it can be concluded that obtained TSF data based PARAFAC model is more efficient than EEEMF based PARAFAC model in resolving the spectral as well as contribution profiles of the present multifluorophoric mixture. The model was further tested by using it to predict the amounts of the CY, BaP, and PY in the six samples of validation set. Blank subtracted TSF data in EEMF layout of dimension $6\times151\times251$ were subjected to the developed PARAFAC model. Obtained score values of the three factors were further subjected to the developed full rank PCR model. Predicted amounts of the fluorophores are given in Table 4. In majority of the samples predicted amounts of BaP, CY, and PY are much more close to their actual amounts. This is also evident from the smaller RMSEP values (given in Table 4) for all the three components in TSF based model. It is to be noted that unlike PARAFAC–EEMF combination, TSF data with PARAFAC facilitated much more accurate quantification of the BaP in both calibration and validation sets.

Above obtained results clearly show that TSF data sets were analysed more efficiently than the EEMF data sets using PARAFAC for the present case of BaP–CY–PY mixtures. EEMF and TSF data of these mixtures were further subjected to the N-PLS analysis. Unlike PARAFAC which involves decomposition of the independent data set without using any information from the dependent data set, N-PLS decomposes both dependent and independent data sets simultaneously and tries to achieve maximum correlation between them. The main objective of N-PLS analysis is to create calibration model which could be used for the quantification of various components present in the mixtures.

Table 4
Predicted amounts of BaP, CY, and PY in test samples of validation set.

Fluorophores	Model	Predicted amounts (ng/ml)						
		Test samples (validation set)						
		Sample 24	Sample 25	Sample 26	Sample 27	Sample 28	Sample 29	RMSEP (ng/ml)
BaP	Model A	3.25	2.87	2.11	2.51	3.14	2.57	1.10
	Model B	3.41	3.72	1.76	2.87	3.85	3.69	0.70
	Model C	3.03	2.85	3.21	2.28	3.11	2.54	1.19
	Model D	3.96	4.34	2.60	3.48	4.35	4.24	0.53
	Model E	3.20	2.85	2.28	2.48	3.15	2.55	1.11
	Model F	3.63	3.63	1.87	2.82	3.79	3.58	0.67
CY	Model A	3.16	1.44	4.50	2.89	2.26	1.79	0.73
	Model B	3.17	1.44	3.45	2.94	2.14	1.62	0.61
	Model C	3.25	1.45	3.98	2.97	2.26	1.80	0.59
	Model D	1.84	0.43	1.55	1.57	1.06	0.59	1.83
	Model E	3.19	1.46	4.35	2.92	2.27	1.80	0.68
	Model F	3.39	1.91	3.03	3.08	2.35	1.81	0.44
PY	Model A	1.82	3.97	−0.15	2.48	2.17	2.82	1.93
	Model B	1.79	3.46	1.70	2.55	2.16	3.02	1.19
	Model C	1.76	3.99	0.25	2.45	2.21	2.84	1.78
	Model D	3.15	4.82	3.71	4.03	3.37	4.30	0.89
	Model E	1.77	3.96	0.06	2.45	2.16	2.83	1.85
	Model F	1.58	3.02	2.16	2.41	1.97	2.89	1.11

(i) Model A – PAAFAC analysis of blank subtracted EEMF data with missing values (NaN).
(ii) Model B – PARAFAC analysis of blank subtracted TSF data with missing values (NaN).
(iii) Model C – N-PLS analysis of blank subtracted EEMF data with missing values (NaN).
(iv) Model D – N-PLS model on blank subtracted TSF data with missing values (NaN).
(v) Model E – MCR-ALS analysis of blank subtracted unfolded-EEMF data with zeros.
(vi) Model F – MCR-ALS analysis of blank subtracted unfolded-TSF data sets.

3.2.2. *N-PLS analysis of EEMF and TSF data sets of BaP-CY-PY mixtures*

3.2.2.1. N-PLS analysis of EEMF data set. Subtraction of blank EEMF alone was not found enough to eliminate the Rayleigh scattering (i.e. non-trilinear part) from the EEMF data sets of BaP–CY–PY mixtures. Presence of non-trilinear part was also evident from the fact that PARAFAC failed to model the blank subtracted EEMF data sets adequately. Thus, N-PLS analyses on blank subtracted EEMF data sets which still have the non-trilinear part were not performed. Nevertheless, N-PLS analysis was performed on the blank subtracted EEMF data sets of dimension $23 \times 151 \times 251$ with missing values (NaN) in the wavelength range of $\lambda_{em} \leq (\lambda_{ex}+20)$ and $\lambda_{em} \geq (2 \times \lambda_{ex}-20)$, with the actual amounts of the CY, PY, and BaP. The N-PLS model of three factors were created; one factor per fluorophore. N-PLS model explained 92.98% and 91.56% variance of the independent (i.e. spectral) and dependent (i.e. concentration) data sets, respectively. The regression equation relating the actual and predicted amounts of the fluorophores are given in Table 3. Slope of the regression equation for the CY, PY and BaP were found to be greater than what was obtained from PARAFAC analysis on EEMF data sets. It shows that N-PLS model is more robust compare to the PARAFAC model for all the three fluorophores of BaP–CY–PY mixtures.

The obtained N-PLS model was tested using the validation set of six samples. Predicted amounts and RMSEP values of CY, BaP, and PY are summarised in Table 4. The RMSEP value of CY was found to be very small compare to the PY and BaP. Overall results of the N-PLS analysis on calibration and validation sets show that the obtained model is robust for all the three components. In the next part of the work, N-PLS analysis was performed on the TSF data sets to see if a better calibration model can be obtained.

3.2.2.2. N-PLS analysis of TSF data set. N-PLS analysis on the blank subtracted TSF data sets in EEMF layout of dimension $23 \times 151 \times 251$ with missing values (NaN) was performed with three factors. It was possible to explain 89.70% variance of the spectral data set and 96.51% variance of the concentration data sets. Amount of spectral variance explained by the N-PLS model of TSF data was relatively smaller than the N-PLS model of EEMF data set. However, it captured more variance of the dependent data sets compare to the N-PLS model of EEMF data sets. Regression equations of the N-PLS model are summarised in Table 3. It can be seen that the calibration parameters for the CY and PY are similar to which were obtained from the N-PLS analysis on EEMF data sets. However, for BaP there was signification improvement in slope from 0.78 (EEMF based N-PLS model) to 0.96. The model was further used to predict the concentrations of the fluorophores in the six samples of validation set. The predicted amounts and RMSEP values for CY, BaP, and PY are summarised in Table 4. Despite good calibration parameters, for an unknown reason the predicted amounts of CY were low and it was consistent in all the six samples of the validation set. However, predictions for the PY and BaP were much better than the predictions obtained from N-PLS model of EEMF data sets. Slope of regression equation and RMSEP values indicates that the N-PLS model of TSF data is relatively more reliable than the N-PLS model of EEMF data set for the present case of BaP–CY–PY mixtures.

Similar to PARAFAC, MCR-ALS analysis is another curve resolution technique which is used to retrieve the pure concentration and spectral profile of a fluorophore from the multifluorophoric mixture. In the next part of the current work, the MCR-ALS analyses were performed on the EEMF and TSF data sets and their outcomes were analysed and compared with each other.

3.2.3. *MCR-ALS analysis of EEMF and TSF data sets of BaP-CY-PY mixtures*

3.2.3.1. MCR-ALS analysis on EEMF data set. Despite the fact that PCA predicted one as the optimum factor for the chemometric analysis, MCR-ALS analysis was performed on the blank subtracted unfolded-EEMF data set of dimension 23×3926 with three factors with an initial estimate of concentration profile. Non-negative constrains were imposed on concentration and spectral modes. Factors F1, F2, and F3 of MCR-ALS model explained 41.95%, 7.40%,

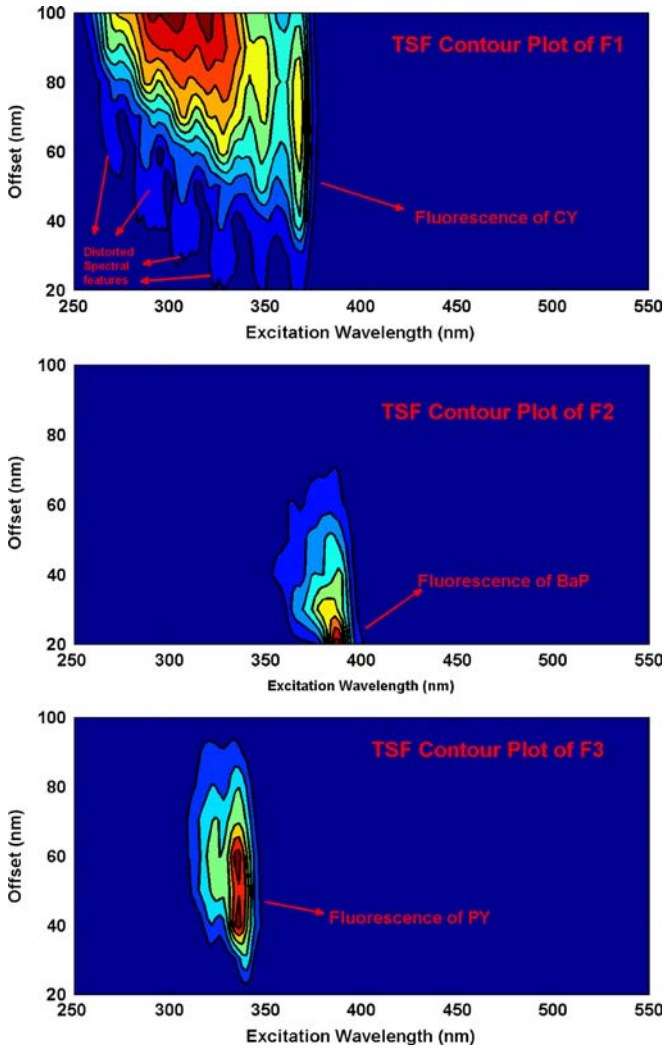

Fig. 7. TSF spectra of F1(CY), F2 (BaP), and F3(PY) retrieved from the PARAFAC analysis on blank subtracted TSF data in EEMF like layout with missing values (NaN).

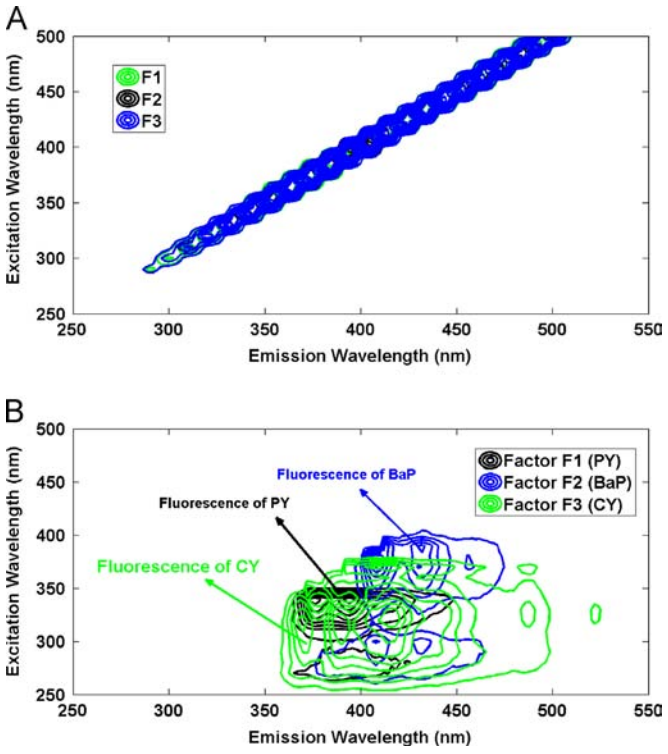

Fig. 8. [A] EEMF spectra of F1–F3 obtained from the MCR-ALS analysis of blank subtracted EEMF data set and [B] EEMF spectra of F1(PY), F2(BaP), and F3(CY) retrieved from the MCR-ALS analysis on blank subtracted EEMF data containing zero in the wavelength range of $\lambda_{em} \leq (\lambda_{ex}+20)$ and $\lambda_{em} \geq (2 \times \lambda_{ex}-20)$.

and 44.81% variances of the data sets, respectively. Cumulatively these three factors captured ~94% variance of the data sets. Loading vectors (i.e. contains spectral information) of F1, F2, and F3 of dimensions 1×3926 were reshaped to obtain matrices of dimension 151×26 (emission × excitation). The reshaped data were further plotted, shown in Fig. 8(A), to obtain the resolved EEMF spectra of F1–F3. It mainly contains the diagonal band of Rayleigh scattering. Similar to PARAFAC, MCR-ALS too could not model the blank subtracted EEMF data sets. In PARAFAC and N-PLS analysis, Rayleigh scattering signals were eliminated from the blank subtracted data sets by using the missing values (NaN). However, MCR-ALS analyses on the EEMF data with NaN were not found to converge with three or any other number of factors. Thus, for MCR-ALS analysis Rayleigh scatterings were eliminated by using the zero values in wavelength range where $\lambda_{em} \leq (\lambda_{ex}+20)$ and $\lambda_{em} \geq (2 \times \lambda_{ex}-20)$. Use of zero in the fluorescence data is generally not well accepted mainly due to two reasons (i) it may lead to the loss of fluorescence information particularly when the fluorescence of any fluorophore falls on the scattering signals or lies very close to it and (ii) in a multifluorophoric mixture there could be some fluorophores which may have fluorescence beyond second order Rayleigh scattering also. Before proceeding further PCA analysis was performed on EEMF data to find optimum number of factors. RMSECV was found to be minimum for a 3 component PCA model, shown in Fig. 4(C). Thus, there is correlation between the actual and predicted rank of the data sets. A three factor MCR-ALS model was created; it explained 90.18% variance of the spectral data. F1, F2, and F3 explained 29.51%, 13.75%, and 46.93% variances of the data set. Loading vectors of F1–F3 were reshaped to obtain their resolved EEMF spectra, shown in Fig. 8(B). EEMF profile of F1, F2, and F3 were found to resemble with PY, BaP, and CY, respectively. MCR-ALS resolved EEMF spectra of BaP, CY, and PY are much pure than the PARAFAC resolved spectra. The score values of F1–F3 with the actual concentrations of BaP, CY, and PY were subjected to create the full rank (3 factor) PCR model. Regression equations relating the actual and predicted amounts of BaP, CY, and PY are summarised in Table 3. Slope of the regression equations are ~0.9 for CY and PY. Similar to PARAFAC and N-PLS analysis, here also there is a significant difference between the measured and predicted amounts of BaP. It is evident from the small slope (~0.7) of regression equation relating the actual and predicted amounts of BaP.

Like PARAFAC and N-PLS models, MCR-ALS model was also tested by the validation set of same six samples. Score values of these six samples obtained from MCR-ALS analysis were subjected to PCR model. The predicted amounts and RMSEP values of the fluorophores are summarised in Table 4. Obtained results show that MCR-ALS is most robust for CY, followed by PY and BaP. Nevertheless, MCR-ALS analysis separated the EEMF spectra of BaP, CY, and PY very efficiently from EEMF data of BaP–CY–PY mixtures.

3.2.3.2. MCR-ALS analysis of TSF data set. One important thing which is necessary to be mentioned here is that TSF data sets

are most suited for the bilinear analysis techniques such as PCA and MCR-ALS analysis. It is so because, it does not require any zero filling or setting variables to the missing for the elimination of non-trilinear or non-bilinear part. Rayleigh scattering is eliminated at the data acquisition stage itself and Raman scattering can easily be minimised by blank subtraction. Thus, pre-processing or data tampering is minimal in the bilinear analysis of TSF data sets compared to bilinear analysis of EEMF data sets.

Blank subtracted unfolded-TSF data set of dimension 23×1359 (sample $\times$ (excitation $\times \Delta\lambda$)) was subjected to the MCR-ALS analysis with (i) three factors, and (ii) an initial estimate of concentration profile, and (iii) non-negative constraints on sample and spectral modes. MCR-ALS model of three factors explained 93.51% variance of the spectral data set. These three factors F1, F2, and F3 separately explained 32.87%, 22.49%, and 38.15% variance of the data sets. Loading vectors of F1–F3 of dimension 1×1359 were reshaped to create the TSF data sets of dimension 151×9 (excitation $\times \Delta\lambda$). The resolved TSF contour plots of F1, F2, and F3, shown in Fig. 9, resembles with BaP, CY, and PY, respectively. It can be seen that all the three resolved spectra are pure. Unlike PARAFAC resolved TSF spectra (Fig. 7), MCR-ALS resolved spectral profiles do not have any distortion and it is true for all the three fluorophores. It is mainly due to the fact that (i) there were no missing variables (NaN) to be approximated in the data sets and (ii) MCR-ALS technique is relatively soft in analysing the data sets. The obtained spectral profile of BaP, CY and PY proves that TSF spectra of a mixture can be resolved more efficiently using MCR-ALS analysis than PARAFAC.

The score values of F1–F3 and the actual concentrations of the fluorophores were subjected to the PCR analysis. The regression equation relating the actual and predicted amounts are summarised in Table 3. The calibration model was found to be robust for all the three fluorophores. Model was tested by the same six samples, which were used throughout the present study. Predicted amounts and RMSEP values of the BaP, CY, and PY in validation set are summarised in Table 4. The results were found to be well within the acceptable limits for all the three components.

It is to be noted that in all the three models of EEMF data i.e. PARAFAC, N-PLS and MCR-ALS, amounts of BaP were not predicted much accurately. It is due to the fact that in EEMF spectra of BaP, shown in Fig. 2(B), Raman scattering of water is severely overlapped with the fluorescence of BaP. On the hand, in the TSF spectra of BaP, shown in Fig. 2(C), water Raman signals are away from the wavelength range over which the BaP shows maximum fluorescence. This separation of Raman signals from the BaP fluorescence in TSF mode is the reason for more accurate estimation of BaP with the TSF data than with EEMF data in all the three models. It shows that even if Raman scattering signals (which lack bilinearity and trilinearity) can be minimised or eliminated by blank subtraction but it can still cause lack of linearity between the fluorescence intensity and the concentration of a fluorophore which has direct affect on the outcomes of chemometric methods. Thus, in quite a few cases non-linearity caused by Raman scattering effects can be minimised by collecting the fluorescence data in TSF spectroscopy mode.

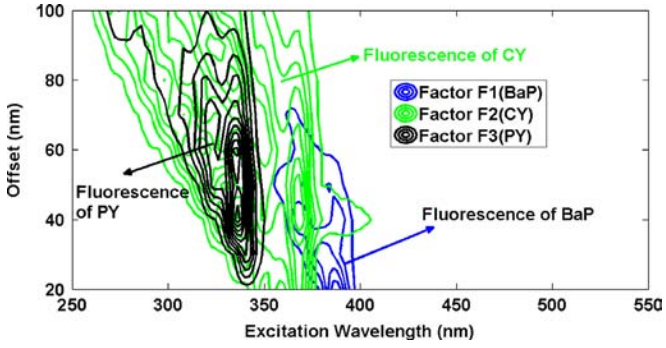

Fig. 9. TSF spectra of F1(BaP), F2(CY), and F3(PY) retrieved from the MCR-ALS analysis on blank subtracted TSF data sets.

In the present case of BaP–CY–PY mixtures, TSF was found to provide better qualitative information than the EEMF spectroscopy. In addition, the resolved spectra and concentration profiles of BaP, CY, and PY obtained from the chemometric analyses of TSF data sets were more accurate than the resolved spectra and concentration profile obtained from the chemometric analysis of EEMF data sets. The obtained results show that TSF is a potential alternate to the EEMF spectroscopy, particularly when someone has to analyse dilute samples where concentration of the fluorophores are of the order of ng/ml or less than that.

4. Conclusions

In the present work, a comprehensive comparative study for the EEMF and TSF spectroscopic data set was performed by taking dilute aqueous multifluorophoric mixtures of BaP–CY–PY. The Rayleigh scattering, a non-trilinear part of EEMF data sets is generally eliminated using blank subtraction, zero or missing values, or interpolation methods; each of these methods have their own limitations. There is no single well accepted recipe for the elimination of Rayleigh scattering from EEMF data sets. On the other hand, $\Delta\lambda$, a purely instrumental parameter in TSF spectroscopy, enables the removal of Rayleigh scattering from the TSF data right at the data acquisition step. It was also found that extent of mixing of Raman scattering signals with fluorescence may vary in EEMF and TSF spectroscopy.

EEMF and TSF data sets for the dilute aqueous mixtures of BaP–CY–PY were subjected to three chemometric techniques PARAFAC, N-PLS, and MCR-ALS and their outcomes were analysed to compare the quality of the EEMF and TSF data sets. In chemometric analysis it is essential that actual (i.e. number of fluorophores present in the mixtures) and mathematical rank should be same. In the present case of BaP–CY–PY mixtures, mathematical rank of EEMF data set was found to differ from the actual rank; whereas, there was no such ambiguity in the selection of optimum number of factors for the chemometric analyses of TSF data sets. The spectral and concentration profiles of BaP, CY, and PY which were retrieved using the PARAFAC analysis on the TSF data were more close to their actual profiles than that obtained from the PARAFAC analysis of EEMF data. It was due to the fact that PARAFAC analysis was performed on TSF data in EEMF like layout and from there SF spectra at offsets of interest were extracted, therefore, the artifact which could appear in the PARAFAC resolved TSF spectra because of the wrong approximation of missing values in the EEMF layout were minimised. In addition, N-PLS model of TSF data was found to be more accurate than the N-PLS model of EEMF data in predicting the amounts of BaP, CY, and PY in majority of the BaP–CY–PY mixtures. TSF data sets of dilute multifluorophoric mixtures was found to have significant advantage over EEMF data, particularly, when they are analysed using the bilinear analysis techniques such as MCR-ALS. It is because TSF data sets do not require any mathematical data pre-treatment to eliminate non-bilinear part from the data sets. It was also found that PARAFAC, N-PLS, and MCR-ALS models of TSF data gave almost quantitative estimation of BaP, whereas, PARAPAC, N-PLS, and MCR-ALS model of EEMF data sets made poor estimation of BaP. It was due to the fact that in EEMF, Raman scattering and BaP fluorescence were severely overlapped, whereas, in TSF spectroscopy, Raman scattering and BaP fluorescence were well separated. The presence of Raman scattering signals (non-bilinear or non-trilinear) which are highly intense in the dilute aqueous mixtures of BaP–CY–PY

causes the lack of linearity between the fluorescence intensity and the concentration of fluorophores which cannot be removed or minimised by blank subtraction. This lack of linearity for BaP is more in EEMF whereas it is minimal in TSF spectroscopy. Thus, for quite a few fluorophores it is possible that influence of Raman scattering on the fluorescence can be minimised at the instrumental level by collecting the fluorescence data in TSF spectroscopy mode. The obtained results of PARAFAC, N-PLS, and MCR-ALS analyses clearly shows that for the dilute multifluorophoric mixtures TSF data provides more quantitative and qualitative information than EEMF data sets. In summary, TSF is a potential alternate to the EEMF spectroscopy, in particular, for the analysis of dilute aqueous solutions.

Acknowledgements

Keshav Kumar is thankful to the Council of Scientific and Industrial Research (CSIR) New Delhi for providing the fellowship. The authors thank CSIR for the financial support to carry out the work.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2013.09.002.

References

[1] I.M. Warner, J.B. Callis, E.R. Davidson, M. Gouterman, G.D. Christian, Anal. Lett. 8 (1975) 665–681.
[2] I.M. Warner, J.B. Callis, E.R. Davidson, G.D. Christian, Clin. Chem. 22 (1976) 1483–1492.
[3] I.M. Warner, J.B. Callis, G.D. Christian, E.R. Davidson, Anal. Chem. 49 (1977) 564–573.
[4] I.M. Warner, G.D. Christian, E.R. Davidson, Anal. Chem. 49 (1977) 2155–2159.
[5] D.W. Johnson, J.B. Callis, G.D. Christian, Anal. Chem. 49 (1977) 747A–757A.
[6] D. Patra, A.K. Mishra, Anal. Bioanal. Chem. 373 (2002) 304–309.
[7] M.M.D. Sierra, M. Giovanela, E. Paralanti, E.J. Soriano-Sierra, Chemosphere 58 (2005) 715–733.
[8] D. Patra, A.K. Mishra, Anal. Chim. Acta 454 (2002) 209–215.
[9] E.M. Carstea, A. Baker, M. Bieroza, D. Reynolds, Water Res. 44 (2010) 5356–5366.
[10] M.Z. Bieroza, J. Bridgeman, A. Baker, Drinking Water Eng. Sci. 3 (2010) 63–70.
[11] C.L. Muller, A. Baker, R. Hutchinson, I.J. Fairchild, C. Kidd, Atmos. Environ. 42 (2008) 8036–8045.
[12] P.G. Coble, S.A. Green, N.V. Blough, R.G. Gagosian, Nature 348 (1990) 432–435.
[13] A.F. Zuluaga, U. Utzinger, A. Durkin, H. Fuchs, A. Gillenwater, R. Jacob, B. Kemp, J. Fan, R.R. Kortum, Appl. Spectrosc. 53 (1999) 302–311.
[14] B. Li, P.W. Ryan, M. Shanahan, K.J. Leister, A.G. Ryder, Appl. Spectrosc. 65 (2011) 1240–1249.
[15] J.L. Beltrán, R. Ferrer, J. Guiteras, Anal. Chim. Acta 373 (1998) 311–319.
[16] H. Wang, Y. Zhang, X. Xiao, Anal. Sci. 26 (2010) 1271–1276.
[17] H.B. Wang, Y.J. Zhang, X. Xiao, S.H. Yu, W.Q. Liu, Anal. Methods 3 (2011) 688–695.
[18] Y. Yamashita, R. Jaffé, Environ. Sci. Technol. 42 (2008) 7374–7379.
[19] C.A. Stedmon, R. Bro, Limnol. Oceanogr. Methods 6 (2008) 572–579.
[20] Z. Wang, W. Liu, N. Zhao, H. Li, Y. Zhang, W. Si-Ma, J. Liu, J. Environ. Sci. 19 (2007) 787–791.
[21] A.G. Ryder, J. Fluoresc. 14 (2004) 99–104.
[22] A.K. Sarma, A.G. Ryder, Energy Fuels 20 (2006) 783–785.
[23] T. Dramićanin, B. Dimitrijević, M.D. Dramićanin, Appl. Spectrosc. 65 (2011) 293–297.
[24] T. Dramićanin, L. Lenhardt, I. Zeković, M.D. Dramićanin, J. Fluoresc. 22 (2012) 1281–1289.
[25] K.I. Poulli, G.A. Mousdis, C.A. Georgiu, Food Chem. 105 (2007) 369–375.
[26] A. Nevin, D. Comelli, G. Valentini, R. Cubeddu, Anal. Chem. 81 (2009) 1784–1791.
[27] K. Kumar, A.K. Mishra, Anal. Methods 3 (2011) 2616–2624.
[28] A.P. Teixeira, T.M. Duarte, M.J.T. Carrondo, P.M. Alves, Biotechnol. Bioeng. 108 (2011) 1852–1861.
[29] J.H. Rho, J.L. Stuart, Anal. Chem. 50 (1978) 620–625.
[30] J.B.F. Llyod, Nat. Phys. Sci. 231 (1971) 64–65.
[31] T. Vo-Dinh, Anal. Chem. 50 (1978) 396–401.
[32] D. Patra, A.K. Mishra, Trends Anal. Chem. 21 (2002) 787–798.
[33] D.A. Skoog, D.M. West, F.J. Holler, Fundamentals of Analytical Chemistry, seventh ed., Saunders College Publishing, Philadelphia, 1996.
[34] J.R. Lakowicz, Fluorescence Spectroscopy, third ed., Springer, New York, 2006.
[35] K. Kumar, A.K. Mishra, Anal. Chim. Acta 755 (2012) 37–45.
[36] A. de Juan, R. Tauler, J. Chemom. 15 (2001) 749–772.
[37] K. Kumar, A.K. Mishra, Chemom. Intell. Lab. Syst. 116 (2012) 78–86.
[38] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[39] K. Kumar, A.K. Mishra, J. Fluoresc. 22 (2012) 339–347.
[40] A.A. de Silva, D.D. Keukeleire, D.R. Cardoso, D.W. Franco, Anal. Methods 4 (2012) 642–646.
[41] R. Bro, N. Viereck, M. Toft, H. Toft, P.I. Hansen, S.B. Engelsen, Trends Anal. Chem. 29 (2010) 281–284.
[42] H. Shinzawa, M. Nishida, W. Kanematsu, T. Tanaka, K. Suzuki, I. Noda, Analyst 137 (2012) 1913–1921.
[43] L. Shintu, S. Caldarelli, J. Agric. Food Chem. 53 (2005) 4026–4031.
[44] J.A. Rodrigues, A.S. Barros, B. Carvalho, T. Brandão, A.M. Gil, J. Chromatogr. A 1218 (2011) 990–996.
[45] N.J. Nielsen, D. Ballabio, G. Tomasi, R. Todeschini, J.H. Christensen, J. Chromatogr. A 1238 (2012) 121–127.
[46] J. Kuligowski, D. Carrión, G. Quintás, S. Garrigues, M. de la Guardia, Food Chem. 131 (2012) 353–359.
[47] M. Bassbasi, A. Hafid, S. Platikanov, R. Tauler, A. Oussama, Fuel 104 (2013) 798–804.
[48] R. Kramer, Chemometric Techniques for Quantitative Analysis, Marcel Decker, New York, 1998.
[49] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Winding, R.S. Koch, PLS_Toolbox 4.0. Eigen-Vector Research, 2006.
[50] R. Bro, Multi-way Analysis in the Food Industry, Theory Algorithms and Applications (PhD thesis), University of Amsterdam, 1998.
[51] S. Wold, K. Esbensen, P. Geladi, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.
[52] R.A. Harshman, UCLA Work. Pap. Phon. 16 (1970) 1–84.
[53] J.D. Carol, J. Chang, Pyschometrika 35 (1970) 283–319.
[54] R.A. Harshman, Comput. Stat. Data Anal. 18 (1994) 39–72.
[55] R. Bro, J. Chemom. 10 (1996) 47–61.
[56] R. Tauler, B. Kowalski, S. Fleming, Anal. Chem. 65 (1993) 2040–2047.
[57] R. Tauler, Chemom. Intell. Lab. Syst. 30 (1995) 133–146.
[58] T. Azzouz, R. Tauler, Talanta 74 (2008) 1201–1210.
[59] H. Abdollahi, R. Tauler, Chemom. Intell. Lab. Syst. 108 (2011) 100–111.
[60] A.C. Olivieri, J. Chemom. 19 (2005) 253–265.
[61] M.C. Bauza, G.A. Ibañez, R. Tauler, A.C. Olivieri, Anal. Chem. 84 (2012) 8697–8706.
[62] W. Guo, J. Xu, J. Wang, Y. Wen, J. Zhuo, Y. Yan, J. Environ. Sci. 22 (2010) 1728–1734.
[63] T. Ohno, A. Amirbahaman, R. Bro, Environ. Sci. Technol. 42 (2008) 186–192.
[64] R.M. Cory, D.M. McKnight, Environ. Sci. Technol. 39 (2005) 8142–8149.
[65] C.A. Stedmon, S. Markager, R. Bro, Mar. Chem. 82 (2003) 239–254.
[66] Y. Ni, S. Su, S. Kokot, Anal. Chim. Acta 580 (2006) 206–215.
[67] A.M. de la Peña, N. Mora Díez, D.B. Gil, A.C. Olivieri, G.M. Escandar, Anal. Chim. Acta 569 (2006) 250–259.
[68] J. Nie, H. Wu, X. Wang, Y. Zhang, S. Zhu, R. Yu, Anal. Chim. Acta 628 (2008) 24–32.
[69] O. Divya, A.K. Mishra, Talanta 72 (2007) 43–48.
[70] O. Divya, A.K. Mishra, Anal. Chim. Acta 592 (2007) 82–90.
[71] R. Bro, H.A.L. Kiers, J. Chemom. 17 (2003) 274–286.
[72] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 50 (1978) 1108–1113.
[73] R.D. Jiji, G.G. Andersson, K.S. Booksh, J. Chemom. 14 (2000) 171–185.
[74] R.D. Jiji, K.S. Booksh, Anal. Chem. 72 (2000) 718–725.
[75] L.G. Thygesen, Å. Rinnan, S. Barsberg, J.K.S. Møller, Chemom. Intell. Lab. Syst. 71 (2004) 97–106.
[76] R.D. Jiji, G.A. Cooper, K.S. Booksh, Anal. Chim. Acta 397 (1999) 61–72.
[77] R. Bro, Chemom. Intell. Lab. Syst. 46 (1999) 133–147.
[78] G. Tomasi, R. Bro, Chemom. Intell. Lab. Syst. 75 (2005) 163–180.
[79] M. Bahram, R. Bro, C. Stedmon, A. Afkhami, J. Chemom. 20 (2006) 99–105.
[80] B. Walczak, D.L. Massart, Chemom. Intell. Lab. Syst. 58 (2001) 15–27.
[81] B. Walczak, D.L. Massart, Chemom. Intell. Lab. Syst. 58 (2001) 29–42.