



PARAFAC analysis of front-face fluorescence data: absorption and scattering effects assessed by means of Monte Carlo simulations

Thierry Aussenac, Lyes Lakhal, Victor Acha

► To cite this version:

Thierry Aussenac, Lyes Lakhal, Victor Acha. PARAFAC analysis of front-face fluorescence data: absorption and scattering effects assessed by means of Monte Carlo simulations. *Chemometrics and Intelligent Laboratory Systems*, 2012, 116, pp.112-122. <hal-04559737>

HAL Id: hal-04559737

<https://hal.science/hal-04559737v1>

Submitted on 25 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

**PARAFAC analysis of front-face fluorescence data: absorption and scattering effects
assessed by means of Monte Carlo simulations**

Lyes Lakhal ^{a,*}, Victor Acha ^b, Thierry Aussenac ^a

^a Institut Polytechnique LaSalle Beauvais, 19 rue Pierre Wagué, BP 30313, 60026 Beauvais,
France

^bHydrogeochemistry and Soil-Environment Interactions (HidrISE) team, Institut
Polytechnique LaSalle-Beauvais, 19 rue Pierre Wagué, B.P. 30313, 60026 Beauvais, France

* Corresponding author. Tel.: + 33 (0)3 44 06 89 68; Fax: + 33 (0)3 44 06 25 26.

E-mail: lyes.lakhal@lasalle-beauvais.fr

Abstract

Three-way fluorescence data originating from mixtures of fluorophores embedded in turbid media such as biological media get strongly modulated by wavelength dependent absorption and scattering phenomena. Thus the consistent resolution and quantitative determination of the mixture becomes a difficult task. In this study two chemometric methodologies frequently used to deal with this type of data were applied to fluorescence simulated data sets qualitatively similar to those measured in biological samples: Parallel Factor Analysis (PARAFAC) that does require the fulfillment of trilinearity, and multivariate curve resolution–alternating least squares (MCR–ALS) which decomposes the data according to a model lacking this structure. Monte Carlo simulations were used to simulate fluorescence excitation-emission matrices (EEMs) of known fluorescent mixtures under separated and simultaneous variations of the absorption parameter μ_a and the scattering parameter μ_s . PARAFAC and constrained MCR-ALS models were then fitted to the simulated data. Both algorithms failed to recover the true profiles. The results obtained with PARAFAC and MCR-ALS models are similar and the recovered profiles exhibit severe distortions due to the absorption and scattering effects. Finally, qualitative and quantitative effects of the absorption and scattering on the fluorescence data were assessed and discussed.

Key words: Fluorescence Spectroscopy, Absorption, Scattering, Monte Carlo, Trilinearity, PARAFAC.

1. Introduction

Fluorescence spectroscopy combined with Parallel Factor Analysis (PARAFAC) [1] is becoming a valuable tool in analytical chemistry [2, 3, 4]. This method allows to identify and quantify different fluorescent compounds such as environmental, therapeutic drugs and metabolites, and food molecules. The process involves exciting samples over a range of wavelengths and recording the fluorescence emission over a different range of wavelengths. When the structure of the data set is trilinear, the PARAFAC decomposition provides an estimation of the relative concentrations, and pure emission as well as excitation spectra of the individual analytes. The trilinear structure of the data means that the fluorescence signals should be linear in concentration, additive and hence independent of each other, and that the spectral response of a given analyte must remain the same in all samples.

These assumptions are valid only in the case of highly diluted samples. For intact biological samples the internal structure of the data is probably non-trilinear due to perturbations influencing the measured spectra and that cannot be modeled by adding one or more additional components in the PARAFAC decomposition. Multivariate curve resolution–alternating least squares (MCR–ALS) method [5, 6, 7, 8, 9] is probably the best chemometric technique for the analysis of this type of data as it allows (small) deviations to the trilinear model in all the components or in some of them.

Depending on the nature of the sample and the geometry of spectra acquisition, there are two distinct phenomena responsible for the non-trilinear structure of fluorescence data. In the case of purely absorbing substances (non-scattering) fluorescence measurements are performed using right angle optics, i.e. the excitation light is focused at the center of the sample and the fluorescence emitted within a certain solid angle is measured at the right angle to the excitation beam direction. The non-trilinear structure of data is due to inner filter effects [10, 11] whose modeling has been the subject of several papers [12, 13, 14, 15]. Based on the assumption of

an exponential attenuation of the intensity, several models relating the true (bilinear) fluorescence signal to the measured fluorescence have been developed and experimentally validated. These models require the acquisition of the sample absorbance [11, 16] or the fluorescence of the original sample diluted several times [17].

In the case of strongly scattering and absorbing media, front-face fluorescence detection mode is used in which fluorescence is viewed from the same side as the excitation beam. The optical path in this case is sufficiently low so that the reabsorption of the excitation light and emitted fluorescence by the fluorophores can be ignored [18, 19]. However, contrary to what is often stated in the literature [20], the acquisition of spectra from turbid media with front-face fluorescence detection mode does not necessarily guarantee the conditions of trilinearity. For instance, biological samples such as plant and animal tissues or milk are turbid with diverse fluorophores, scatters and absorbing molecules. Light incident on and emitted from such media experiences multiple scattering and absorption events before it re-emerges from surface. The combined outcome of absorption and scattering can influence the intrinsic fluorescence and makes the PARAFAC loadings difficult to interpret.

Christensen et al. [21] discussed the factors that can affect the fluorescence emission signals from intact foods. This discussion remains incomplete because the authors (i) do not rigorously describe the phenomena responsible for the optical distortion of fluorescence spectra and (ii) do not give a physical model linking the observed and intrinsic fluorescence of a mixture of fluorophores in such media.

Several research groups have applied PARAFAC for analysis of fluorescence data measured from intact food systems with front-face optics. Airado-Rodriguez et al. [22] used PARAFAC for identification of fluorophores present in wine. PARAFAC was used to monitor the evolution of naturally occurring and neo-formed fluorescent compounds in oils during thermal treatment [23]. Synchronous fluorescence spectroscopy coupled with PARAFAC was used for the

determination of heterocyclic aromatic amines in grilled meat [24] and for the characterization of milk molecular changes during mild heating [25]. The same strategy was proposed by Rizkallah et al., [26] for the identification and rapid quantitative estimation of neoformed contaminants in industrially processed cookies. There are numerous other examples on the use of front-face fluorescence data and PARAFAC in food analysis: detection and characterization of active photosensitizers spectrally in butter [27], evaluation of yogurt during storage [28], evaluation of light-induced oxidation of semi hard cheese [29], and quality monitoring of dry-cured Parma ham [30]. However, one must be particular careful when fitting the PARAFAC model to fluorescence data measured from intact food samples. Light scattering and the presence of absorbing pigments may give a model inadequacy, influencing the estimated model parameters.

In this work, Monte Carlo (MC) simulations were used to simulate fluorescence EEMs qualitatively similar to those observed in biological samples and intact food systems. Each simulated EEM included contributions from three fluorophores and took into account the absorption and scattering effects. PARAFAC and MCR-ALS are applied to these data and the recovered profiles were compared with the pure profiles used in the simulation (excitation spectra, emission spectra, and true concentrations).

The paper is structured as follows: The physical principles governing the interaction of light with turbid media are briefly described in section 2. Mathematical models describing the propagation of excitation and fluorescent light through a turbid medium are discussed in section 3. MC modeling used to simulate the propagation of excitation light and the resulting fluorescence is presented in section 4. The basic of PARAFAC and MCR-ALS models is briefly discussed in section 5. The simulation of fluorescence data sets in different optical environments is described in section 6. Finally, PARAFAC and MCR-ALS models were

applied to each data set and the deterioration of the loadings due to absorption and scattering is assessed and discussed in section 7.

2. Optical properties

Light propagation in turbid media can be described by the Radiance Transport Equation with three optical specified parameters: absorption parameter μ_a [m^{-1}], scattering parameter μ_s [m^{-1}], and the anisotropy factor g [31].

The absorption parameter μ_a is defined as the probability for a photon to be absorbed per unit length. Thus, the probability of absorption in infinitesimal distance ds is $\mu_a ds$. The mean free path for an absorption event, i.e. the mean distance a photon travels before it is absorbed, is $1/\mu_a$.

Similarly, the scattering parameter μ_s is defined as the probability per unit length for a photon to be scattered. The probability of scattering in infinitesimal distance ds is then $\mu_s ds$ and the mean free path for a scattering event is $1/\mu_s$.

When an incident photon along a direction described by the unit vector \hat{s} experiences a scattering event (Fig. 1), the angular probability of this photon to be scattered into \hat{s}' direction is given by the normalized phase function $f(\hat{s}, \hat{s}')$. It can be assumed that the probability distribution is a function of the angle between the incident and scattered photon only, and that it does not depend on the angle of incidence relative to the scatterer. For light transport in a biological medium the Henyey-Greenstein function is most often used [32]:

$$f(\cos \theta) = (1/4\pi) \frac{1 - g^2}{(1 + g^2 - 2g \cos \theta)^{3/2}}. \quad (1)$$

The scaling factor normalizes $f(\cos \theta)$ such that:

$$\int_{-1}^{+1} f(\cos \theta) d(\cos \theta) = 1. \quad (2)$$

The anisotropy can be characterized in terms of the mean cosine of the scattering angle which is called the anisotropy factor g :

$$g = \int_{-1}^{+1} f(\cos \theta) \cos \theta d(\cos \theta). \quad (3)$$

The value of g ranges between -1 (for total back scattering) and 1 (for total forward scattering), while $g = 0$ corresponds to isotropic scattering.

3. Model of Excitation and Emission Matrices (EEMs) in absorbing and scattering medium

Identification and quantification of fluorescent compounds in complex mixtures is performed by means of PARAFAC decomposition applied to hyphenated fluorescence spectroscopy data [33]. The data set under consideration is a three-way array or third-order tensor in which several EEMs are stacked. A single EEM is obtained by scanning the excitation wavelength range, producing an emission spectrum for each excitation wavelength. Thus, one dimension of the tensor corresponds to the excitation wavelength (λ_i), the second corresponds to the emission wavelength (λ_j) and the third to the sample index (k). The PARAFAC decomposition of the third-order tensor is often unique under minor assumptions [34]; this is due to the fact that the third-order tensor is not merely a collection of EEMs, but there is actually an internal relationship between each of the EEMs.

In this section, an expression for the fluorescence EEM measured from the surface of homogeneous absorbing and scattering medium containing a uniform distribution of fluorophores is developed.

3.1 Optically diluted homogeneous medium

For a set of diluted samples containing several fluorophores, the normalized fluorescence intensity of the k^{th} sample detected at wavelength λ_j when excited at wavelength λ_i can be described as a trilinear expression [3],

$$EEM_{Dil}(\lambda_i, \lambda_j, k) = \sum_{f=1}^F \chi_f(k) \varepsilon_f(\lambda_i) \varphi_f(\lambda_j), \quad (4)$$
$$1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K.$$

where $\chi_f(k)$ is the wavelength independent factor of fluorophore f in the sample k that explains how the EEM depends on the concentration, $\varepsilon_f(\lambda_i)$ is the molar absorptivity of fluorophore f at excitation wavelength λ_i and $\varphi_f(\lambda_j)$ is the fluorescence emission intensity of fluorophore f at emission wavelength λ_j . F is the number of fluorophores, I the number of excitation wavelengths, J the number of emission wavelengths and K the number of samples.

3.2 Turbid medium

The fluorescent light in a turbid medium is absorbed and scattered before emerging from the surface. Thus measurements of fluorescence intensity detected at the surface depend on the optical properties of both the fluorophores and the medium.

In general, a one-dimensional model of fluorescence must describe the attenuation of the excitation beam as it passes through the turbid sample and the attenuation of the fluorescence as it travels from the point of generation to the sample surface. These effects can be mathematically described by recognizing that a turbid sample can be treated as a dilute solution if its thickness is small compared to the photon mean free path (i.e., the mean distance traveled by a photon before being absorbed or scattered). We divide a turbid sample of thickness L_{Turb} into a number of thin layers of thickness dz along the optical axis. This allows us to calculate

the contribution to the fluorescence $dEEM_{Tur}$, collected from the front surface of a thin layer with thickness dz and located at depth z (Fig. 2), as follows [35]:

$$dEEM_{Tur}(\lambda_i, \lambda_j, k) = (2/L_{Tur}) \left[\sum_{f=1}^F \chi_f(k) \varepsilon_f(\lambda_i) \varphi_f(\lambda_j) \right] \left[H_{in}(\lambda_i, z, k) H_{out}(\lambda_j, z, k) \right], \quad (5)$$

$$1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K.$$

In this expression, H_{in} describes the fraction of the incident excitation light that reaches the layer of concern, and H_{out} describes the fraction of fluorescence generated at depth z that reaches the front surface. Note that for a dilute solution with isotropically emitting fluorophores, H_{in} is 1 and H_{out} is 1/2.

The total fluorescence emitted from the front surface of a sample k is obtained by integrating over the depth z :

$$EEM_{Tur}(\lambda_i, \lambda_j, k) = (2/L_{Tur}) \int_0^{L_{Tur}} \left[\sum_{f=1}^F \chi_f(k) \varepsilon_f(\lambda_i) \varphi_f(\lambda_j) \right] \left[H_{in}(\lambda_i, z, k) H_{out}(\lambda_j, z, k) \right] dz, \quad (6)$$

$$1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K.$$

If fluorophores are homogeneously distributed within the sample, the summation can be taken outside the integral, which yields:

$$EEM_{Tur}(\lambda_i, \lambda_j, k) = (2/L_{Tur}) \left[\sum_{f=1}^F \chi_f(k) \varepsilon_f(\lambda_i) \varphi_f(\lambda_j) \right] \int_0^{L_{Tur}} H_{in}(\lambda_i, z, k) H_{out}(\lambda_j, z, k) dz, \quad (7)$$

$$1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K.$$

Combining Eqs. (4) and (7) yields:

$$EEM_{Tur}(\lambda_i, \lambda_j, k) = EEM_{Dil}(\lambda_i, \lambda_j, k) \cdot TF(\lambda_i, \lambda_j, k), \quad (8)$$

$$1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K.$$

where

$$TF(\lambda_i, \lambda_j, k) = \int_0^{L_{Tur}} H_{in}(\lambda_i, z, k) H_{out}(\lambda_j, z, k) dz, \quad (9)$$

$$1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K.$$

This fundamental result provides the key for relating the "true" fluorescence signal, EEM_{Dil} , which is independent of the optical properties of the medium, to the measured fluorescence of a turbid sample, EEM_{Tur} , by the wavelength dependent Transfer Function (TF).

The form of the TF depends on the details of the light propagation within the sample. Several models of light propagation have been described in the literature and can be used to calculate the dependence of the TF on the sample's optical properties, collection and detection geometry, and mismatch in the index of refraction at the sample surface [36]. Once the form of the TF has been determined, its value can be computed as a function of the excitation and emission wavelengths, given the total optical properties of the turbid sample at these wavelengths and the thickness of the sample.

The MC technique offers a flexible and accurate approach to these problems, since it can handle turbid media with complex geometries and since it can score multiple physical quantities simultaneously. Hence MC simulation can be used to estimate the TF.

4. Fluorescence Monte Carlo Simulation (FMCS)

Modeling fluorescence in turbid media must account for the following stages: (i) propagation of excitation light from the medium's surface into its interior, (ii) absorption by fluorophores, (iii) conversion to fluorescence, (iv) emission of the fluorescent light from the fluorophores, and (v) propagation of that light back up to the surface (we neglect here re-emission of the absorbed light). Besides, we assume that the absorption and emission spectra of the fluorescent molecules do not overlap. Hence, the excitation and emission processes take place at different wavelengths denoted by λ_i and $\lambda_j > \lambda_i$, respectively.

Consider a short light pulse at the excitation wavelength λ_i , which enters the medium in the direction perpendicular to its surface, as illustrated in Fig. 3. The solid lines show simulated excitation photon paths. When it is absorbed by a fluorophore, a new fluorescence photon is emitted at the wavelength λ_j (dashed line). The fluorescence photon propagates through the medium and is eventually absorbed or reaches the surface and is observed.

The modeled medium is represented by a semi-infinite slab with infinite x, y dimensions and homogeneously distributed interaction (absorption or scattering) centers, with thickness L_{Tur} , refractive index n_{medium} , optical properties at the excitation wavelength $[\mu_a, \mu_s, g](\lambda_i)$ and optical properties at the emission wavelength $[\mu_a, \mu_s, g](\lambda_j)$.

Three coordinate systems are employed in the FMCS: (a) a Cartesian coordinate system (x, y, z) is used to trace the photon movements, (b) a cylindrical coordinate system (r, z, ϕ) is used to record the internal photon absorption and fluorescence as a function of r and z , and (c) a moving spherical coordinate system, of which the z -axis is dynamically aligned with the photon propagation direction, is used for sampling the propagation direction of a photon packet.

Photon packets are traced inside the slab as they are absorbed and scattered in the interaction centers (marked by circles in Fig. 3), analogous to a random walk in three dimensions.

FMCS consists of two separate MC simulations. The first simulation deals with the excitation light traveling from the light source to the fluorophore. The second simulation deals with the emitted fluorescence photon traveling from the fluorophore to the surface.

FMCS determines separately (a) the absorption probability distribution, and (b) the emission probability distribution. The combination of these two probabilities provides the TF.

We use an open source MC simulation package provided by Wang et al. [37] to determine the absorption probability distribution $A(\lambda_i, r, z)$, defined as the probability per unit volume for an excitation photon to be absorbed at a radial distance r and a depth z from the injection point $(0,$

0). The software (both executables and source codes) along with a thorough manual is available at [38].

This software was modified according to Swartling et al. [39] to determine the emission probability distribution $E(\lambda_j, r, z)$, defined as the probability per unit area that a photon at position $(0, z)$ will exit the surface of the medium at position $(r, 0)$.

After $A(\lambda_i, r, z)$ and $E(\lambda_j, r, z)$ have been simulated, the one-dimensional photon absorption function $H_{in}(\lambda_i, z)$ and photon fluorescence function $H_{out}(\lambda_j, z)$ are obtained by numerical integration over the radial distance r :

$$H_{in}(\lambda_i, z) = \int_0^{\infty} A(\lambda_i, r, z) 2\pi r dr, \quad (10)$$

$$H_{out}(\lambda_j, z) = \int_0^{\infty} E(\lambda_j, r, z) 2\pi r dr, \quad (11)$$

Finally, the TF follows from Eq. (9).

5. Multiway data analysis

Both PARAFAC [1, 2, 3] and MCR-ALS [5, 6, 7, 8, 9] models have been discussed in detail elsewhere and only a brief description is presented here.

5.1 Parallel factor analysis (PARAFAC)

Let $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ be a three-way fluorescence data array. Then an F-component PARAFAC model can be expressed as in Eq.12, where a_{if} and b_{jf} are the elements of the loading matrices $\mathbf{A} \in \mathbb{R}^{I \times F}$ and $\mathbf{B} \in \mathbb{R}^{J \times F}$ respectively, and c_{kf} is the element of the scores matrix $\mathbf{C} \in \mathbb{R}^{K \times F}$. x_{ijk} represents an entry of $\underline{\mathbf{X}}$ in the i^{th} row, j^{th} column and k^{th} tube and e_{ijk} are the elements of the residual array $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$, which contains the variance not captured by the model.

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk}. \quad (12)$$

In the context of the three-way fluorescence data analysis, the F columns of \mathbf{A} represent pure excitation spectra, the F columns of \mathbf{B} represent pure emission spectra and the F columns in \mathbf{C} contain the relative concentrations of each analyte in the different samples.

There are several techniques for finding a valid PARAFAC model for describing fluorescence excitation-emission data. The appropriate number of components is determined based on several different criteria. For example, the variance explained by the model, the visual appearance of loadings, the number of iterations of the algorithm and the core consistency diagnostic (CORCONDIA) [40] are used. Among these techniques, CORCONDIA has been commonly applied in literature [41]. However, for complex data, the determination of the number of factors by CORCONDIA remains elusive [42]. Therefore, it is often suggested that several diagnostic tools are used together rather than a single method [40].

5.2 MCR-ALS

The strategy to perform MCR-ALS decomposition of the three-way fluorescence data array $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ is to unfolded $\underline{\mathbf{X}}$ along the mode that is suspected of breaking the trilinear structure [5, 6]. Both emission-column-wise and excitation-column-wise matrix augmentations are possible. In the emission-wise augmentation used here, an emission- augmented matrix $\mathbf{X}_{\text{aug}} \in \mathbb{R}^{I \times J}$ is created by setting one EEM $\in \mathbb{R}^{I \times J}$ on top of the others and keeping the common emission wavelengths in the same column. The MCR-ALS decomposition of $\mathbf{X}_{\text{aug}} \in \mathbb{R}^{I \times J}$ is performed according to the expression [7, 8]:

$$x_{ij} = \sum_{f=1}^F d_{if} b_{fj} + e_{ij}. \quad (13)$$

where x_{ij} is the ij^{th} element in the augmented matrix \mathbf{X}_{aug} , d_{if} is the if^{th} element in the column-wise augmented matrix $\mathbf{D} \in \mathbb{R}^{I \times I}$ of excitation spectra, b_{fj} is the fj^{th} element in $\mathbf{B} \in \mathbb{R}^{J \times F}$, and e_{ij} is the residual term. In other words, \mathbf{D} matrix contains the sets of excitation spectra (one set for each EEM) for each factor and \mathbf{B} matrix contains the emission spectra common to all samples. A single set of excitation profiles may then be selected for comparison or all recorded sets may be averaged to create one set of excitation profiles. The intensity at each wavelength is then related to the concentration of each component in the mixture.

6. Description of simulated data

In order to assess the effect of the absorption and scattering independently and in combination, MC simulations were performed to simulate fluorescence EEMs of a fluorophores mixture emitting in different optical environments: (i) optically clear (dilute) media (ii) purely scattering media containing Intralipid and (iii) absorbing and scattering media containing Intralipid and β -carotene.

Intralipid is a fat emulsion that is used clinically as an intravenously administered nutrient. In biomedical optics, it is used for providing the scattering component in a tissue phantom to investigate propagation of light in tissue [43, 44].

Three fluorophores with fairly similar spectral properties were used in this simulation. The three compounds are Vitamin A (ViTA), Vitamin B6 (VitB6) and the Reduced Nicotinamide Adenine Dinucleotide (NADH). Their excitation and emission spectra, downloaded from FoodFluor Database at [45], were mixed according to (Eq. 4) to produce a set of 20 EEMs. The relative concentrations profiles (proportions of the mixtures) were chosen so that the uniqueness of the PARAFAC decomposition is guaranteed (random and linearly independent profiles). These data are referred to as data set 1.

These EEMs were then affected by scattering and absorption effects via MC simulated TFs according to (Eq.9). The TFs were simulated using the absorption parameter (μ_a) of β -carotene

and the scattering parameters (μ_s and g) of Intralipid. These were obtained from [38] and shown in Fig.4.

To investigate scattering effect, MC simulations were performed using the scattering parameters (μ_s , g) of the Intralipid and fixing the absorption parameter μ_a to 0. The simulations were conducted as follows: The first simulation is run with parameters μ_s and g of

Fig. 4. The scattering coefficient, μ_s , is then varied by linearly increasing the concentration of Intralipid and a new simulation is launched. The added concentration of Intralipid was chosen in such way to achieve an increase of 100% at the last simulation. At the k^{th} simulation, a new TF is calculated and multiplied point by point with the k^{th} EEM of data set 1 to get the k^{th} EEM data set 2.

To examine the effect of absorption and scattering in combination, the values of μ_a and μ_s were simultaneously varied. The absorption parameter μ_a was varied by altering the concentration of β -carotene in such way to reach an increase of about 100% at the last simulation. At the k^{th} simulation, a new TF is generated and applied to the k^{th} EEM du data set 1 to get the k^{th} EEM of data set 3.

Examples of fluorescence EEMs simulated in different optical environments are shown in Fig.5 together with the corresponding simulated TFs.

7. Results and discussion

The simulated data sets were stacked in a three-way arrays (excitation \times emission \times samples), of which the dimensions corresponded to 25 excitation wavelengths (mode one), 44 wavelengths (mode two) and 20 samples (mode three) respectively. PARAFAC and MCR-ALS methods were then used to decompose the three data sets into several spectral profiles of the contributing components.

PARAFAC modeling was carried out using the N-way Toolbox for MATLAB [46]. As convergence criteria a value of 1×10^{-6} and a maximum number of iterations of 2500 were used. The unconstrained models were chosen because an attempt to fit a PARAFAC models with non-negativity constraints in the spectral mode resulted in no significant improvement compared to the unconstrained option (not shown). The calculations associated to the MCR-ALS method were performed using several programs implemented in MATLAB and obtained from [47]. The MCR-ALS decomposition of the three data sets was done using a non-negativity constraints for both spectral modes performed with the Fast non-negative least-squares algorithm. The structure of the data set 1 was assumed trilinear while non-trilinear structure was selected for the MCR-ALS decomposition of data sets 2 and 3. The matrix loading **B** of the PARAFAC model was taken as the initial estimate for the emission spectra.

Unconstrained PARAFAC models of the fluorescence data sets were fitted using from one to seven components. The results of the model evaluation criteria are shown in Table 1.

For data set 1, three components is obviously the ideal number. The variance explained by the model was close to 100% and the core consistency test showed a CORCONDIA of 100 %. When a four-component model was fitted the CORCONDIA dropped significantly to around 32%. This result is consistent with the data set 1 constituted from only the three fluorophores. For data set 2, the explained variance is close to 100 % using from three to seven components. The analysis using CORCONDIA indicates that three components are necessary, because the utilization of more factors leads to a great decrease of CORCONDIA and hence of the trilinearity of the data modeled. Three components give a CORCONDIA value of 100% (a perfect trilinear model) whilst, when using four or more components, this value diminishes to values below to 32%.

Four components provide a valid PARAFAC model For Data set 3. The four-component model explains approximately 99.9% of the variation in the data and gives a CORCONDIA of 87.4%.

The CORCONDIA of the five-component model of 19% indicates that this model might not be stable.

The explained variance and the CORCONDIA criteria show that the intrinsic EEMs of data sets 2 and 3 are constituted by three and four components respectively. This is not consistent with a priori knowledge. The data set 2 was constituted from 3 fluorophores plus scatter, what gives at least four factors. The data set 3 was constituted from 3 fluorophores plus absorber and scatter, what gives at least five components. Consequently, the four- component model for data set 2 and the five-component model for data set 3 deserve further analysis.

As mentioned above, the MCR-ALS procedure decomposes an unfolded fluorescence three-way array into two matrices: one matrix containing the excitation spectra as a function of the sample index k (here $1 \leq k \leq 20$), and the second matrix containing the emission spectra which are common to all samples. Thus for clarity reasons only the excitation spectra corresponding to the first sample (the less affected by absorption and scattering) will be shown hereafter.

Spectral loadings and scores profiles obtained from the three-component PARAFAC model were fitted to the data set 1 (Fig. 6) and data set 2 (Fig. 7). The pure excitation and emission spectra of VitA, VitB6 and NADH as well as the excitation and emission spectra obtained by MCR-ALS were also plotted for comparison. The scores were plotted against the true concentrations and a linear fit was performed. The deformation induced by absorption and scattering on the linear relationship between the PARAFAC scores (fluorescence intensity) and fluorophores concentration was evaluated by the Mean Relative Error (MRE %) (Table 2).

The PARAFAC and MCR-ALS emission and excitation spectra of the three components extracted from data sets 1 and 2 are quite similar and perfectly matched to the pure spectra (Fig. 6, and 7). Thus PARAFAC and MCR-ALS were able to perfectly resolve the three fluorophores even in presence of scattering effects. Accordingly, the TF of Intralipid may not have had any effect on the trilinear structure of the data set 2.

In a pure scattering medium, modulation of fluorescence by the TF is limited to a scaling by a factor $\alpha \geq 1$, at least for the values of μ_s that were considered. Fig. 5C shows a nearly constant TF, and Fig. 5B shows an EEM that is very similar to the corresponding unaffected EEM (Fig. 5A). Hence, the emission and excitation profiles were accurately recovered.

For the concentration mode, the data set 1 PARAFAC scores are proportional to the concentrations of fluorophores. As a result the predicted fluorophore concentrations have a good agreement with their true concentrations.

As for data set 2, the scattering produced by Intralipid resulted in an overall increase in fluorescence intensity. Despite of some dispersion of predicted concentrations around the linear fit, good agreements between the predicted and the true concentration were obtained. The maximum prediction MRE was less than 3.5 %.

As mentioned above, the pure scattering samples (data set 2), at least in the considered range of μ_s values, the TF modulation is limited to scaling by a factor α . In this range, the sample to sample TF variations are relatively small. Thus the PARAFAC scores correspond more or less to the true PARAFAC scores, increased by a constant factor α .

The estimated spectral loadings and scores from fitting the four-component PARAFAC model to data set 3 are illustrated in Fig. 8. The profiles recovered in the MCR-ALS analysis as well as the pure excitation and emission spectra of the analytes are also shown. The PARAFAC and MCR-ALS were found to be similar. While no scattering effects were observed, it is clear that absorption did introduce spectral distortions in both spectral loadings obtained from data set 3. By comparing PARAFAC and MCR-ALS spectral loadings with pure spectra of VitA, VitB6 and NADH it was determined that only VitB6 (2nd component) was well matched, although some differences in the profiles. However, VitA (1st component) and NADH (3rd component) did not matched.

The recovery of the true spectral profiles from data set 3 using PARAFAC and MCR-ALS becomes difficult because the absorption and scattering effects destroy not only the trilinear but also the bilinear structure of the data. Therefore, without a prior knowledge, the chemical interpretation of the PARAFAC and MCR-ALS loadings is a hard task. Even if the profiles look reasonable, the presence of β -carotene avoids VitA and NADH to be identified. In real applications, components 1 and 3 might have been identified as other fluorophores.

The observed spectral deformations are likely due to the modulation by the TF where the β -carotene absorption characteristic plays clearly a role. The β -carotene absorption peaks produce a narrow valley centered at 280 nm (excitation and emission) and a wide valley centered at 450 nm (emission) in the TF (Fig. 5E). The effects of these valleys can be clearly observed in the fluorescence EEM of Fig. 5D. The extended valley along the 450 nm emission line flattens the emission spectra of VitA and NADH from 354 nm to 490 nm by including an emission peak shift toward higher wavelengths. The narrow valley centered at 280 nm in the TF is responsible of the distortion of the excitation spectra between 260 and 380 nm. On the other hand, the excitation and emission spectra of VitB6 are less affected because they are located outside the β -carotene absorption bands.

For the concentration mode, there is no correlation between the PARAFAC scores and fluorophore concentration vectors. For the 3rd component (NADH) concentrations and PARAFAC scores varies in the opposite direction. The MRE in the predicted concentrations ranged from 29.6% (1st component, VitA) to 233 % (3rd component, NADH).

In absorbing and scattering samples, there is a variable reduction of the emitted fluorescence that becomes more pronounced with increasing levels of absorption. As mentioned above, the absorption parameter μ_a increases linearly in subsequent samples (data set 3). On the other hand, the concentrations of the three fluorophores are random in all the data sets. If the concentration of the fluorophore remains constant between two successive samples, the measured

fluorescence intensity decreases due to a higher amount of β -carotene. On the other hand, if the fluorophore concentration increases, two different effects work in opposite directions: The increase of the fluorophore concentration tends to increase the signal intensity while the absorption parameter μ_a of β -carotene tends to decrease the signal. In this case the evolution of the fluorescence intensity is unpredictable as it can be seen for NADH (3rd component) in Fig. 8C.

Figs. 9 and 10 show the loadings and scores of the four- and five-component models fitted to the data set 2, and 3, respectively. For data set 2, the first three components have similar loadings and scores as the three components estimated using the three-component model. Besides components 1, 2 and 3, a fourth component emerged which had similar excitation and emission profiles as component 1. The increase of component 4 is related to the increase of the sample index which is probably due to the scattering process which tends to enhance the fluorescence signal.

Fig 10 shows the loadings and scores of the five-component model fitted to data set 3; The five component model didn't improve as much the fitting. Compared to the four-component model, the first four components have similar loadings and scores; however, the fifth component has no chemical meaning.

The effects absorption and scattering on fluorescence measurements cannot be eliminated by adding more components to the PARAFAC model. This fact clearly indicates that the absorption and scattering effects need to be removed from fluorescence data to get an accurate estimation of fluorophore excitation and emission spectra.

7. Conclusions

Due to the non-bilinear (consequently non-trilinear) structure of the multi-way fluorescence data, resolution of the pure component spectral and concentration profiles from mixtures of

fluorophores embedded in absorbing and scattering samples cannot be accomplished with the use of the current chemometric methodologies using the second order advantage.

By modeling the excitation and fluorescence light propagation in absorbing and scattering media, the measured EEM was related to the intrinsic (bilinear) EEM by means of a transfer function (TF). The former is due only to fluorophores and the latter depends upon absorption and scattering properties at both excitation and emission wavelengths.

Based on this theoretical formalism, a series of Monte Carlo simulations were run to generate fluorescence EEMs which are qualitatively similar to those observed from biological samples.

PARAFAC and MCR-ALS decompositions of these EEMs provided deformed and hence non-reliable component loadings. This demonstrates that the absorption and scattering effects destroyed the bilinear (and consequently the trilinear) structure of the data. Furthermore, no correlation was found between the true concentrations and the concentrations predicted by PARAFAC scores. Simulation results showed also that the pure profiles cannot be recovered by adding more components to the model. In order to obtain interpretable component loadings and correctly interpret the spectra, it is necessary to disentangle the effects of absorption and scattering from the measured fluorescence EEMs.

References

- [1] R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an explanatory multi-modal factor analysis, UCLA Work. Pap. Phon. 16 (1970) 1-84.
- [2] R. Bro, Review on Multiway Analysis in Chemistry - 2000-2005. Crit. Rev. Anal. Chem. 36 (2006) 279-293.
- [3] A. Smilde, R. Bro, P. Geladi, Multi-way Analysis. Applications in the Chemical Sciences, Wiley, Chichester, 2004.

- [4] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri and R.J. Poppi, Second and third-order multivariate calibration: Data, algorithms and applications, *Trends. Anal. Chem.* 26 (2007) 752-765.
- [5] M. C. G. Antunes, J. C. G. E. da Silva, Multivariate curve resolution analysis excitation-emission matrices of fluorescence of humic substances, *Anal. Chim. Acta.* 546 (2005) 52-59.
- [6] A. de Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, *J. Chemom.* 15 (2001) 749-772.
- [7] R. Tauler, Multivariate curve resolution applied to second order data, *Chemom. Intell. Lab. Syst.* 30 (1995) 133-146.
- [8] R. Tauler, A. Smilde and B.R. Kowalski, Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution, *J. Chemom.* 9 (1995) 31-58.
- [9] J Jaumot, R Gargallo, A Dejuan, R Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemom. Intell. Lab. Syst.* 76 (2005) 101-110.
- [10] C.A. Parker, W.J. Barnes, Some experiments with spectrofluorimeters and filter fluorimeters, *The Analyst.* 82 (1957) 606-618.
- [11] J. J. Mobed, S.L. Hemmingsen, J.L. Autry, L.B. McGown, Fluorescence characterization of IHSS humic substances: Total luminescence spectra with absorbance correction, *Environ. Sci. Technol.* 30 (1996) 3061-3065.
- [12] J.F. Holland, R.E. Teets, P.M. Kelly, A. Timnick, Correction of right-angle fluorescence measurements for the absorption of excitation radiation, *Anal. Chem.* 49 (1977) 706-710.

- [13] M.C. Yappert, J.D. Ingle, Correction of Polychromatic Luminescence Signals for Inner Filter Effects, *Appl. Spectrosc.* 43 (1989) 759-767.
- [14] A. Credi, L. Prodi, From observed to corrected luminescence intensity of solution systems: an easy-to-apply correction method for standard spectrofluorimeters, *Spec. Acta A-Mol. Biomol. Spec.* 54 (1998) 159-170.
- [15] A. Andrade-Eiroa, E. Vázquez-Blanco, P. López-Mahía, S. Muniategui-Lorenzo, D. Prada-Rodríguez, Modeling Of Inner Filter Effect In Synchronous Spectrofluorimetry By Using Partial Least Squares, *Analisis.* 28 (2000) 148-154.
- [16] Q. Gu, J. E. Kenny, Improvement of Inner Filter Effect Correction Based on Determination of Effective Geometric Parameters Using a Conventional Fluorimeter, *Anal. Chem.* 81 (2009) 420-426.
- [17] X. Luciani, S. Mounier, R. Redon, A. Bois, A Simple Correction Method Of Inner Filter Effects affecting FEEM and its application to the PARAFAC decomposition, *Chemometr Intell Lab Syst.* 96 (2009) 227-238.
- [18] R.E. Hirstch, Front-Face Fluorescence Spectroscopy Of Hemoglobins, *Methods. Enzymol.* 232 (1994) 231-246.
- [19] M. Zandomeneghi, L. Carbonaro, Chiara Caffarata, Fluorescence of Vegetable Oils: Olive Oils, *J. Agric. Food Chem.* 53 (2005) 759-766.
- [20] C.M. Andersen, G. Mortensen, Fluorescence spectroscopy: a rapid tool for analyzing dairy products, *J. Agric. Food. Chem.* 56 (2008) 720-729.
- [21] J. Christensen, L. Nørgaard, R. Bro, S.B. Engelsen, Multivariate Autofluorescence of Intact Food Systems, *Chem. Rev.* 106 (2006) 1979-1994.

- [22] D. Airado-Rodríguez, T. Galeano-Daz, I. Durn-Mers, J. P. Wold, Usefulness of Fluorescence Excitation-Emission Matrices in Combination with PARAFAC, as Fingerprints of Red Wines, *J. Agric. Food Chem.* 57 (2009) 1711-1720.
- [23] P. Valderrama, P.H. Marçó, N. Locquet, F. Ammari, D.N. Rutledge, A procedure to facilitate the choice of the number of factors in multi-way data analysis applied to the natural samples: Application to monitoring the thermal degradation of oils using front-face fluorescence spectroscopy. *Chemom. Intell. Lab. Syst.* (in press).
- [24] A. Sahar, S. Portanguen, A. Kondjoyan, Dufour E, Potential of synchronous fluorescence spectroscopy coupled with chemometrics to determine the heterocyclic aromatic amines in grilled meat, *Eur. Food. Res. Tech.* 231 (2010) 803-812.
- [25] T. Boubellouta, E. Dufour. Effects of Mild Heating and Acidification on the Molecular Structure of Milk Components as Investigated by Synchronous Front-Face Fluorescence Spectroscopy Coupled with Parallel Factor Analysis, *Appl. Spectrosc.* 62 (2008) 490-496.
- [26] J. Rizkallaha, F. J. Morales, L. Ait-ameur, V. Fogliano, A. Hervieu, M. Courel, I. Birlouez Aragon, Front face fluorescence spectroscopy and multiway analysis for process control and NFC prediction in industrially processed cookies, *Chemom. Intell. Lab. Syst.* 93 (2008) 99-107.
- [27] J. P. Wold, R. Bro, A. Veberg, F. Lundby, A. N. Nilsen, J. Moan, Active Photosensitizers in Butter Detected by Fluorescence Spectroscopy and Multivariate Curve Resolution, *J. Agric. Food Chem.* 54 (2006) 10197-10204.
- [28] J. Christensen, E. Miquel Becker, C.S. Frederiksen, Fluorescence spectroscopy and PARAFAC in the analysis of yogurt, *Chemom. Intell. Lab. Syst.* 75 (2005) 201-208.
- [29] C.M. Andersen, M. Vishart, V.K. Holm, Application of Fluorescence Spectroscopy in the Evaluation of Light-Induced Oxidation in Cheese, *J. Agric. Food. Chem.* 53 (2005) 9985-9992.

- [30] J.K.S. Møller, G. Parolari, L. Gabba, J. Christensen, L.H. Skibsted, Monitoring chemical changes of dry-cured Parma ham during processing by surface autofluorescence spectroscopy. *J. Agric. Food. Chem.* 51 (2003) 1224-1230.
- [31] J. Swartling, Biomedical and Atmospheric Applications of Optical Spectroscopy in Scattering Media, Doctoral thesis, Lund Institute of Technology, Lund, 2002.
- [32] S.L. Jacques , C.A. Alter , S.A. Prahl, Angular dependence of HeNe laser light scattering by human dermis, *Lasers. Life Sci.* 1 (1987) 309-333.
- [33] R. Bro, PARAFAC: Tutorial and applications, *Chemom. Intell. Lab. Syst.* 38 (1997) 149-171.
- [34] A. Stegeman, N.D. Sidiropoulos, On Kruskal's uniqueness condition for the Candecomp/Parafac Decomposition, *Lin. Alg. Appl.* 420 (2007) 540-552.
- [35] R. Richards-Kortum, Fluorescence spectroscopy of turbid media, in: A. J. Welch and M. J. C. van Gemert (Eds.), *Optical-Thermal Response of Laser Irradiated Tissue*, Plenum Press, New York, 1995, pp. 667-708.
- [36] N. Ramanujam, Fluorescence Spectroscopy In Vivo, in: R.A. Meyers (Eds.), *Encyclopedia of Analytical Chemistry*, Wiley, Chichester, 2000, pp. 20-56.
- [37] L.H. Wang, S.L. Jacques, L. Zheng, MCML – Monte Carlo Modeling of light transport in multi-layered tissues, *Comput. Methods. Programs. Biomed.* 47 (1995) 131-146.
- [38] <http://omlc.ogi.edu/spectra/>
- [39] J. Swartling, A. Pifferi, A.M.K. Enejder, S. Andersson-Engels, Accelerated Monte Carlo models to simulate fluorescence spectra from layered tissues, *J.Opt. Soc. Am.* 20 (2003) 714-727.

- [40] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, *J. Chemom.* 17 (2003) 274-286.
- [41] C. M. Andersen, and R. Bro, Practical aspects of parafac modelling of fluorescence excitation-emission Data, *J. Chemom.* 17 (2003) 200-215.
- [42] R. Bro, N. Viereck, M. Toft, H. Toft, P. I. Hansen, S. B. Engelsen, Mathematical chromatography solves the cocktail party effect in mixtures using 2D spectra and PARAFAC, *TrAC, Trends Anal. Chem.* 29 (2010) 281-284.
- [43] HG van Staveren, CJM Moes, J van Marle, SA Prahl, MJC van Gemert, Light scattering in Intralipid-10% in the wavelength range of 400-1100 nanometers, *Appl. Opt.* 30 (1991) 4507-4514.
- [44] J.R. Lakowicz, J.D. Dattelbaum, I.Gryczynski, Intensity measurements in scattering media, *Sens. Actuators, B.* 60 (1999) 1-7.
- [45] <http://www.models.life.ku.dk>.
- [46] C. A. Andersson, R. Bro, The N-way toolbox for MATLAB, *Chemom. Intell. Lab. Syst.* 52 (2000) 1-4.
- [47] <http://www.ub.edu/mcr/welcome.html>.

FIGURE CAPTIONS

Figure 1: A scattering event causes a deflection at angle θ from the original forward trajectory.

Figure 2: Geometry for the fluorescence model in a turbid medium. H_{in} represents the probability that an excitation photon will reach a thin layer dz , located at depth z below the sample surface. Similarly, H_{out} represents the probability that an isotropically emitted fluorescence photon generated in the layer dz will reach the front surface of the sample.

Figure 3: Schematic illustration of fluorescence excitation and emission process from a turbid medium. Solid line represents a path of an excitation photon and the dotted line is a path of emission photon.

Figure 4: Optical properties used in the Monte Carlo simulation: (A) absorption coefficient μ_a of β -carotene, (B) scattering coefficient μ_s and (C) anisotropy factor g of Intralipid.

Figure 5: Typical simulated EEMs in different optical environments. (A) clear medium (data set 1), (B) purely scattering medium (data set 2) and (D) absorbing and scattering medium (data set 3). The EEMs (B) and (D) were obtained from the EEM (A) and the TFs (C) and (E) respectively according to Eq.8.

Figure 6: (A) Excitation spectra and (B) emission spectra recovered with three-component PARAFAC (\circ) and MCR-ALS (\square) models fitted to data set 1 together with the pure excitation and emission spectra (*). The 1st, 2nd and the 3rd components are assigned to VitA, VitB6 and NADH respectively. (C) PARAFAC scores plotted versus the true concentrations.

Figure 7: (A) Excitation spectra and (B) emission spectra recovered with three-component PARAFAC (\circ) and MCR-ALS (\square) models fitted to data set 2 together with the pure excitation and emission spectra (*). (C) PARAFAC scores plotted versus the true concentrations. See Fig 6 for the 1st, 2nd and the 3rd components assignment.

Figure 8: (A) Excitation spectra and (B) emission spectra recovered with three-component PARAFAC (solid lines) and MCR-ALS (dotted lines) models fitted to data set 3 together with the pure excitation and emission spectra (dashed lines). (C) PARAFAC scores plotted versus the true concentrations. See Fig 6 for the 1st, 2nd and the 3rd components assignment.

Figure 9: (A) Excitation spectra and (B) emission spectra recovered with four-component PARAFAC model (solid lines) fitted to data set 2 together with the pure excitation and emission spectra (dashed lines). (C) PARAFAC scores plotted versus the true concentrations. See Fig 6 for the 1st, 2nd and the 3rd components assignment.

Figure 10: (A) Excitation spectra and (B) emission spectra recovered with five-component PARAFAC model (solid lines) fitted to data set 2 together with the pure excitation and emission spectra (dashed lines). (C) PARAFAC scores plotted versus the true concentrations. See Fig 6 for the 1st, 2nd and the 3rd components assignment.

Table 1: Explained variance and CORCONDIA vs the number of components for PARAFAC models of the simulated fluorescence data sets with 1-7 components.

Table 2: The mean relative error (MRE %) of concentration prediction based on the linear fit.

Figure 1

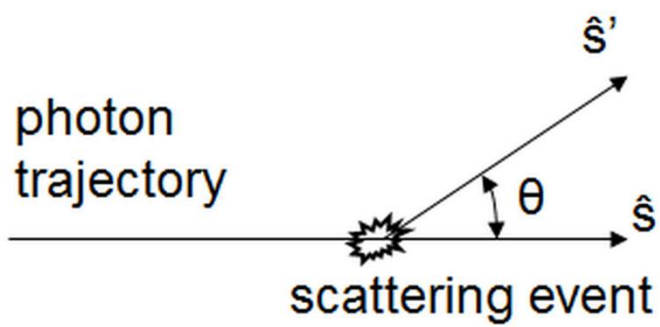


Figure 2

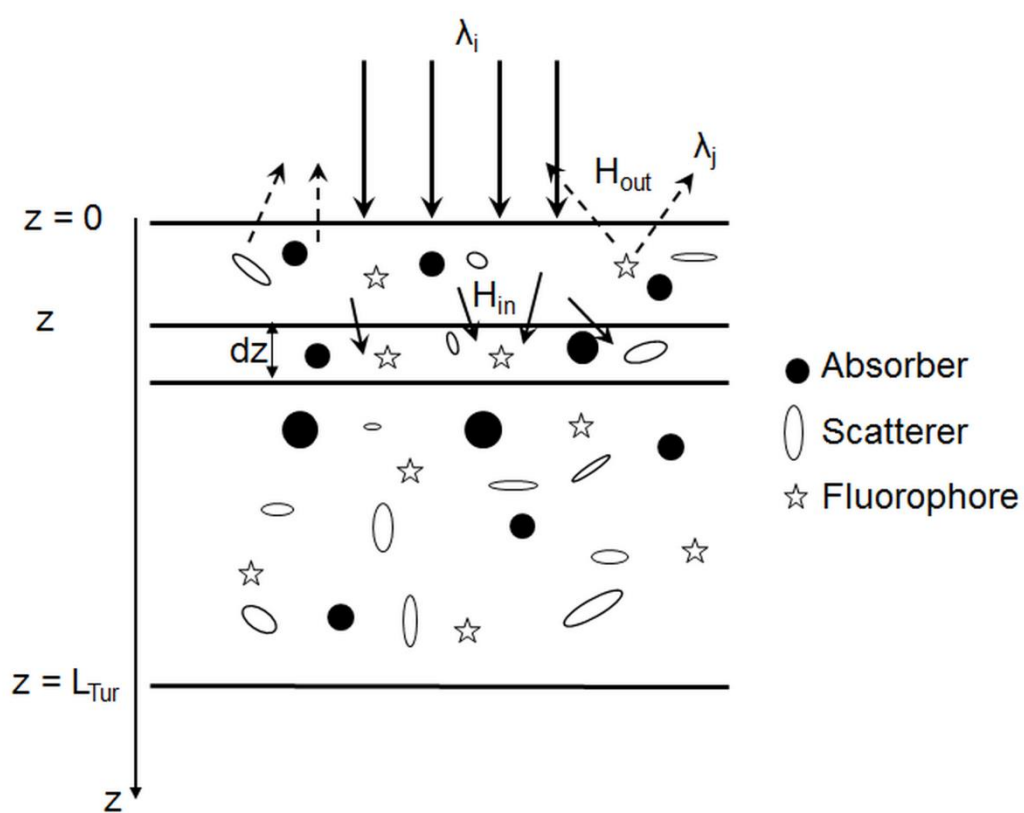


Figure 3

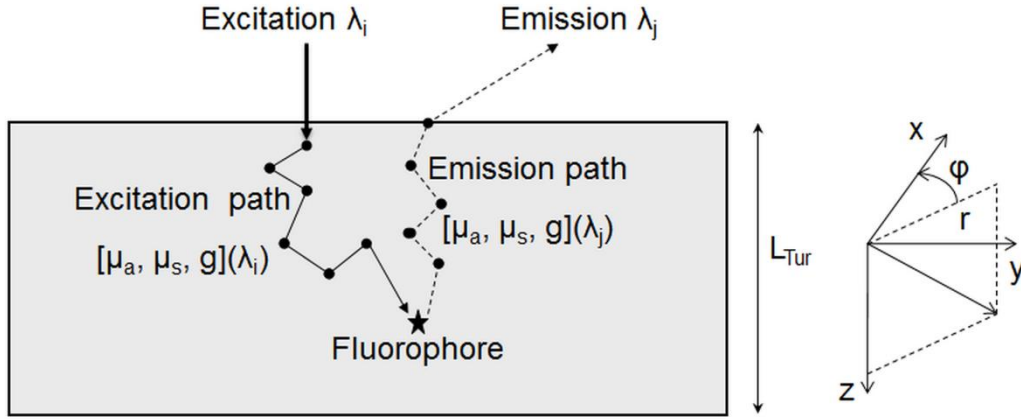


Figure 4

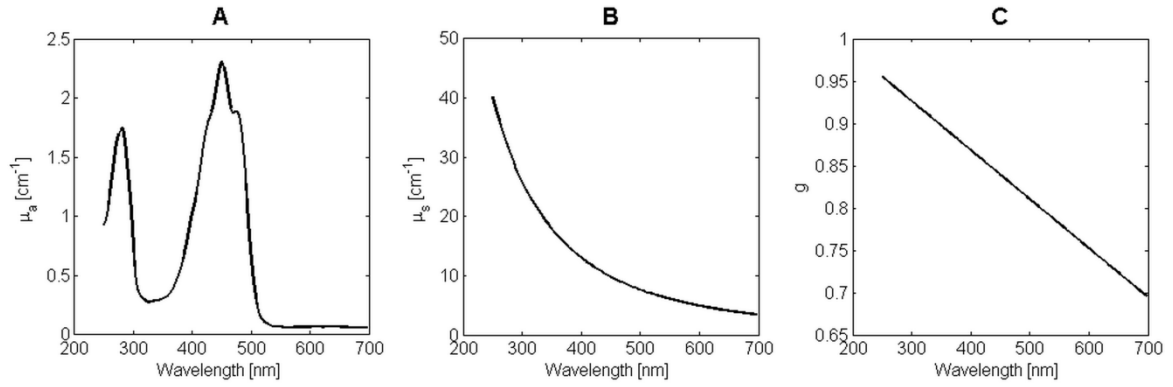


Figure 5

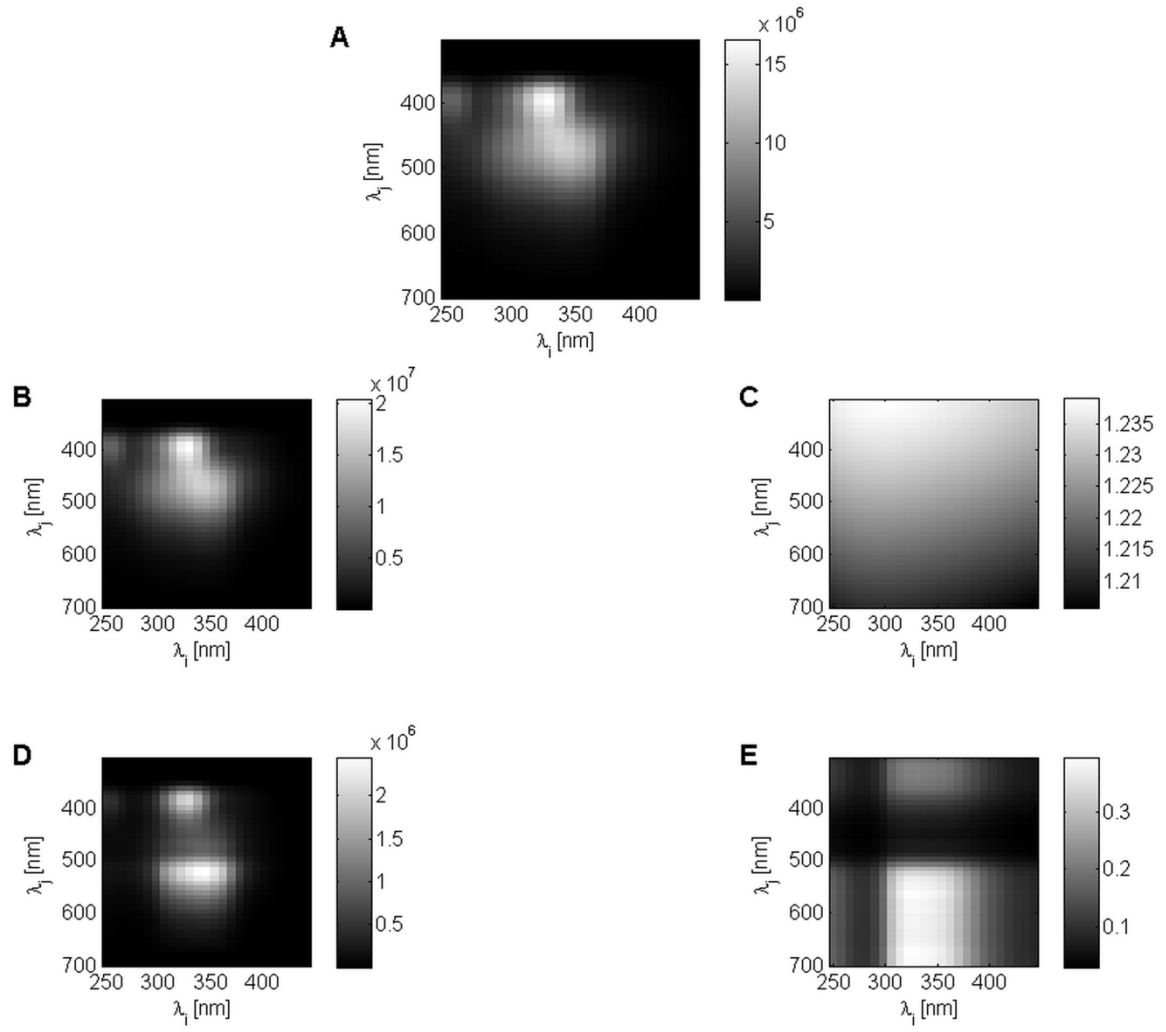


Figure 6

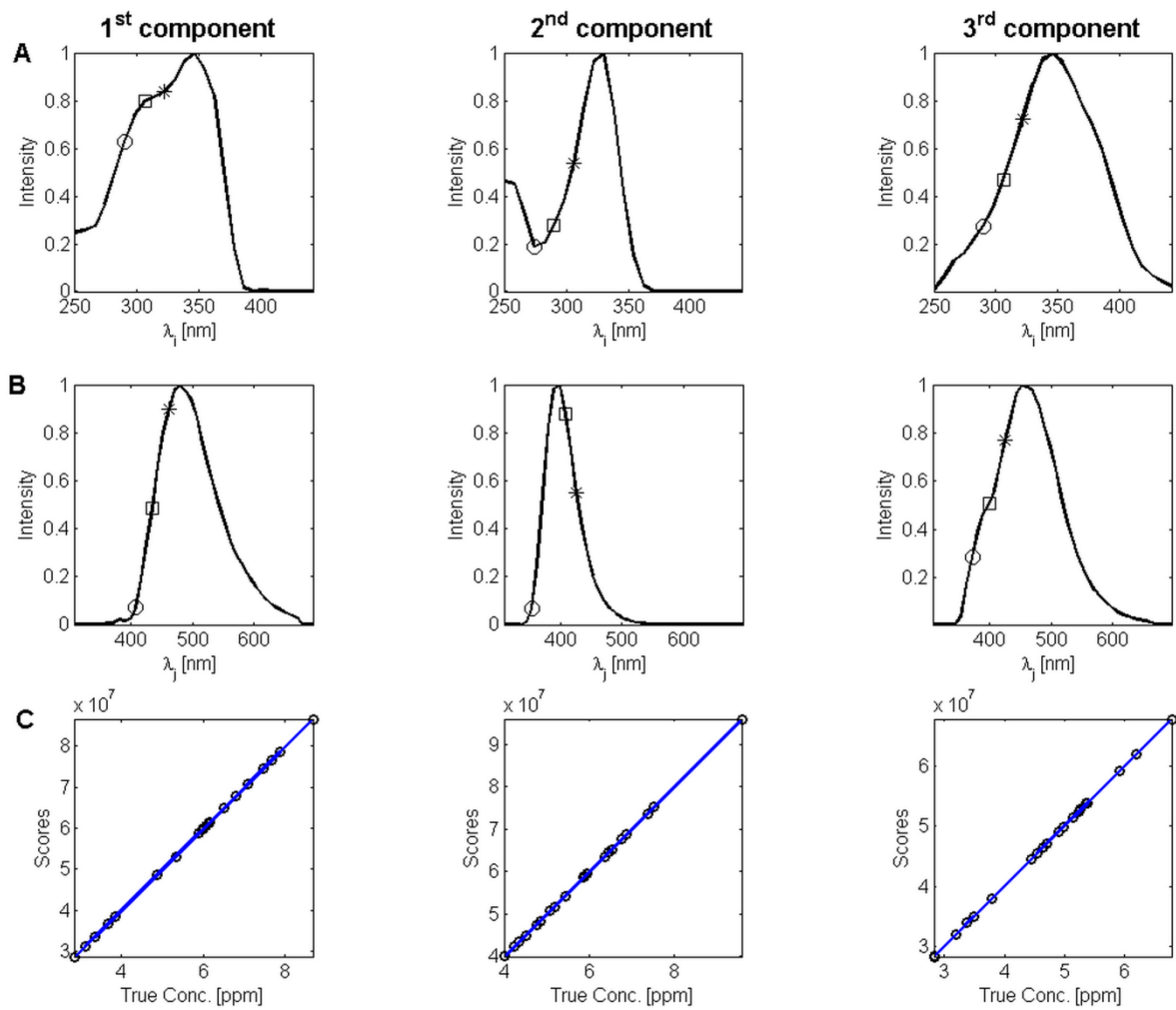


Figure 7

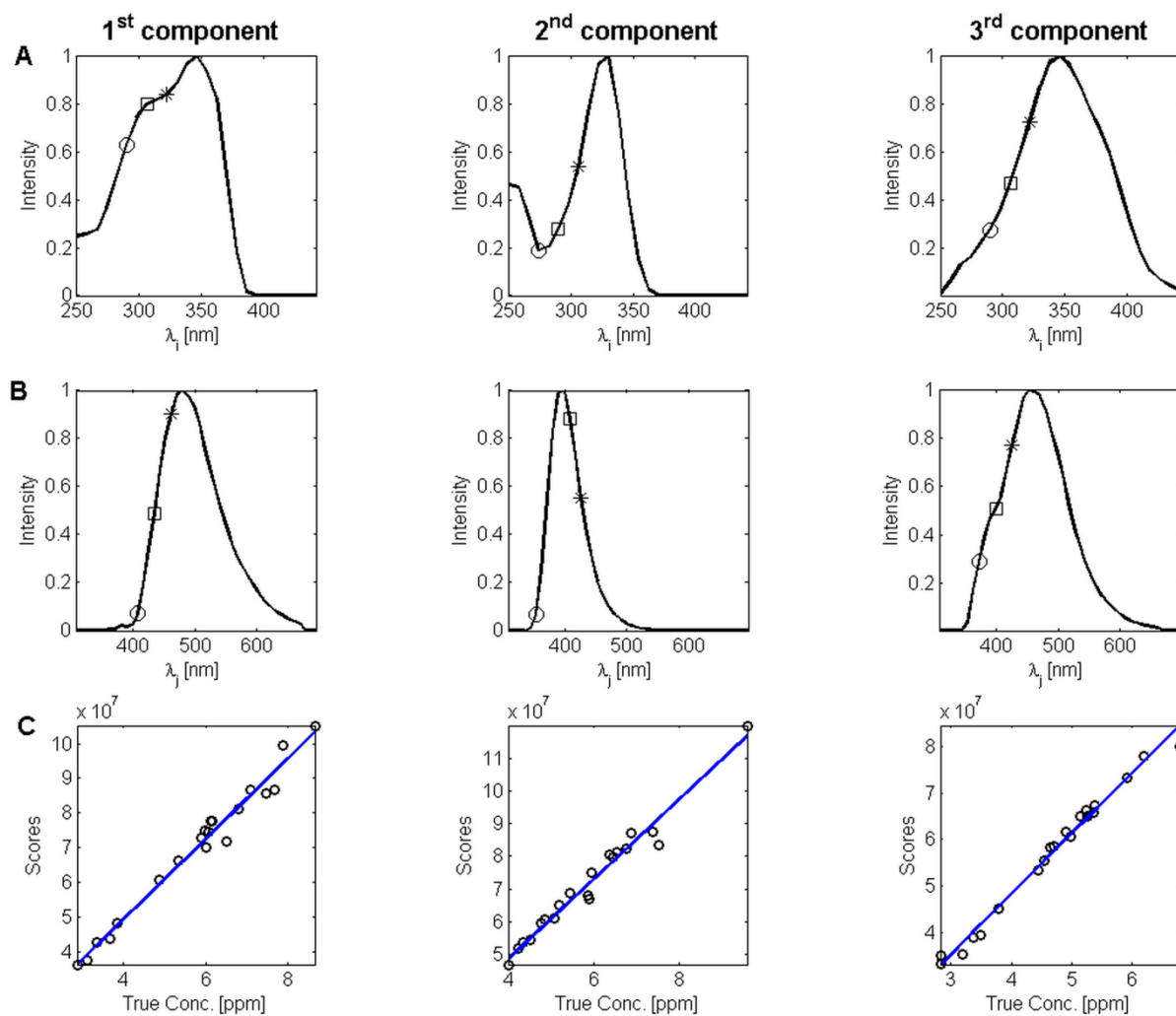


Figure 8

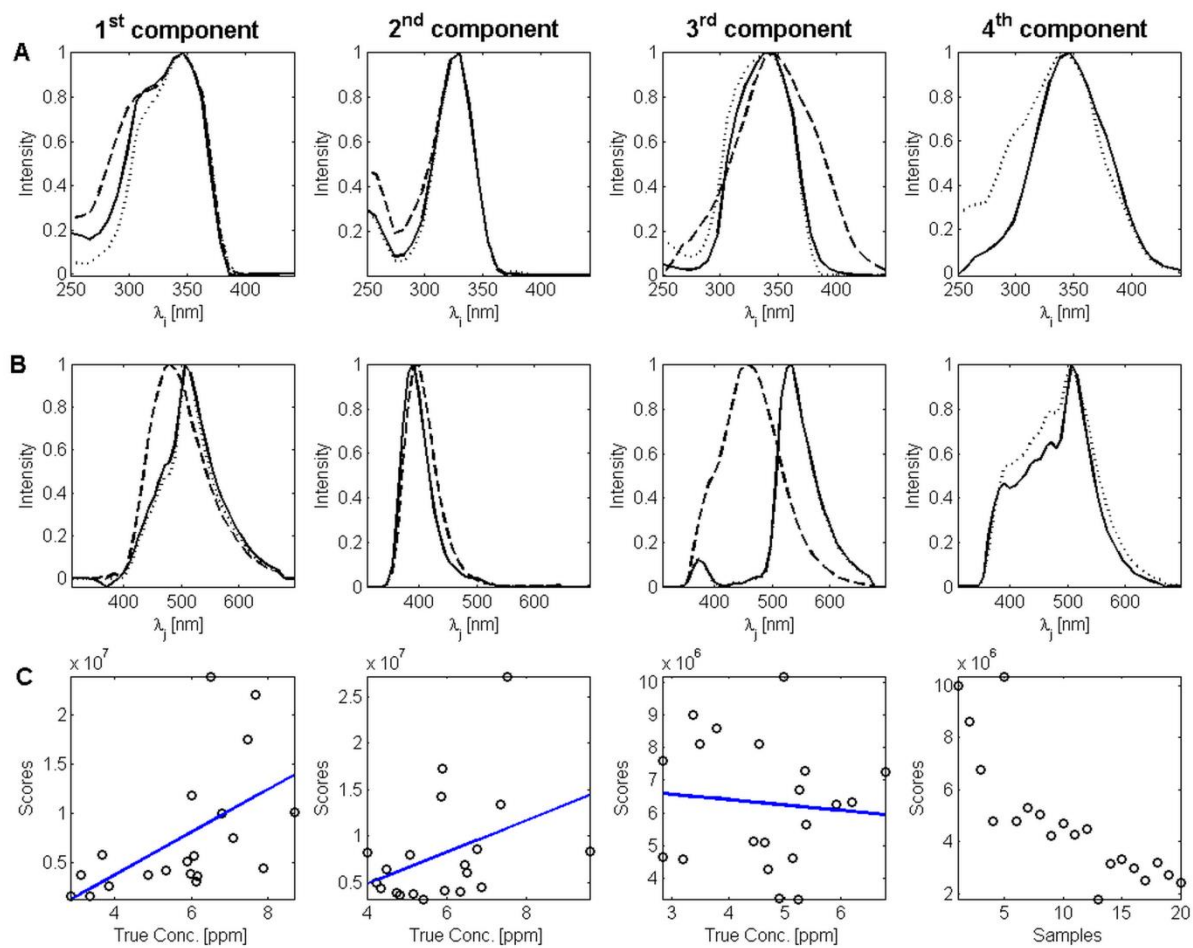


Figure 9

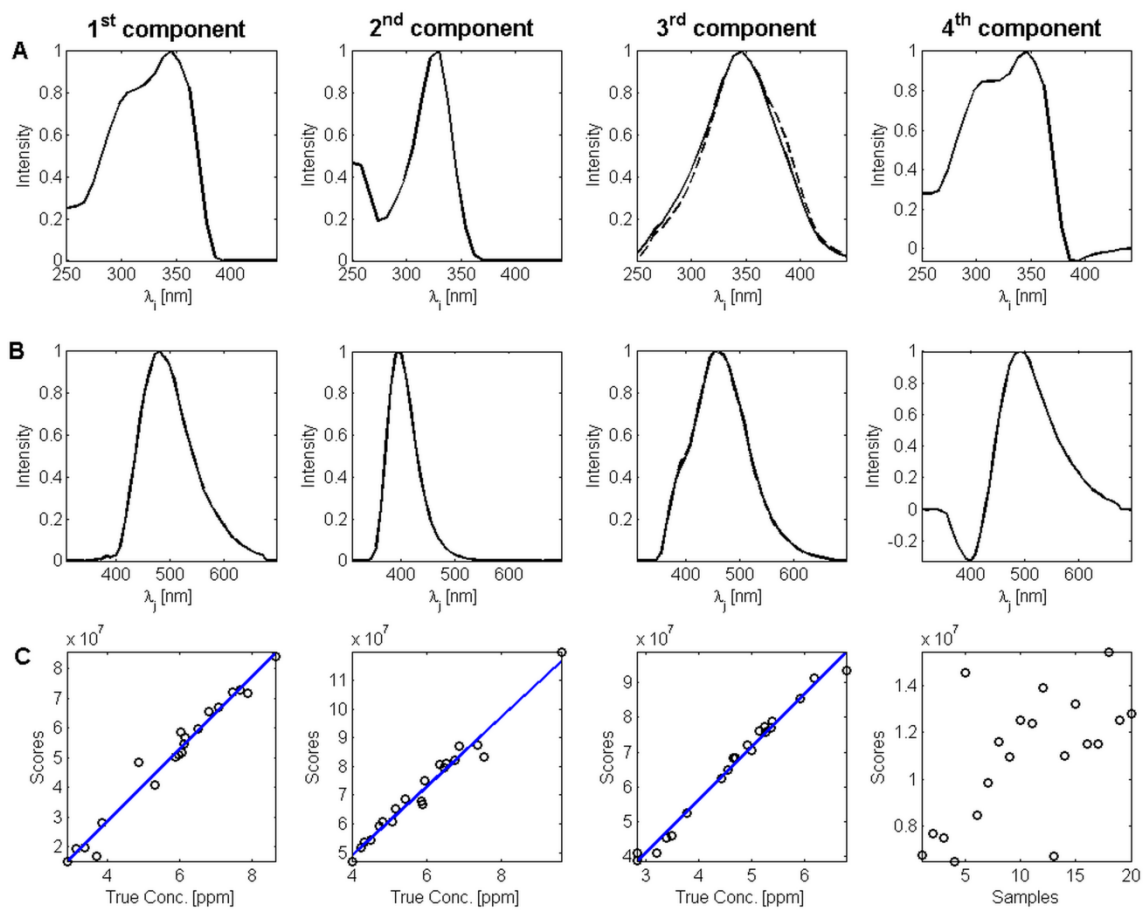


Figure 10

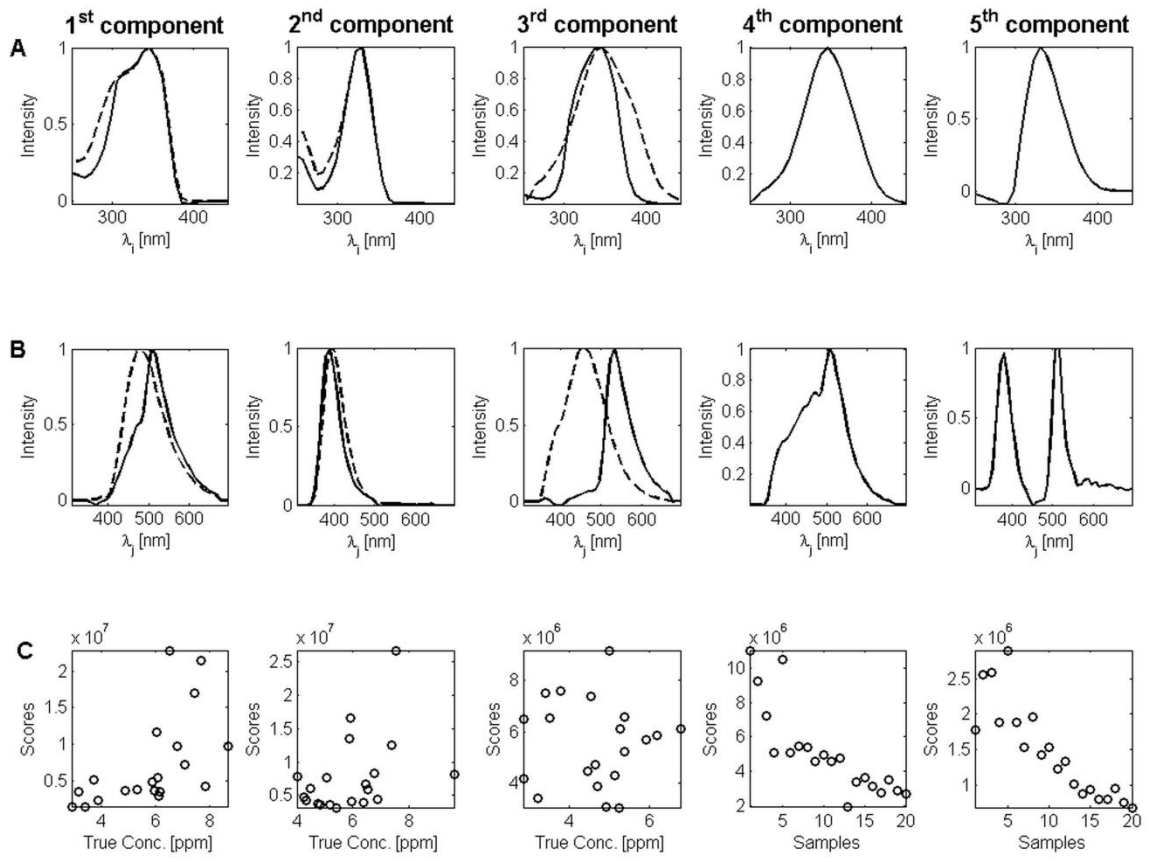


Table 1

Components	1	2	3	4	5	6	7
Data set 1							
Explained variance (%)	95.49	99.75	100.00	100.00	100.00	100.00	95.49
CORCONDIA (%)	100.00	99.97	100.00	32.26	14.74	6.55	3.08
Data set 2							
Explained variance (%)	95.49	99.76	100.00	100.00	100.00	100.00	100.00
CORCONDIA (%)	100.00	99.97	100.00	31.70	16.75	7.03	4.21
Dataset 3							
Explained variance (%)	94.67	99.43	99.87	99.98	99.99	99.99	100.00
CORCONDIA (%)	100.00	99.99	89.76	87.43	19.65	0.41	0.15

Table 2

Components	1st	2nd	3rd
Data set 2			
MRE (%)	3.30	3.11	2.21
Dataset 3			
RE (%)	29.58	38.04	233.86