Journal of Fluorescence (2019) 29:1183–1190
https://doi.org/10.1007/s10895-019-02435-8

ORIGINAL ARTICLE

# Non-negative Factor (NNF) Assisted Partial Least Square (PLS) Analysis of Excitation-Emission Matrix Fluorescence Spectroscopic Data Sets: Automating the Identification and Quantification of Multifluorophoric Mixtures

Keshav Kumar[1]

Received: 27 April 2019 / Accepted: 3 September 2019 / Published online: 10 September 2019


**Abstract**
Excitation-emission matrix fluorescence spectroscopy is simple and sensitive techniques that generate the composite fluorescence fingerprints. EEMF can be used for the identification and quantification of the fluorophores without involving any pre-separation step provided a suitable data analysis approach is applied. In the present work, non-negative factor (NNF) assisted partial least square (PLS) analysis is used for the analysis of EEMF data sets acquired for the dilute aqueous mixtures of fluorophores. The proposed approach allows automatic selection of the optimum number of factors for NNF analysis by incorporating the Akaike information criterion. The proposed approach also incorporates the spectral correlation analysis for the automatic identification of the NNF retrieved EEMF spectral profiles. The NNF retrieved contribution values along with their real concentration values are subjected to PLS analysis to develop a calibration model. The proposed approach was successfully tested using EEMF data acquired for the dilute aqueous mixtures of Catechol, Hydroquinone, Indole, Tryptophan and Tyrosine. The results were evaluated using the various statistical parameters and each of them found to well within the expected limits. In summary, NNF assisted PLS analysis of EEMF technique allows automatized analysis of the multifluorophoric mixtures with minimum user inputs.

**Keywords** Excitation-emission matrix fluorescence · Multifluorophoric mixture · Spectral correlation · Non-negative factor analysis · Partial least square analysis

## Introduction

Excitation-emission matrix fluorescence (EEMF) spectroscopy is a simple, sensitive and fast analytical technique [1–5]. An EEMF spectrum is a collection of emission and excitation spectra collected at varying excitation and emission wavelengths, respectively. EEMF technique can be used to create *fingerprints* of the fluorescent molecules (i.e. fluorophores). The *fingerprints* can be used to create the database and subsequently might be used for the identification of the fluorophores in the samples of unknown compositions. EEMF technique has found its application in the fields of petrochemicals, agricultural, pharmaceutical, and clinical fields [1–11]. The fluorescence intensity of a fluorophore in the EEMF spectroscopy can be described using the Eq. 1

$$I(\lambda_{ex}, \lambda_{em}) = kClE_{ex}(\lambda_{ex})E_{em}(\lambda_{em}) \quad (1)$$

In the above equation, $I(\lambda_{ex}, \lambda_{em})$ is the intensity of a fluorophore in EEMF mode, $E_{ex}$ is the excitation profile, $E_{em}$ is the emission profile, $C$ is the concentration, $l$ is the path length, $\lambda_{ex}$ is the excitation wavelength, $\lambda_{em}$ is the emission wavelength and K is the constant related to instrumental configuration. The above equation is valid as long as the Beer-Lambert law is obeyed [4, 5]. The above equation suggests that one can extract the excitation spectrum from EEMF spectrum by collecting the fluorescence intensity at varying excitation wavelength and constant emission wavelength. Similarly, one can extract the emission spectrum by collecting the fluorescence intensity at varying emission wavelength and constant excitation wavelength. Moreover, EEMF data has a

✉ Keshav Kumar
keshavkumar29@gmail.com

[1] Hochschule Geisenheim University, 65366 Geisenheim, Germany

trilinear structure that mainly arises due to the fact shape of the excitation and emission spectrum for each fluorophore is invariant to the change in the emission and excitation wavelengths, respectively [4, 5, 12, 13]. The adherence of trilinearity in EEMF spectroscopic data set makes them truly compatible with each of the available chemometric techniques. The compatibility has been successfully explored to develop efficient analytical methods for analyzing different kind of multifluorophoric mixtures. Several of such applications could be seen in the recent review article by Kumar et al [4].

Non-negative factor (NNF) analysis is a chemometric technique that essentially carries matrix factorization of a given data set with the user specified number of factors [14–18]. NNF based factorization of the given data set X of dimension I × J (sample ×variable) with $F$ factors can be summarized using the Eq. 2.

$$X_{ij} = \sum_{f=1}^{F} W_{if} H_{fj} \tag{2}$$

In the above equation, $X_{ij}$ is the element of the matrix X, $W_{if}$ is the element of the matrix W of dimension (I × F) and $H_{fj}$ is the element of the matrix H of dimension (F × J). The matrix W contains the concentration profiles and the matrix H contains the spectral profiles for each of the user specified factors. It can be seen that NNF analysis can be a useful technique to extract the pure spectrum and the corresponding concentration values for each of fluorophores in multifluorophoric mixtures.

Partial least square (PLS) analysis is one of the most commonly used technique to develop a calibration model [19–23]. PLS algorithm reduces the dimension of the predictor data set by finding few uncorrelated components that not only captures the maximum variation of the predictor data set but also maximizes the correlation between the predictor and predicted data matrices. PLS algorithm serve as a useful tool for the developing a robust calibration model that allows simultaneous quantification of several analytes [4, 19–23] without involving any pre-separation step.

From the above discussion, it is clear that EEMF generate the composite fluorescence fingerprints for the multifluorophoric mixtures that are weighted by the concentration of the fluorophores. With the help of suitable chemometric approach, the composite fluorescence fingerprints can be separated and individual fingerprints could further be easily processed towards establishing their identity with the help of a database. The concentration related information for each fluorophores buried in the EEMF data sets can be used to develop a calibration model for their quantification. Often, the identification of the fluorophore is time consuming and computationally challenging task. To address this, present work proposes a computational approach by combining NNF with PLS algorithm analysis for the analysis of EEMF spectral data sets that enables identification and quantification of the fluorophores present in multifluorophoric mixtures with minimum user inputs. In the proposed approach, EEMF spectral data sets are first subjected to NNF technique to retrieve pure EEMF spectral profiles. In the next step, using a distance metric approach the best match for the retrieved pure EEMF spectral from the database is obtained so that their identity can be confirmed. The NNF retrieved concentration related information are subjected to PLS analysis to develop a robust calibration model for making their accurate and precise estimations. To carry out the present work, EEMF spectral data sets acquired for the dilute aqueous mixtures of five fluorophores of biological relevance Catechol, Hydroquinone, Indole, Tryptophan and Tyrosine are taken as the test case. The selected fluorophores has significant spectral overlap in EEMF spectra.

## Theory

### Non-negative Factor (NNF) Analysis

NNF model can be fitted on a given data set using the iterative algorithm [14–18] summarized using the Eqs. 3–5.

$$H^{n+1} = H^{n} \frac{(W^{n})^{T} X}{(W^{n})^{T} W^{n} H^{n}} \tag{3}$$

$$W^{n+1} = W^{n} \frac{X\left(H^{n+1}\right)^{T}}{W^{n} H^{n+1} \left(H^{n+1}\right)^{T}} \tag{4}$$

$$L(W,H) = \|X - WH\|_{F}^{2} \tag{5}$$

In the above equation $n$ is the iteration number, for $n = 1$, $W$ and $H$ are initialized with non-negative numbers. The $W^{n}$ and $H^{n}$ contains the updated values for the $W$ and $H$ matrices at the $n^{th}$ iteration respectively. A series of iterations are carried out until the loss function $L(W,H)$ given in Eq. 5 is minimized.

### Akaike Information Criterion (AIC) for Finding the Optimum Number of Factor

In the Akaike information criterion (AIC) [24–27] based approach for finding the optimum number of factor, NNF model is fitted with using the different number of factors (K). The set of factors that minimizes AIC value is considered as the optimum. The AIC (K) value is calculated for each of these models using the Eq. 6

$$AIC\ (K) = -\sum_{I,J} X(I,J) \log\left(\sum_{k=1}^{K} W(I,k) H(k,J)\right) \tag{6}$$

The AIC criteria provides an unambiguous approach for selecting the optimum number of factors for carrying out the chemometric analysis.

## Material and Methods

A calibration and validation set is created by selecting 32 dilute samples containing Catechol, Hydroquinone, Indole, Resorcinol, Tryptophan and Tyrosine in different proportions. The calibration set consist of 26 samples labeled as S1-S26 and the validation set consist of 6 samples labeled as V1-V6. The concentrations of these five fluorophores in each of these samples are reported in Table 1. The EEMF data sets for these samples are collected using the Varian Eclipse spectrofluorimeter. The excitation and emission wavelengths are varied in the range of 230–320 nm and 230–500 nm in steps of 5 and 2 nm, respectively. Both excitation and emission slit widths were kept at 5 nm. The scan rate were set to 1920 nm/min and PMT detector voltage were set to 600 V, respectively. For detailed discussion about the sample preparation and data acquisition one can refer to the work reported by Bro and co-workers [28, 29]. The optical density (OD) values of these samples were less than 0.87 suggesting there were no interferences due to the inner filter effects and the analyzed samples obey the Beer-Lambert law (i.e. fluorescence intensities of the fluorophores are directly proportional to their concentration).

**Table 1** Concentration of fluorophores in calibration and validation set samples

| Sample | Catechol×$10^{-4}$ M | Hydroquinone×$10^{-4}$ M | Indole×$10^{-4}$ M | Tryptophan×$10^{-4}$ M | Tyrosine×$10^{-4}$ M |
|---|---|---|---|---|---|
| S1 | 0.870 | 0.112 | 0.026 | 0.074 | 0 |
| S2 | 0.435 | 0.028 | 0 | 0.009 | 0.121 |
| S3 | 0 | 0.056 | 0.006 | 0.019 | 0.031 |
| S4 | 0.870 | 0 | 0.051 | 0.074 | 0.121 |
| S5 | 0.870 | 0.224 | 0.006 | 0 | 0.061 |
| S6 | 0.218 | 0 | 0.006 | 0.037 | 0.061 |
| S7 | 0.870 | 0.112 | 0.026 | 0 | 0.030 |
| S8 | 0 | 0.224 | 0.006 | 0.037 | 0.015 |
| S9 | 0.435 | 0.224 | 0 | 0.074 | 0.121 |
| S10 | 0 | 0.056 | 0.013 | 0.019 | 0.121 |
| S11 | 0.109 | 0.224 | 0.051 | 0 | 0.030 |
| S12 | 0.432 | 0.029 | 0 | 0.010 | 0.123 |
| S13 | 0 | 0.057 | 0.006 | 0.019 | 0.031 |
| S14 | 0.863 | 0 | 0.050 | 0.077 | 0.123 |
| S15 | 0.863 | 0.115 | 0.025 | 0 | 0.030 |
| S16 | 0.108 | 0 | 0.012 | 0.010 | 0.015 |
| S17 | 0 | 0.23 | 0.006 | 0.038 | 0.015 |
| S18 | 0.216 | 0.029 | 0.025 | 0 | 0.061 |
| S19 | 0.432 | 0.23 | 0 | 0.077 | 0.123 |
| S20 | 0.432 | 0.115 | 0.012 | 0.038 | 0 |
| S21 | 0 | 0.058 | 0.012 | 0.019 | 0.123 |
| S22 | 0.108 | 0.029 | 0 | 0.077 | 0.015 |
| S23 | 0.216 | 0.057 | 0.050 | 0.019 | 0 |
| S24 | 0.432 | 0.115 | 0.050 | 0.038 | 0 |
| S25 | 0.108 | 0 | 0.015 | 0.010 | 0.061 |
| S26 | 0 | 0.230 | 0 | 0.077 | 0.061 |
| V1 | 0.109 | 0 | 0.013 | 0.009 | 0.015 |
| V2 | 0.218 | 0.028 | 0.026 | 0 | 0.060 |
| V3 | 0.435 | 0.112 | 0.013 | 0.037 | 0 |
| V4 | 0.109 | 0.028 | 0 | 0.074 | 0.015 |
| V5 | 0.216 | 0 | 0.006 | 0.038 | 0.061 |
| V6 | 0 | 0.057 | 0.025 | 0.010 | 0.031 |

## Results and Discussion

### EEMF Characteristic of Dilute Aqueous Mixtures of Selected Fluorophores

EEMF spectral profiles of the sample S1 is shown in Fig. 1a. It can be observed that apart from the significantly overlapped fluorescence of the fluorophores there are two intense diagonal bands originating due to 1st and 2nd order Rayleigh scattering signals. These bands, clearly marked in Fig. 1a, are quite intense and do not contain any fluorescence related information and must be eliminated from EEMF data sets with a suitable approach before processing them with NNF technique. In the present work, both 1st and 2nd order Rayleigh bands from EEMF spectral data set is eliminated using a home based computational routine that substitute all the variables appearing in the spectral range spanned over (i) $\lambda_{em} \leq \lambda_{ex} + 10$ and (ii) $\lambda_{em} \geq 2^* \lambda_{ex}-10$ with zero. In order to ensure that both the scattering bands are completely eliminated from EEMF, variables within the range of 10 nm right side of the 1st order Rayleigh scattering and left side of the 2nd order Rayleigh scattering bands are also set to zero. Raman scattering is another redundant signal that do not contain any fluorescence information and are often deeply buried in the fluorescence regions and difficult to correct. In the present work, the issue of Raman scattering is minimized using blank subtraction. The Rayleigh and Raman scattering corrected EEMF spectrum of S1 is shown in Fig. 1b. The scattering corrected EEMF data sets can safely be used for further analysis.

### Data Arrangement

EEMF data sets for these 26 samples arranged in a three-way array of dimension 26 × 136 × 19 (Sample × Emission-wavelength × Excitation-wavelength) are suitably unfolded to generate a two way array 26 × 2584 (Sample × (Emission-wavelength × Excitation-wavelength)) for the NNF analysis. The outputs of NNF retrieved spectral profiles are further rearranged to generate EEMF spectral profiles. The data rearrangement is carried out using the home written computational routine.

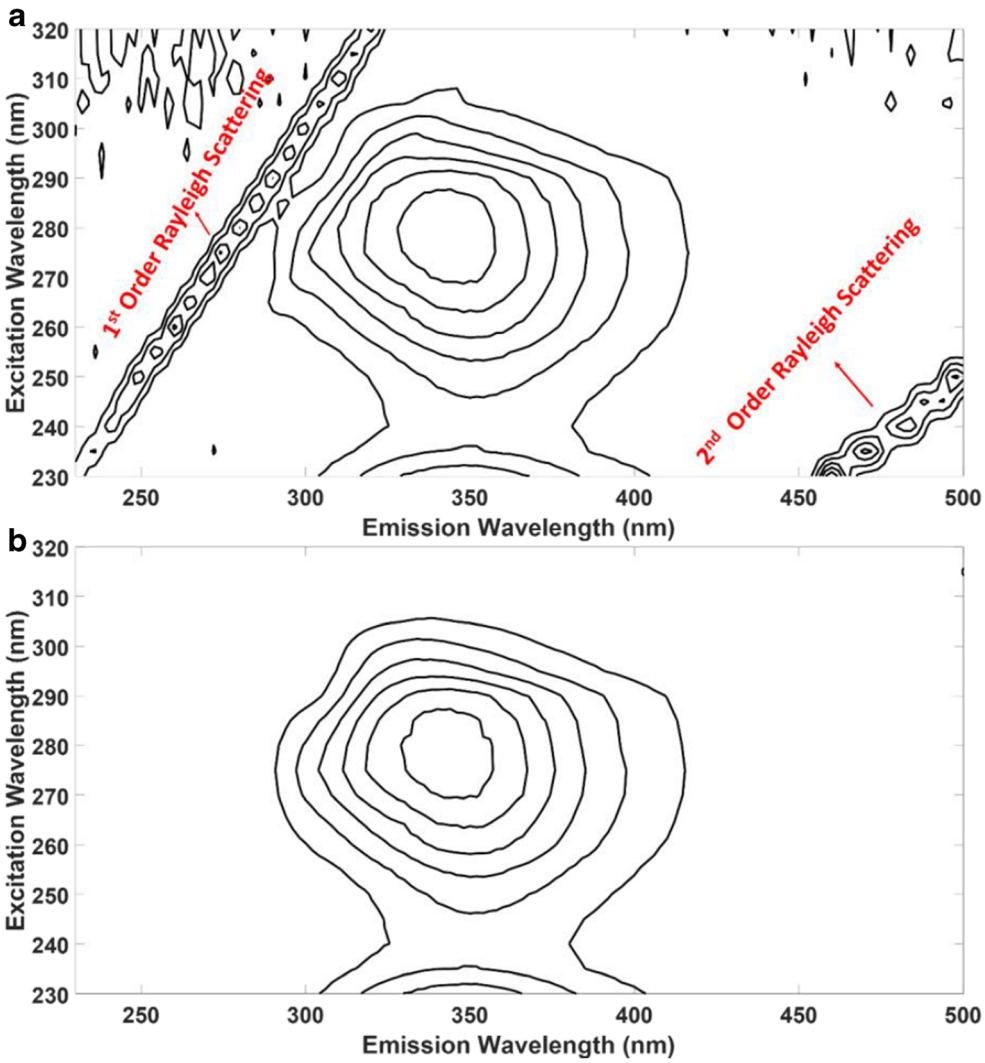


**Fig. 1** **a** Raw EEMF spectra and **b** Processed EEMF spectra of sample S1. The preprocessing involves elimination of Rayleigh and Raman scattering signals. 1st and 2nd order Rayleigh scattering signals are eliminated using a computational routine and Raman scattering signals are mitigated using the blank subtraction

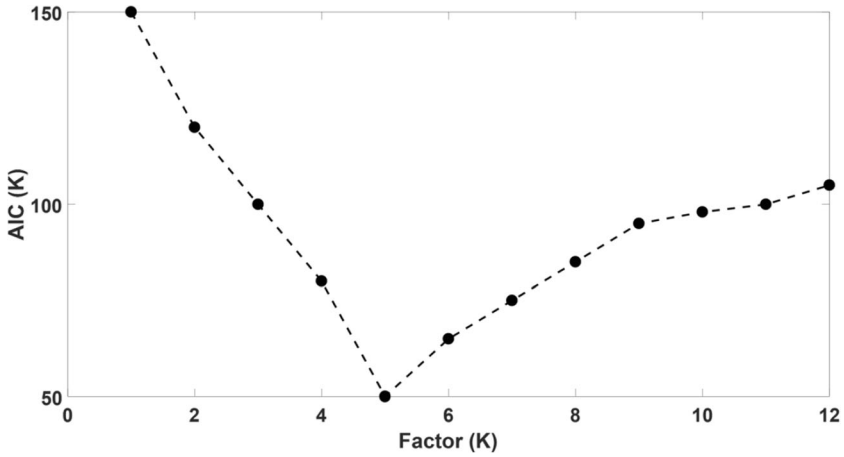


**Fig. 2** AIC (K) versus factor (K) plot. The curve clearly suggest that NNF model with five factor component is optimum to analyze EEMF data sets acquired for dilute aqueous mixtures of Catechol, Hydroquinone, Indole, Tryptophan and Tyrosine

## Akaike Information Criterion (AIC) Analysis for Finding the Optimum Number of Factor

EEMF data set of dimension 26× 2584 is subjected to AIC assisted NNF algorithm. The AIC (K) value is calculated for NNF model developed with different number of factors (K). The obtained AIC (K) values are plotted against K, the obtained curve is shown in Fig. 2. It can clearly be seen that NNF model developed with five factors provides the minimum AIC value. Thus, it must be preferred selected over other models developed with different set of components. The suggested number of components is in fine agreement with the number of fluorescent molecules present in the dilute aqueous mixtures.

## NNF Analysis on EEMF Data Sets

NNF model of five factors (one for each fluorophore) developed on rearranged EEMF data sets of dimension 26 × 2584 are found to make more than 90% fit of the data set. Each of the five factors labeled as Fac1, Fa2, Fac3, Fac4 and Fac5 individually make a fit of 21.5%, 19.3%, 18.5%, 17.6%, and 16.6% respectively. The number of iterations and time elapsed before the convergence for NNF algorithm is found to be 3211 and 141 s. The retrieved two-dimensional EEMF spectral data for Fac1-Fac5 are shown in Fig. 3. In order to establish the identity between the retrieved spectral profiles of Fac1-Fac5 and the experimentally acquired pure spectral profiles of the Catechol, Hydroquinone, Indole, Resorcinol, Tryptophan and Tyrosine a parameter called spectral correlation (SC) is introduced. The parameter is essentially based on the Pearson correlation and it can be calculated using the Eq. 7

$$SC = \frac{\sum_{j=1}^{J}\left(h_{jk}-\overline{h_k}\right)\left(x_{jr}-\overline{x_r}\right)}{\sqrt{\sum_{j=1}^{J}\left(h_{jk}-\overline{h_k}\right)^2}\sqrt{\sum_{j=1}^{J}\left(h_{jk}-\overline{x_r}\right)^2}} \tag{7}$$

In the above equation, $h_k$ and $x_r$ are the $K^{th}$ NNF retrieved and $r^{th}$ experimentally acquired spectral profiles. It is to be noted that both experimentally and NNF retrieved spectral data sets are normalized to unit area. Using the calculated SC values, a color-coded plot, shown in Fig. 4, is developed to find the best match between the Fac1-Fac5 and five

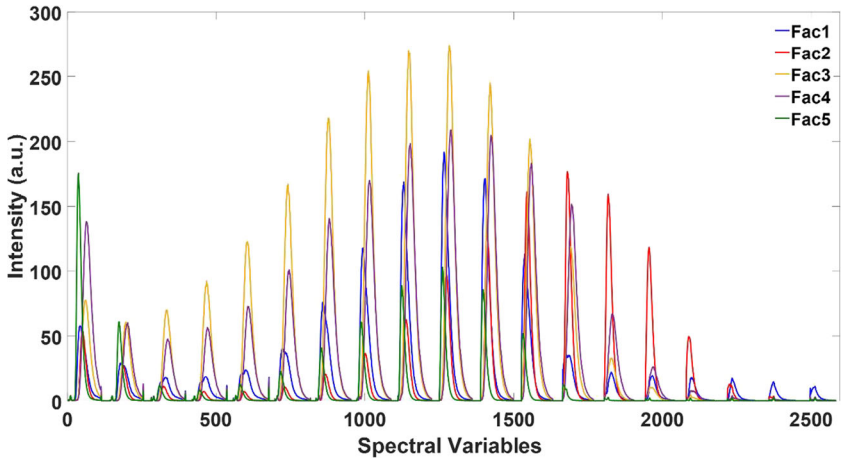


**Fig. 3** NNF retrieved spectral data sets of Fac1-Fac5. The spectral data sets of each of these factors can be reshaped and subjected to further analysis

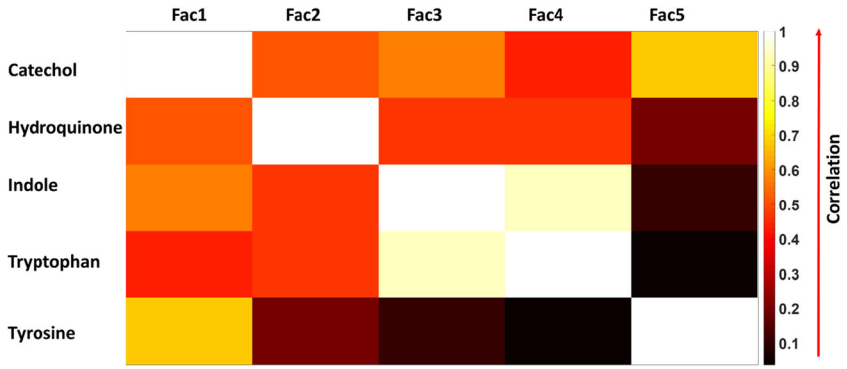


**Fig. 4** Spectral correlation (SC) value based indemnification chart, it clearly shows that best matches for Fac1, Fac2, Fac3 and Fac4 are Catechol, Hydroquinone, Indole, Tryptophan, Tyrosine are respectively

fluorophores. It is found that best matches for the Fac1, Fac2, Fac3, Fac4 and Fac5 are Catechol, Hydroquinone, Indole, Tryptophan and Tyrosine, respectively. The experimental and NNF retrieved EEMF spectral profiles are shown in Fig. 5. The obtained results clearly show that there is perfect match between the experimental and NNF retrieved EEMF spectral profiles of each of the five fluorophores. The obtained results clearly suggest that NNF model allows successful identification of the fluorophores in an automatized manner. Having identified the fluorophores their NNF retrieved

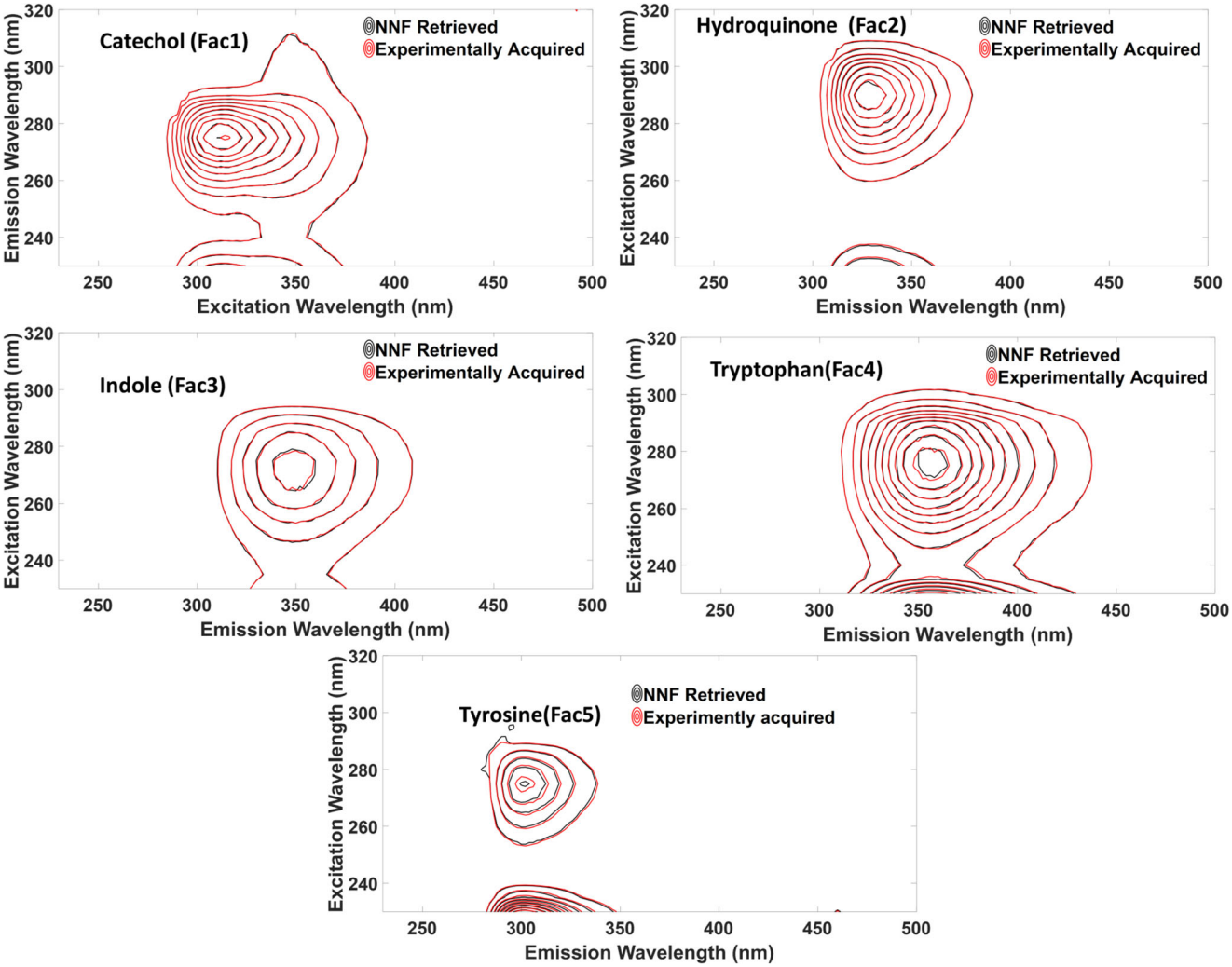


**Fig. 5** NNF and experimentally acquired EEMF spectral profiles Catechol (Fac1), Hydroquinone (Fac2), Indole (Fac3), Tryptophan (Fac4) and Tyrosine (Fac5). It can be seen that there is close correspondence between the experimentally acquired and NNF retrieved EEMF spectral profiles for each of the five fluorophores

**Table 2** The regression eq. Y = m*x + c, relating the actual and predicted concentrations of Catechol (Fac1), Hydroquinone (Fac2), Indole(Fac3), Tryptophan(Fac4) and Tyrosine(Fac5). Y is the predicted concentration and X is the actual concentration, m is slope and c is intercept. $R^2$ is the square of correlation coefficient between the actual and predicted concentrations, RMSEC is measure of calibration error and RMSEP is a measure of prediction error

| Fluorophore | Regression equation | $R^2$ | RMSEC (%) | RMSEP (%) |
|---|---|---|---|---|
| Catechol (Fac1) | $Y = 0.98*X + 1.01*10^{-7}$ | 0.98 | 0.77 | 0.78 |
| Hydroquinone (Fac2) | $Y = 0.99*X + 4.52*10^{-7}$ | 0.99 | 0.35 | 0.55 |
| Indole(Fac3) | $Y = 0.98*X + 4.11*10^{-7}$ | 0.98 | 0.89 | 1.02 |
| Tryptophan(Fac4) | $Y = 0.97*X + 7.18*10^{-7}$ | 0.97 | 1.35 | 1.21 |
| Tyrosine(Fac5) | $Y = 0.98*X + 5.19*10^{-7}$ | 0.98 | 1.01 | 1.04 |

contribution values along with their real concentration values are subjected to PLS modeling with five factors one for each fluorophore. The developed PLS model is found to capture 99% variances of the predictor and predictand data sets. The root mean square of calibration (RMSEC) values [19] are found to be less than 2% for each of the five fluorophores. The precise RMSEC values are summarized in Table 2. The regression equation correlating the actual and predicted concentrations are summarized in Table 2. The square of the correlation coefficients ($R^2$) values [19], reported in Table 2, are found to be close to unity. Each of these statistical parameters clearly suggests that PLS model make precise quantification of the fluorophores in all the samples.

In order to validate the developed procedure a validation set of six samples are used. The experimentally acquired EEMF data sets of these six samples 6 × 136 × 19 are reshaped as a two dimensional data sets 6 × 2584 and are further projected in the space spanned by the pure spectral profiles of Fac1-Fac5. The projection of validation set is summarized in Eq. 8.

$$H_{val} = \frac{(W_{val})^T Y}{(W_{val})^T W_{val}} \text{ where } W_{val} = \frac{Y(H_{cal})^T}{H_{cal}(H_{cal})^T}, \quad (8)$$

In the above equation, Y is the validation EEMF data set. The matrix $H_{val}$ contains the spectral data and the matrix $W_{val}$ contains the contribution values for each of the five factors. The matrix $H_{cal}$ is the spectral profile of each of the five factors of the calibration set. Using the spectral correlation map, identity of each of the five fluorophores is established. The contribution values i.e. $W_{val}$ are subjected to the PLS model already developed for the calibration set. The prediction of the concentrations of the fluorophores of each of the six samples of validation set is found to be in close agreement with their actual concentrations. The root square error of prediction (RMSEP) values [19] for each of the five fluorophores are found to be less than 2%, precise values are reported in Table 2, suggests that developed model can make precise and accurate prediction even for the validation set samples. Thus, obtained results clearly suggest that PLS assisted NNF analysis of EEMF spectral data sets can provide an identification and quantification of the fluorophores in automatized manner.

## Conclusion

The present work provides an analytical procedure by combining NNF assisted PLS analysis with EEMF technique for the identification and quantification of the fluorophores in an automatized manner with minimum user inputs. The utility of the proposed approach is successfully validated using the dilute aqueous mixtures of five fluorophores Catechol, Hydroquinone, Indole, Tryptophan and Tyrosine. The results of the proposed approach are successfully evaluated using various statistical parameters (RMSEC, RSMSEP and $R^2$) each of them found to well within the desired limits.

## References


1. Rho JH, Stuart JL (1978) Automated three-dimensional plotter for fluorescence measurements. Anal Chem 50:620–625
2. Freegarde M, Hatchard CG, Parker CA (1971) Oil spilt at sea: its identification, determination, and ultimate fate. Lab Pr 20:35–40
3. Warner IM, Callis JB, Davidson ER, Goutermann M, Christian GD (1975) Fluorescence analysis: a new approach. Anal Lett 8:665–681
4. Kumar K, TaraiM MAK (2017) Unconventional steady-state fluorescence spectroscopy as an analytical technique for analyses of complex-multifluorophoric mixtures. TrAC Trends Anal Chem 97:216–243
5. Kumar K, Mishra AK (2013) Analysis of dilute aqueous multifluorophoric mixtures using excitation-emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy:a comparative evaluation. Talanta 117:209–220
6. Patra D, Sireesha LK, Mishra AK (2001) Characterization and investigation of polycyclic aromatic compounds present in petrol, diesel, kerosene and 2T oil using excitation emission matrix fluorescence. Indian J Chem 40:374–379
7. Sikorska E, Romaniuk A, Khmelinskii IV, Herance R, Bourdelande JL, Sikorski M, Koziol J (2004) Characterization of edible oils using total luminescence spectroscopy. J Fluoresc 14:25–35
8. Sierra MM, Giovanela M, Parlanti E, Soriano-Sierra EJ (2005) Fluorescence fingerprint of fulvic and humic acids from varied origins as viewed by single-scan and excitation/emission matrix techniques. Chemosphere 58:715–733
9. Baker A (2001) Fluorescence excitation - Emission matrix characterization of some sewage-impacted rivers. Environ Sci Technol 35: 948–953
10. Brewer M, Utzinger U, Silva E, Gershenson D, Bast RC, Follen M, Richards-Kortum R (2001) Fluorescence spectroscopy for in vivo characterization of ovarian tissue. Lasers Surg Med 29:128–135

11. Sterenborg HJCM, Motamedi M, Wagner J, Duvic M, Thomsen S, Jacques SL (1994) In vivo fluorescence spectroscopy and imaging of human skin tumors. Lasers Med Sci 9:191–201
12. Kumar K, Mishra AK (2012) Application of parallel factor analysis to total synchronous fluorescence spectrum of dilute multifluorophoric solutions: addressing the issue of lack of trilinearity in total synchronous fluorescence data set. Anal Chim Acta 755:37–45
13. Kumar K, Kumar Mishra A (2015) Parallel factor (PARAFAC) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets in excitation-emission matrix fluorescence (EEMF) layout: certain practical aspects. ChemomIntell Lab Syst 147:121–130
14. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791
15. Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. Adv Neural Inf Process Syst:556–562
16. Paatero P, Tapper U (1994) Positive matrix factorization: a non???Negative factor model with optimal utilization of error estimates of data values. Environmetrics 5:111–126
17. Paatero P (1997) Least squares formulation of robust non-negative factor analysis. Chemom Intell Lab Syst 37:23–35
18. Devarajan K (2008) Nonnegativematrix factorization: an analytical and interpretive tool in computational biology. PLoS Comput Biol 4:e1000029
19. Kramer R (1998) Chemometric techniques for quantitative analysis. Marcel Dekker, New York
20. Wold S, Ruhe A, Wold H, Dunn WJ (1984) The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J Sci and Stat Comp 5:735–743
21. Geladi P, Kowalski B (1986) Partial least square regression: a tutorial. Anal Chim Acta 185:1–17
22. Lorber A, Wangen LE, Kowalski BR (1987) A theoretical foundation for the PLS algorithm. J Chemom 1:19–31
23. Varmuza K, Filzmoser P (2008) Introduction to multivariate statistical analysis in chemometrics. Taylor & Francis Group, Boca Raton
24. Hanselmann M, Kirchner M, Renard BY, Amstalden ER, Glunde K, Heeren RM, Hamprecht FA (2008) Concise representation of mass spectrometry images by probabilistic latent semantic analysis. Anal Chem 80:9649–9658
25. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723
26. Burnham KP, Anderson DR (2002) Model selection and multimodel inference, a practical-theoretic approach. Springer, New York
27. Kumar K (2019) Application of Akaike information criterion assisted probabilistic latent semantic analysis on non-trilinear total synchronous fluorescence spectroscopic data sets: automatizing fluorescence based multicomponent mixture analysis. Anal Chim Acta 1062:60–97
28. Bro R, Rinnan Å, Faber NM (2005) Standard error of prediction for multilinear PLS 2. Practical implementation in fluorescence spectroscopy. Chemom Intell Lab Syst 75:69–76
29. Rinnan Å (2004) Application of PARAFAC on spectral data. Royal Veterinary and Agricultural University-DK, (PhD Thesis