Journal of Fluorescence (2019) 29:185–193
https://doi.org/10.1007/s10895-018-2327-z

ORIGINAL ARTICLE

Discrete Wavelet Transform (DWT) Assisted Partial Least Square (PLS) Analysis of Excitation-Emission Matrix Fluorescence (EEMF) Spectroscopic Data Sets: Improving the Quantification Accuracy of EEMF Technique

Keshav Kumar[1]

Received: 28 September 2018 / Accepted: 19 November 2018 / Published online: 28 November 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
In the present work, it is shown that quantitative estimation efficiency of the partial least square (PLS) calibration model can be significantly improved by pre-processing the EEMF with discrete wavelet transform (DWT) analysis. The application of DWT essentially reduces the volume of data sets retaining all the analytically relevant information that subsequently helps in establishing a better correlation between the spectral and concentration data matrices. The utility of the proposed approach is successfully validated by analyzing the dilute aqueous mixtures of four fluorophores having significant spectral overlap with each other. The analytical procedure developed in the present study could be useful for analyzing the environmental, agricultural, and biological samples containing the fluorescent molecules at low concentration levels.

Keywords Excitation-emission matrix fluorescence · Partial least square analysis · Wavelet analysis · Discrete wavelet analysis · Fluorophores

Introduction

Over the years, the technological developments have provided the desired sensitivity and selectivity to the modern fluorimeters that essentiality has made it possible to analyze the mixtures of fluorophores in a simple and swift manner. Excitation-emission matrix fluorescence (EEMF) spectroscopy is the most commonly used fluorescence technique [1–5]. EEMF has been successfully used for analyzing the samples of biological, clinical, pharmaceutical, petrochemicals and environmental origin [1–5]. The popularity of EEMF technique as an analytical tool can be attributed to the following two reason (i) most of the fluorimeters have the required software that allows the acquisition of the EEMF spectrum without much inputs form the user and (ii) it allows capturing the fluorescence response of all the fluorophores in a single spectrum [1–5]. EEMF spectrum essentially consists of emission and excitation spectra acquired at different excitation and emission wavelengths, respectively. EEMF technique serve as a useful tool to "fingerprint" the fluorescent molecules in a reliable manner.

It is true that with the recent advancement in the spectrofluorometric technology the acquisition of EEMF spectral profiles have become much easier. However, the analyses of large volume of EEMF data sets remain a computational and interpretive challenge. To make meaningful interpretation in a swift manner, over the years several chemometric approaches have been successfully used. Partial least square (PLS) algorithm [6–10] is one of the most commonly used chemometric technique that has been used with EEMF to develop a reliable and robust calibration model for quantifying the analytes without involving any pre-separation step. The PLS algorithm search for the significant factors that describes the maximum variance of the spectral datasets and maximizes the correlation between the spectral and concentration information of the fluorophores. The PLS modeling requires that set of (i) spectral variables and (ii) latent variables be optimized. The optimizations of these two parameters are interrelated, as a result, it becomes a time consuming and conceptually and computationally challenging task. The optimization is mainly targeted towards minimizing the error of prediction for the

✉ Keshav Kumar
keshavkumar29@gmail.com

[1] Institute for Wine analysis and Beverage Research, Hochschule Geisenheim University, 65366 Geisenheim, Germany

concentrations of fluorophores in both calibration and validation set samples. The optimization of the latent variables can be simplified to great extent provided optimum set of analytically useful spectral variables can be selected with a suitable variable selection approach.

The optimization of the spectral variables and subsequent latent variables selection towards developing a robust calibration model becomes difficult with the increase in the number of spectral variables. To deal with this, over the years, a number of variable selection approaches that helps in finding the set of analytically useful spectral variables have been introduced in the area of the chemometrics. However, with the usage of these techniques their limitations have also been experienced that limits their application to a great extent. For example, the approaches such as variable influence on projection (VIP) method, PLS-Beta method and selectivity ratio (SR) method are found to have limited applications [11–14]. It is mainly so because these methods carries independent evaluation of each variables towards fitting the PLS modeling whereas the spectral data sets are highly correlated. The Genetic Algorithm (GA) [14–23] that is essentially inspired from Darwin theory of "survival of the fittest" attempts to find the set of most informative spectral variables that can safely be processed for developing the robust PLS calibration model. However, the application of GA is found to suffer on the practical grounds because it involves a heavy load of calculation. It mainly arises because successful application of GA demands optimization of several parameters such as (i) Population size, (ii) Maximum generation, (iii) Mutation rate, (iv) Window width, (v) Convergence, (vi) Initial terms and (vii) Cross-over term [14–23]. Thus, it can be realised that an approach that significantly reduces the computational burden and allows swift selection of the analytically useful variables is desirable as a pre-processing step to speed up the analysis.

Over the years, wavelet analysis has found its application in the fields of analytical and bio-analytical chemistry [24–40]. Wavelets can be expressed as oscillating waves that has the zero moments [24–40]. Wavelet analysis essentially allows the transformation of a signal from one domain to other [24–40]. For example, wavelet analysis can efficiently represent a signal from time domain to frequency domain. Wavelet analysis has an advantage that it ensures that processed signals are localised in both time and frequency scales [24–40]. Wavelet transformation can mainly be carried out with two different approaches (i) discrete wavelet transformation (DWT) and (ii) continuous wavelet transformation (CWT) analysis [24–40]. The DWT analysis is preferred over the CWT because most of the instruments including the fluorimeters generate the data sets that are discrete in nature. DWT has been successfully used to analyse the chromatographic, spectroscopic and mass-spectrometric data sets acquired for different kind of samples [41–45]. DWT analysis with a suitable mother wavelet (i.e. basis function) decomposes a signal or spectrum at different levels into its high and low frequency component. High frequency component essentially captures the noise component of the signal and low frequency component captures the spectral features of the signal. With each level of decomposition, DWT analysis involves reduction of the number of data points by half without losing any analytical information. Thus, it can be realised that application of DWT on EEMF spectral data sets with a suitable mother wavelet can significantly reduce the volume of data sets and this will subsequently reduce the computational burden in optimising the various parameters required to develop a robust PLS calibration model.

The objective of the present work is to assess if the application of DWT analysis on EEMF can help the PLS algorithm to maximise the correlation between the spectral and concentration data matrices. In other words to see if the application of DWT analysis on EEMF data set as pre-processing technique improve the quantification efficiency of the PLS calibration model. To carry out the present study the EEMF spectral data sets acquired for dilute aqueous mixtures of four fluorophores namely Catechol, Hydroquinone, Indole and Tryptophan are taken as the test case. It is believed that proposed analytical procedure could be useful for analyzing the environmental, agricultural, and biological samples containing the fluorescent molecules at low concentration levels.

Theory

DWT Analysis

As discussed above, DWT analysis [24–40] decomposes a signal S(t) as a sum of two series that are complementary of each other. One of the series contains the approximation (A) coefficients and the other series describes the detailed (D) coefficients of the decomposed signal. The coefficients A and D represent the signal and noise components of the original signal, respectively. The approximation (A) coefficients are obtained by convoluting the signal s(x) with discrete scaled-shifted scaling (Φ) function. The detailed (D) coefficients are obtained by convoluting the s(x) with discrete scaled-shifted wavelet (Ψ) function. Mathematically, the calculation of approximation (A) and detailed (D) coefficients can be summarised using the set of eqs. 1–4.

$$A(j,k) = \sum_{x=0}^{X-1} s(x)\Phi_{j,k}(x) \quad (1)$$

$$\Phi_{j,k}(x) = 2^{\frac{j}{2}}\Phi\left(2^{j}x-i\right), j = 0.....i = 0...., 2^{j}-1 \quad (2)$$

$$D(j,k) = \sum_{x=0}^{X-1} s(x)\Psi_{j,k}(x) \quad (3)$$

$$\Psi_{j,k}(x) = 2^{\frac{j}{2}}\Psi\left(2^{j}x-i\right), j = 0.....i = 0...., 2^{j}-1 \quad (4)$$

It is important to mention that above convolution steps are followed by the dyadic decimation (i.e. down sampling). The approximation coefficients can further be convoluted with discrete scaled-shifted scaling (Φ) function to obtain

approximation and decomposition coefficients at next level. The procedure can be repeated till the desired level of decomposition of the signal is achieved. The above-discussed DWT based decomposition of the signal is summarised in Fig. 1. Some of the commonly used mother wavelets that can be used to convolute the signals are haar, db2-db10, sym1-sym8 and coif1-coif5. These wavelets differ from each other in their properties (i) number of vanishing moments, (ii) symmetry, (iii) period etc. It is to be noted that each of these wavelet functions are appropriate for a certain applications. A detailed theoretical discussion of DWT can be seen elsewhere in the literature.

PLS Analysis

In the present work, the PLS algorithm is briefly discussed. A detailed description of PLS algorithm can be seen elsewhere in the literature [6–10]. The PLS model essentially finds a set of orthogonal latent variables that can maximise the covariance between the spectral and concentration data matrix. Mathematically, PLS algorithm can be summarised using the given below eqs. (5–7).

$$X = TP^T + E_1 \quad (5)$$

$$Y = QU^T + E_2 \quad (6)$$

$$U = TB \quad (7)$$

In the above equations, X is the spectral data matrix of dimension I × J where I is the number of samples and J is the number of spectral variables. T is the score matrix of dimension I × F where F is the number of latent variables. P is the loading matrix of dimension J × F. E_1 is the residual matrix of the dimension I × J. The matrix Y is the concentration matrix of dimension I × N where N is the number of analytes (or fluorophores) in the present work, number of analytes and latent variables are equal to each other. Q is the score matrix of dimension I × F. U is the loading matrix of dimension N × F. E_2 is the residual matrix of the dimension I × N. B is the regression matrix of dimension F × N.

Material and Methods

EEMF Spectral Data Sets

In the present work, a calibration set of 25 samples and a validation set of five samples are created. Both calibration and validation set samples contains varying concentration of Catechol, Hydroquinone, Indole and Tryptophan. The samples of calibration set are labelled as S1-S25 and the samples of the validation set are labelled as V1-V5. The concentration of each of the fluorophores in the all the samples of the calibration and validation sets are summarised in Table 1. The instruments and various instrumental parameters used to acquire the EEMF data sets are summarised in Table 2. A more detailed description of the sample preparation and data acquisition can be seen in the literature reported by Bro and co-workers [46, 47].

Computational Platform

The computational work of the present study is carried out on MATLAB platform. All the graphic plots are created using the suitable MATLAB inbuilt commands.

Results and Discussion

DWT Analysis of EEMF Data Sets: Finding a Suitable Wavelet Basis

The Rayleigh and Raman corrected EEMF data sets of calibration sets are rearranged to generate a two away array of dimension 25× 2584 (where 25 (=sample) × (136 (=emission wavelength) × 19 (=excitation wavelength)). The Rayleigh and Raman corrected EEMF data sets of validation sets are rearranged to generate a two-way array of dimension 5× 2584. The unfolded-EEMF profile of calibration set is shown in Fig. 2. As discussed earlier, the objective of the present work is to explore the utility of DWT analysis towards (i) reducing the computational burden and (ii) finding the few most informative variables that can be used for developing a robust calibration model for analysing the dilute aqueous mixtures of fluorophores present in the mixture. Towards this, we need to find a suitable mother wavelet and the appropriate level of decomposition that can be used for carrying out the analysis. The selection of the mother wavelets need to be carried out in a legitimate manner. To achieve this, a parameter called Energy to Shannon Entropy ratio (R) is calculated using a set of eqs. (8) for each spectrum of the calibration set

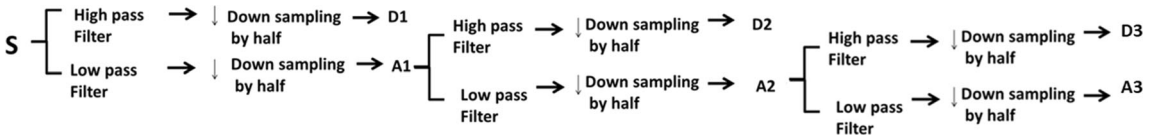

Fig. 1 Schematic of DWT analysis. DWT decompose the signal S at third level and provides A1-A3 approximation and D1-D3 detail coefficients

Table 1 The concentrations of Tryptophan, Hydroquinone, Indole and Catechol in the calibration set samples S1-S25 and validation set samples V1-V5

Sample	Tryptophan ($\times 10^{-4}$)M	Hydroquinone ($\times 10^{-4}$)M	Indole ($\times 10^{-4}$)M	Catechol ($\times 10^{-4}$)M
S1	0.0372	0.0559	0.0128	0.2176
S2	0.0558	0.2238	0.0384	0.8704
S3	0.0384	0.2296	0	0
S4	0.0384	0.1148	0.0125	0
S5	0.0372	0.1119	0	0.4352
S6	0	0.1678	0.0384	0.6528
S7	0	0.1119	0.0384	0.4352
S8	0.0744	0.0559	0.0512	0.2176
S9	0.0384	0	0.0499	0
S10	0	0.2296	0	0.1079
S11	0	0.2238	0	0.8704
S12	0.0558	0.1119	0.0128	0.4352
S13	0	0	0.0256	0
S14	0.0558	0	0.0384	0
S15	0.0186	0.2238	0.0128	0.8704
S16	0.0186	0.2238	0.0512	0.8704
S17	0.0744	0.2238	0.0256	0.8704
S18	0.0744	0	0.0512	0.8704
S19	0.0372	0.1678	0	0.6528
S20	0.0372	0.1119	0.0128	0.4352
S21	0.0372	0.1119	0.0512	0.4352
S22	0.0786	0.1150	0.2490	0.8630
S23	0.0192	0.0574	0.0499	0.2176
S24	0.0186	0.0559	0.0512	0.2176
S25	0.0186	0.1119	0.0256	0.4352
V1	0.0744	0.1678	0.0128	0.6528
V2	0.0744	0.0559	0.0256	0.2176
V3	0.0744	0.0559	0.0384	0.2176
V4	0.0558	0.1678	0.0512	0.6528
V5	0.0186	0.0559	0	0.2176

at various level of decomposition [31, 45]. The wavelet function with appropriate level of decomposition maximising the Energy to Shannon Entropy ratio is used to analyse the spectral data sets as.

$$R = \frac{Energy}{Shannon\ Entropy} = \frac{\sum_{i=1}^{N} |w(s,i)|^2}{-\sum_{i=1}^{N} p_i.log2.p_i} \quad (8)$$

In the above eqs. N is the number of wavelet coefficients, *w(s,i)* represents the wavelet coefficients and p_i is the energy probability distribution of the wavelet coefficients.

The energy and Shannon entropy is calculated for the spectral profiles of each of the 25 samples of the calibration set using each of these 22 wavelets i.e. Haar, db2-db10, sym2-sym8 and coif1-coif5 at various decomposition levels. It is

Table 2 Various parameter used for acquiring the EEMF data on Varian fluorimeter

Instrumental Parameter	Value
Slit width (Excitation and Emission)	5 nm
PMT Voltage	600 V
Scan rate	1920 nm/min
Excitation Wavelength	230–320 nm with a step size of 5 nm
Emission Wavelength	230–500 nm with a step size of 2 nm

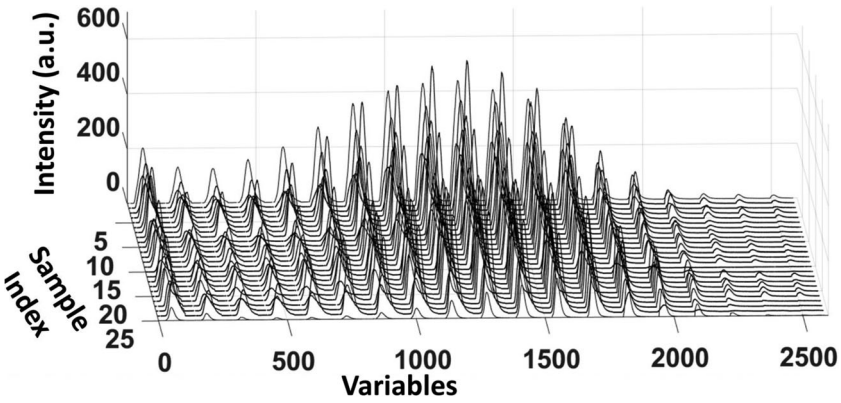

Fig. 2 Unfolded-EEMF spectral profiles of the calibration set samples S1-S25

found that for each of the sample, db4 wavelet at 3rd level of decomposition maximises the R-values. Hence, db4 with third level of decomposition is chosen as the mother wavelet carrying out the analysis. For each spectrum of the calibration set, the obtained Energy, Shannon Entropy and their ratio (R) is reported in Table 3. The db4 wavelet decomposes each spectrum in three different stages. In the first stage, the db4 wavelet decomposes the spectrum into A1 (approximation) and D1 (detail) coefficients. In the second stage, the db4 wavelet decomposes the A1 into A2 and D2 coefficients. In the final step of the analysis, the db4 wavelet decomposes A2 coefficients and provides the A3 and D3 coefficients. The A3 coefficients approximate the spectral data set and D1-D3 describes the noise content or irrelevant portion of the spectral profiles. The db4 based decomposition of the spectra into approximation (A3) and detailed (D1-D3) coefficients are shown in Fig. 3. At each level of decomposition, the spectral data reduces by half. Therefore, with the application of the wavelets, instead of 2584 variables one has to process only

Table 3 Energy, Shannon-Entropy and R (Energy to Shannon-Entropy ratio) obtained from the db4 wavelet analysis with third level of decomposition of unfolded-EEMF data sets of S1-S25 samples

Sample	Energy	Shannon-Entropy	R = Energy/Shannon- Entropy
S1	99.8147236863932	−499.242259869422	−0.199932441040748
S2	99.9057919370756	−499.256867531875	−0.200108998862589
S3	99.1269868162935	−499.607772980466	−0.198409616857921
S4	99.9133758561390	−499.344522109822	−0.200089059621575
S5	99.6323592462254	−499.828813312188	−0.199332964792480
S6	99.0975404049994	−499.293953911980	−0.198475346293637
S7	99.2784982188670	−499.968167153623	−0.198569638511330
S8	99.5468815192050	−499.723077015039	−0.199204091421636
S9	99.9575068354343	−499.953828609369	−0.199933476084118
S10	99.9648885351993	−499.902868218764	−0.199968623687619
S11	99.1576130816776	−499.176542171673	−0.198642373398180
S12	99.9705927817606	−499.305171377024	−0.200219421933992
S13	99.9571669482429	−499.682900519939	−0.200041199817392
S14	99.4853756487228	−499.049777951162	−0.199349604075886
S15	99.8002804688888	−499.965553919497	−0.199614312799154
S16	99.1418863386272	−499.561255640344	−0.198457917260909
S17	99.4217612826263	−499.618441542907	−0.198995379305045
S18	99.9157355251891	−499.234483211851	−0.200137888878140
S19	99.7922073295596	−499.204800098863	−0.199902339300016
S20	99.9594924263929	−499.813127395446	−0.199993731551805
S21	99.6557406991566	−499.510235604212	−0.199506904155051
S22	99.0357116785742	−499.554413799289	−0.198248096589463
S23	99.8491293058688	−499.353686989889	−0.199956727881115
S24	99.9339932477576	−499.290635169142	−0.200151948000994
S25	99.6787351548578	−499.245313318018	−0.199658830029643

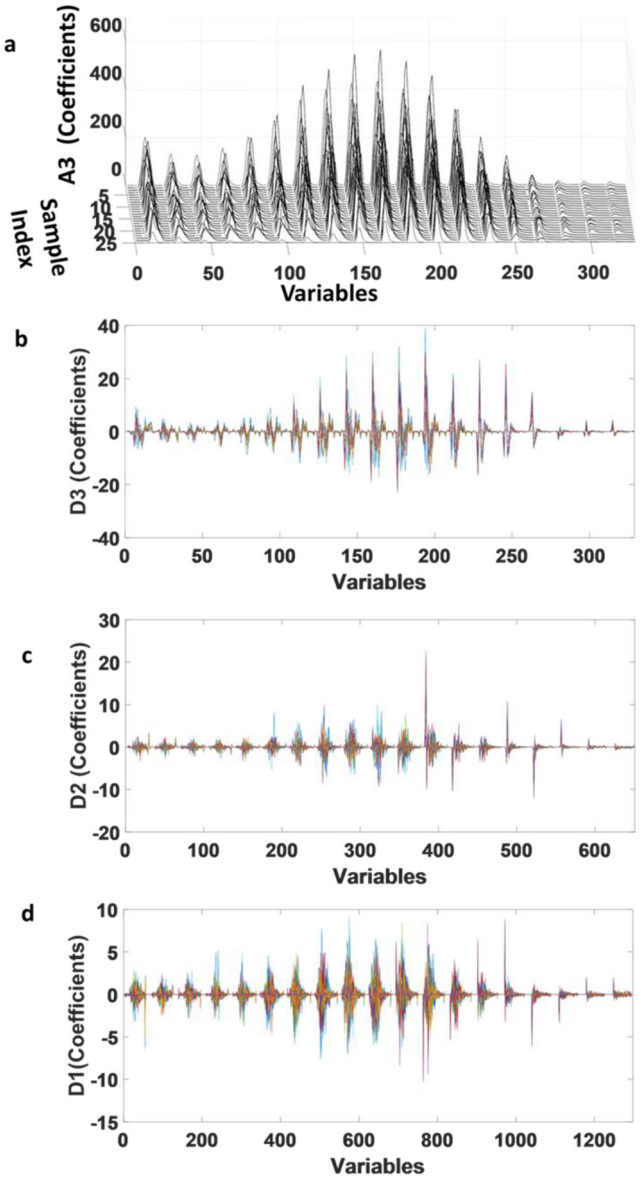

Fig. 3 The db4 decomposition of the spectral profiles at 3rd level of decomposition. **a** The approximation coefficient (A3) obtained at 3rd level, (**b**) The detail coefficient (D3) obtained at 3rd level, (**c**) The detail coefficient (D2) obtained at 2nd level and (**d**) The detail coefficient (D1) obtained at 1st level

329 (A3 coefficients) variables per sample to develop PLS calibration model.

PLS Calibration Analysis on db4 Wavelet Processed Spectral Data Sets

The A3 coefficients retrieved from the db4 wavelet decomposition of the spectral data sets of calibration set are arranged in a matrix X of dimension 25 × 329 and the concentration of each of the four fluorophores are arranged in a matrix Y of dimension 25 × 4. The next step important thing before carrying out the PLS analysis is that one must ensure that mathematical rank of the data set are in correlation with the number of fluorophores present in the samples of calibration set. In other words, in an ideal case the chemical and mathematical rank must be equal to each other. To verify this, the data matrix X is subjected to principal component analysis (PCA) [6]. In this approach, factors explaining more than 2% variance of the spectral data sets are considered significant and must be retained to carry out any multivariate analysis. In the present case, as shown in Fig. 4, the first four factors of the PCA model are cumulatively found to explain >99% variance of the spectral data sets suggesting there are four independent source of variation. The PCA clearly verifies that mathematical rank of the DWT processed spectral data set X correlates well with number of fluorophores present in the calibration set. Thus, PLS analysis must be carried out with four latent variables one for each fluorophore.

The PLS analysis on the db4 processed spectral data sets 25 × 329 (sample × (emission-wavelength× excitation-wavelength)) and concentration data matrix of dimension 25 × 4 (sample × number of fluorophores) are carried out with 4 latent variables. The PLS model is found to explain more than 99.5% variance of both spectral and concentration data matrices. The regression equations relating the actual and predicted concentrations of the analysed fluorophores are summarised in Table 4. The statistical parameters such as root mean square error of calibration (RMSEC) [9, 48] and square of correlation coefficients (R^2) [9, 48] are also reported in Table 4 for each fluorophore. The developed model is further applied on the db4 processed spectral data sets of the validation set and root mean square error of prediction (RMSEP) value [9, 48] is calculated for each of the four fluorophores. The obtained RMSEP values are reported in Table 4. The R^2 values are found to be close unity suggesting that actual and regressed concentration have close correspondence with each other. The RMSEC and RMSEP values are less than 1.0% and 1.5%, respectively for each fluorophores. By evaluating all the statistical parameters obtained from the PLS analysis, it can easily be inferred that developed calibration model makes a precise and accurate estimation of the concentration for each of the four fluorophores in all the samples of calibration and validation sets. The obtained results clearly show that pre-processing the EEMF spectral data sets with a suitable wavelet can substantially reduce the data size while retaining all the analytically relevant spectral variables for developing a robust PLS calibration model that can make accurate and precise concentration estimation for each fluorophore of the dilute mixtures.

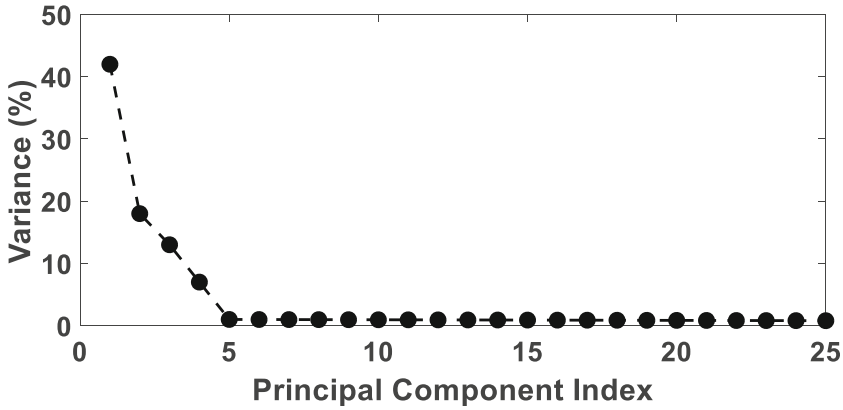

Fig. 4 Amount of variance (%) explained by the different principal components. The results clearly suggest there are four significant factors in the analysed unfolded-EEMF data sets

In order to show the advantage of the DWT assisted PLS analysis of EEMF data sets over traditional way of carrying out PLS analysis. The unfolded-EEMF data sets of dimension 25× 2584 (sample × emission wavelength × excitation wavelength) and concentration data matrix of dimension 25 × 4 (sample × number of fluorophores) are subjected to PLS analysis with four latent variables. The developed PLS model explains ~81% variance of the spectral data set and ~80% variance of the concentration data matrix. The regression equations relating the actual and predicted concentrations of the fluorophores are summarised in Table 4. The RMSEC and R^2 are also reported in Table 4. The RMSEP values obtained for each of the four fluorophores present in the validation set samples are also summarized in Table 4. The comparison of these results clearly suggests that with the application of DWT analysis as a pre-processing technique it is possible to significantly improve the quantitative estimation efficiency of PLS calibration model. For example, with application of DWT analysis he RMSEC and RMSEP values of the PLS model reduces from >12% and > 15% to <1.5% and < 2%, respectively for each of the four fluorophores. Similarly, with the application of DWT analysis the R^2 values improves from <0.85 to >0.99 for each fluorophore. Thus, with the obtained results it can be easily inferred that DWT assisted PLS analysis can be an analytically useful approach for analysing the large volume of highly correlated spectral data sets. The DWT analysis reduces the volume of data sets and retains the analytically useful variables for the subsequent PLS analysis.

Conclusions

The present work shows that analytical accuracy of the PLS calibration model can be significantly improved by pre-processing EEMF spectral data sets with DWT analysis. The application of DWT analysis essentially reduces the volume of data sets retaining all the analytically relevant information. It subsequently helps PLS algorithm in establishing a better correlation between the spectral and concentration data matrices. The proposed DWT assisted PLS analysis approach is successfully tested by analysing EEMF data sets of dilute aqueous mixtures of Catechol, Hydroquinone, Indole and Tryptophan. The proposed approach is also compared with traditional way of carrying out the PLS model. The obtained results of clearly suggests that application of DWT analysis as pre-processing technique significantly improves the analytical efficiency of the PLS analysis of the EEMF data sets. The developed analytical procedure can provide a simple means for analyzing the environmental, agricultural, or biological samples containing the fluorescent molecules at low concentration levels.

Table 4 Comparing calibration models developed using DWT assisted PLS analysis and traditional PLS analysis of EEMF data sets

Models	Fluorophore	Regression Equation	R^2	RMSEC (%)	RMSEP (%)
DWT assisted PLS analysis	Catechol	$Y = 0.998*X + 3.98*10^{-7}$	0.99	0.52	1.21
	Hydroquinone	$Y = 0.998*X + 2.28*10^{-7}$	0.99	0.43	1.25
	Indole	$Y = 0.997*X + 0.85*10^{-7}$	0.99	0.49	1.32
	Tryptophan	$Y = 0.991*X + 2.54*10^{-7}$	0.99	0.75	1.37
Traditional PLS analysis	Catechol	$Y = 0.828*X + 8.96*10^{-6}$	0.83	17.29	17.96
	Hydroquinone	$Y = 0.836*X + 7.98*10^{-6}$	0.82	14.89	18.88
	Indole	$Y = 0.785*X + 8.65*10^{-6}$	0.78	13.10	18.37
	Tryptophan	$Y = 0.801*X + 7.98*10^{-6}$	0.80	12.62	19.10

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, Tarai M, Mishra AK (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. Kramer R (1998) Chemometric techniques for quantitative analysis. Marcel Dekker, New York
7. 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
8. Geladi P, Kowalski B (1986) Partial least square regression: a tutorial. Anal Chim Acta 185:1–17
9. Lorber A, Wangen LE, Kowalski BR (1987) A theoretical foundation for the PLS algorithm. J Chemom 1:19–31
10. Varmuza K, Filzmoser P (2008) Introduction to multivariate statistical analysis in chemometrics. Taylor & Francis Group, Boca Raton
11. Chong IG, Jun CH (2005) Performance of some variable selection methods when multicollinearity is present. Chemom Intell Lab Syst 78:103–112
12. Rajalahti T, Arnebergc R, Bervend FS, Myhra KM, Ulvikd RJ, Kvalheimg OM (2009) Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. Chemom Intell Lab Syst 95: 35–48
13. Hocking RR (1976) The analysis and selection of variables in linear regression. Biometrics 32:1–49
14. Fujiwara K, Sawada H, Kano M (2012) Input variable selection for PLS modelling using nearest correlation spectral clustering. Chemom Intell Lab Syst 118:109–119
15. Sorol N, Arancibia E, Bortolato SA, Olivieri AC (2010) Visible/near infrared-partial least-squares analysis of brix in sugar cane juice A test field for variable selection methods. Chemom Intell Lab Syst 102:100–109
16. Goicoechea HC, Olivier AC (2003) A new family of genetic algorithms for wavelength interval selection in multivariate analytical spectroscopy. J Chemom 17:338–345
17. Mehmood T, Liland KH, Snipen L, Sæbø S (2012) A review of variable selection methods in partial least squares regression. Chemom Intell Lab Syst 118(2012):62–69
18. Shaffer RE, Small GW (1996) Genetic algorithm-based protocol for coupling digital filtering and partial least-squares regression: application to the near-infrared analysis of glucose in biological matrices. Anal Chem 68:2663–2675
19. Ding Q, Small GW (1998) Genetic algorithm-based wavelength selection for the near-infrared determination of glucose in biological matrixes: initialization strategies and effects of spectral resolution. Anal Chem 70:4472–4479
20. Bangalore AS, Shaffer RE, Small GW (1996) Genetic algorithm based method for selecting wavelengths and model size for use with partial least-squares regression: application to near-infrared spectroscopy. Anal Chem 68:4200–4212
21. Xiaobo Z, Jiewen Z, Hanpin M, Jiyong S, Xiaopin Y, Yanxiao L (2010) Genetic algorithm interval partial least squares regression combined successive projections algorithm for variable selection in near-infrared quantitative analysis of pigment in cucumber leaves. Appl Spec 64:786–794
22. Leardi R, González AL (1998) Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemom Intell Lab Syst 41:195–207
23. Arakawa M, Yamashita Y, Funatsu K (2010) Genetic algorithm-based wavelength selection method for spectral calibration. J Chemom 25:10–19
24. Walczak B, D. L. Massart DL (1997) Wavelets-something for analytical chemistry?. TrAC Trends Anal Chem 16:451–463
25. Shao X, Cai W, Pan Z (1999) Wavelet transform and its applications in high performance liquid chromatography HPLC analysis. Chemom Intell Lab Syst 45:249–256
26. Shao X, Cai W, Sun P, Zhang M, Zhao G (1997) Quantitative determination of the components in overlapping chromatographic peaks using wavelet transform. Anal Chem 69:1722–1725
27. Labat D (2005) Recent advances in wavelet analyses: part 1. A review of concepts. J Hydrol 314:275–288
28. Alsberg BK, Woodward AM, Kell DB (1997) An introduction to wavelet transforms for chemometricians: a time-frequency approach. Chemom Intell Lab Syst 37:215–239
29. Perrin C, Walczak B, Massart DL (2001) The use of wavelets for signal Denoising in capillary electrophoresis. Anal Chem 73:4903–4917
30. Pasti L, Walczak B, Massart DL, Reschiglian P (1992) Optimization of signal denoising in discrete wavelet transform. Chemom Intell Lab Syst 48:21–34
31. Gao RX, Yan R (2011) Wavelets: theory and application for manufacturing. Springer, New York
32. Shao XG, Leung AKM, Chau FT (2003) Wavelet: a new trend in chemistry. Acc Chem Res 36:276–283
33. Chau FT, Liang YZ, Gao J, Shao XG (2004) Chrmometrics: from basic to wavelet transform. John Wiley & Sons, Inc., Hoboken
34. Ehrentreich F (2002) Wavelet transform applications in analytical chemistry. Anal Bioanal Chem 372:115–121
35. Barache D, Antoine JP, Dereppe JM (1997) The continuous wavelet transform, an analysis tool for NMR spectroscopy. J Magn Reson 128:1–11
36. Barclay VJ, Bonner RF (1997) Application of wavelet transforms to experimental spectra: smoothing, Denoising, and data set compression. Anal Chem 69:78–90
37. Zhang ZM, Chen S, Liang YZ (2011) Peak alignment using wavelet pattern matching and differential evolution. Talanta 83:1108–1117
38. Chourasia VS, Mittra AK (2009) Selection of mother wavelet and Denoising algorithm for analysis of Foetal Phonocardiographic signals. J Med Technol 33:442–448
39. Walczak B, Massart DL (1997) Noise suppression and signal compression using the wavelet packet transform. Chemom Intell Lab Syst 36:81–94
40. M. Sifuzzaman M, Islam MR, Ali MZ (2009) Application of wavelet transform and its advantages compared to Fourier transform. J Phys Sci 13: 121–134
41. Schwartz M, Meyer B, Wirnitzer B, Hopf C (2015) Standardized processing of MALDI imaging raw data for enhancement of weak analyte signals in mouse models of gastric cancer and Alzheimer's disease. Anal Bioanal Chem 407:2255–2264
42. Lagarrigue M, Alexandrov T, Dieuset G, Perrin A, Lavigne R, Baulac S, Thiele H, Martin B, Pineau C (2012) New analysis workflow for MALDI imaging mass spectrometry: application to the discovery and identification of potential markers of childhood absence epilepsy. J Proteome Res 11:5453–5463
43. Cappadona S, Levander F, Jansson M, James P, Cerutti S, Pattini L (2008) Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry. Anal Chem 80:4960–4968

44. Chen S, Hong D, Shyr Y (2007) Wavelet-based procedures for proteomic MS data processing. Stat Data Anal 52:211–220
45. Kumar K (2017) Discrete wavelet assisted correlation optimised warping of chromatograms: optimizing the computational time for correcting the drifts in peak positions. Anal Methods 9:2049–2058
46. 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
47. Rinnan Å (2004) Application of PARAFAC on Spectral Data. Royal Veterinary and Agricultural University-DK, (PhD Thesis)
48. Wise BM, Gallaghar NB, Bro R, Shaver JM (2006) PLS_Toolbox 4.0. Eigen vector research