Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 124 (2014) 52–58

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

Discrimination of three dimensional fluorescence spectra based on wavelet analysis and independent component analysis

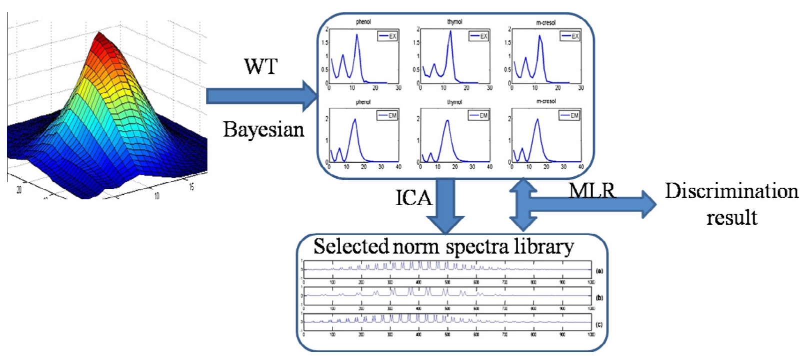
Xiaoya Yu, Yujun Zhang *, Gaofang Yin, Nanjing Zhao, Xue Xiao, Changhua Lu, Yanwei Gao, Wei Zhang

Key Laboratory of Environmental Optics & Technology, Chinese Academy of Sciences, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China

HIGHLIGHTS

- Wavelet analysis and ICA are applied for recognition of overlapped spectra.
- Wavelet analysis extracts the features of the spectra and amplifies differences.
- ICA analysis is used to separate single component before linear regression.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 5 September 2013
Received in revised form 17 November 2013
Accepted 5 December 2013
Available online 21 December 2013

Keywords:
Blind signal separation
Three dimensional fluorescence spectra
Wavelet analysis
Independent component analysis

ABSTRACT

Fluorescence spectroscopy is a rapid and non-destructive method for monitoring water quality. In this work, wavelet analysis, together with independent component analysis (ICA), was applied for component recognition of seriously overlapped, multi-component, three dimensional fluorescence spectra. Wavelet analysis extracts the features of the spectra and amplifies differences among phenolic homologs. ICA analysis in blind signal separation was used to separate single component before multiple linear regression (MLR). The proposed method increases the correct classification rate and enriches the spectra library. As such, it is a useful alternative to traditional techniques in component recognition.

Introduction

Fluorescence spectroscopy is a rapid and non-destructive method used in vivo and in situ water quality monitoring because of its high sensitivity and good expression of features [1,2]. Methods for spectra analysis based on the three-linear model have received significant attention over the past several years. Among known methods, parallel factor analysis (PARAFAC) is the most classical one [3–5]. However, non-multi-linear problems are very common in this field, and such problems may be attributed to [6]: (1) The non-linear relationship between signals and analyte concentrations, (2) the non-multi-linear of signals, and (3) the variation in component profiles across different samples.

Models that allow deviations of multi-linearity in one way or another include: parallel profiles with linear dependencies (PARALIND) [7], multivariate curve resolution couple to ALS (MCR-ALS) [8], non-bilinear rank annihilation (NBRA), [9] unfolded partial least-squares (U-PLS), [10] multi-way PLS (N-PLS), [11] and artificial neural networks (ANN) [12,13]. Models must be selected according to the cause of deviation. For example, U-PLS can be used for three dimensional fluorescence spectra, whereas ICA can be used to analyze second order data, regardless of whether the data is in accordance with the three-linear model or not [14,15].

* Corresponding author. Tel.: +86 551 65593691; fax: +86 551 65593530.
E-mail address: yjzhang@aiofm.ac.cn (Y. Zhang).

1386-1425/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.saa.2013.12.033

Moreover, the non-negativity constraint is unnecessary when using the ICA [16]. When the extracted proportions are negative, the proportions should be multiplied by -1.

Zhang et al. converted three dimensional spectra into two dimensional spectra [17,18], extracted fluorescence features from harmful algal bloom(HAB) species and discriminated the algae at the division and genus level. Wavelet analysis combined with Bayesian discriminant analysis, a method established by MLR, was used to determine discriminant spectra. This method discriminates the algae in the wavelet domain, thereby avoiding the errors caused by tri-linearity deviations. However, this method only discriminates the species in the spectral library. Even worse, operations containing all spectra increase the risk of error. Therefore, blind separation must be executed before discrimination. Thus MLR only takes specific components into account. When the result of blind separation contains the spectra that are not included in the spectral library, these spectra are considered new species and added to the library. Therefore, the proposed process not only increases the correct classification rate, but also enriches the spectral library.

In the present work, wavelet analysis and ICA are used to analyze three dimensional fluorescence spectra and facilitate the monitoring of water quality. In the next section, the theoretical bases of our method are introduced in detail. The section thereafter verifies the feasibility of the proposed method by experiments. Finally, a concise conclusion and some remarks are given.

Theories

Wavelet analysis

Wavelet multi-scale decomposition, also called "mathematics microscope" [19], can refine the intrinsic information of data and extract inner relations. Local representation information in terms of both time and frequency can be extracted by wavelet analysis. The wavelet features of the spectra are the projections of original spectra in the wavelet space [20–22].

The wavelet analysis theory of fluorescence spectra is as follows:

The fluorescence spectra of organic matter is: $H(f), f = 1, 2, \ldots, F$, where F is the number of measured points. First, wavelet analysis refines $H(f)$ into multi-scale signals by selecting an appropriate orthogonal scale base $\Phi_{j,n}(f)$ and corresponding wavelet base. The relationship of scale space Φ_j and wavelet space Ψ_j is as follows:

$$\Phi_j \perp \Psi_j \tag{1}$$

$$\Phi_{j+1} = \Phi_j \oplus \Psi_j \tag{2}$$

Then, the scale component $b_{j,n}$ and wavelet component $e_{k,n}$ can be obtained:

$$b_{j,n} = \sum_{f=1}^{F} H(f)\Psi^*_{j,n}(f) \tag{3}$$

$$e_{k,n} = \sum_{f=1}^{F} H(f)\Phi^*_{k,n}(f) \quad k = 1, 2, \ldots, j \tag{4}$$

The scale component (also called the low-frequency component) represents most of the measured points and contains large scale information. The wavelet component (also called the high-frequency component) reveals few findings in several measured points and contains small scale information. Thus, the scale component was used as standard spectral feature in this paper. Daub4 (Daubechies wavelet with four filter coefficients) [23] was employed as the mother wavelet because it is the most local in terms of time domain [21].

Independent component analysis

ICA is a signal processing technique that aims to recover underlying source signals from a set of mixed signals, based on the assumption that source signals are statistically independent [15]. ICA has been applied to spectroscopic data [24], speech recognition [25], blind signal separation [26], fault detection [27], statistical process monitoring [28], and batch process monitoring [29].

The ICA of a random vector is used to search the linear transformation that minimizes the statistical dependence among its components [15,30]. To design a practical optimization criterion, the expansion of mutual information is used as the function of cumulants.

The following linear statistical model is assumed as:

$$y = Px + v \tag{5}$$

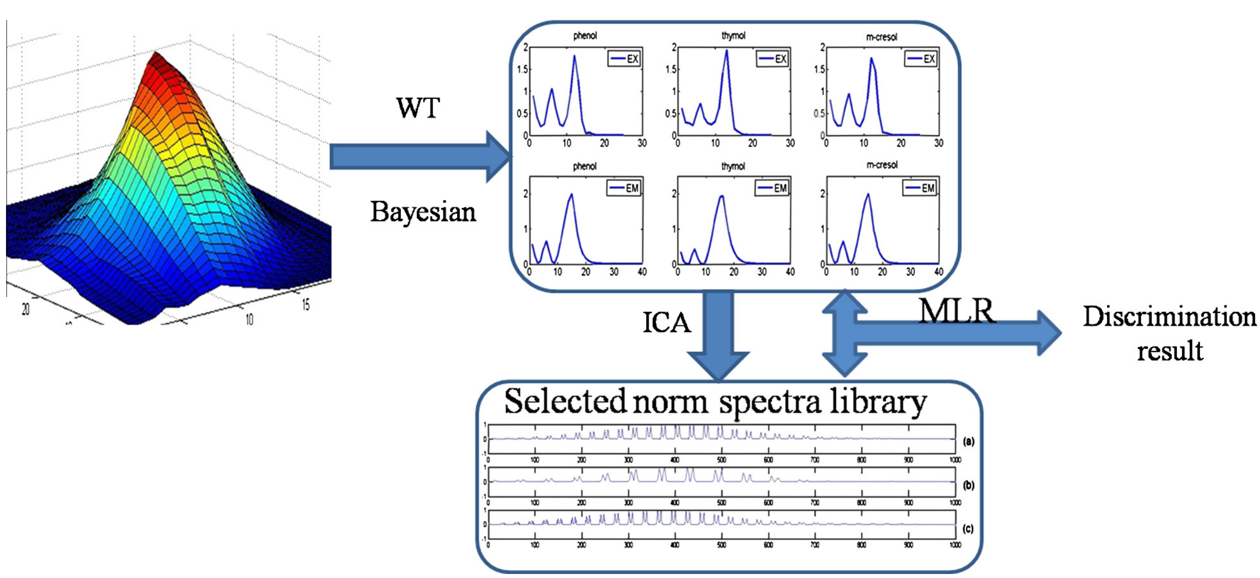


Fig. 1. Main steps of dealing with spectra data.