Talanta 131 (2015) 170–174

Contents lists available at ScienceDirect

Talanta

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

Multilevel analysis of temperature dependent near-infrared spectra

Ruifeng Shan, Yue Zhao, Mengli Fan, Xiuwei Liu, Wensheng Cai, Xueguang Shao *

Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), State Key Laboratory of Medicinal Chemical Biology, Research Center for Analytical Sciences, College of Chemistry, Nankai University, Tianjin 300071, China

ARTICLE INFO

Article history:
Received 14 June 2014
Received in revised form
23 July 2014
Accepted 26 July 2014
Available online 12 August 2014

Keywords:
Near-infrared spectra
Temperature
Multilevel simultaneous component analysis
Quantitative spectra-temperature relationship
Composition of solvent

ABSTRACT

Quantitative spectra-temperature relationship (QSTR) between near-infrared (NIR) spectra and temperature has been used for quantitative determination of the compositions in mixtures. In this work, QSTR is studied using multilevel simultaneous component analysis (MSCA) and the spectral data of the samples with different concentrations measured at different temperatures. MSCA model contains a between-individual model describing the differences between the individuals and a within-individual model capturing the differences within the data of all the individuals. NIR spectra of five different compositions (water–ethanol–isopropanol) measured at seven temperatures were analyzed. A between-temperature model describing the effect of temperature and a within-temperature model describing the variation of concentration were obtained, from which QSTR model is established and quantitative analysis is achieved. Furthermore, the difference between the between-temperature or within-temperature models of different mixtures is used to study the composition of the solvent.

1. Introduction

Near-infrared (NIR) spectra show not only isolated molecular structure and functional groups but also inter- or intra-molecular hydrogen bonds [1]. Variations of temperature change the extent of hydrogen bonding [2], and therefore the NIR spectra measured under different temperatures can reflect these changes. In order to investigate water structure, the dynamic behavior of water was studied using NIR spectra measured under different temperatures [3–5]. Structural analysis of proteins was performed based on the effect of temperature on NIR spectra [6,7]. These studies proved that the technique is helpful to understand the influence of water in biological systems. Moreover, NIR spectra have been adopted to study the interactions in hydrogen bonds based on temperature-induced spectral variations for aqueous solutions of alcohols [8–11]. Temperature dependent NIR spectra were also used to quantitative analysis using the three-way spectral dataset and high order chemomteric approach [12]. The temperature of polyamide 66 was predicted based on the temperature-induced spectral variations [13]. In our previous works, a quantitative spectra-temperature relationship (QSTR) model between NIR spectra and temperature was established using partial least squares (PLS) regression and applied to the quantitative determination of the compositions in mixtures, including water, methanol, ethanol and ethylenediamine [14,15].

Multilevel simultaneous component analysis (MSCA) was proposed by Timmerman to analyze the multivariate data at different levels [16]. The method was developed based on simultaneous component analysis (SCA) that can simultaneously analyze data of the samples belonging to multiple populations [17]. However, SCA model is focused on the variations within the populations but discards the variation between the populations [18–20]. Therefore, MSCA is developed to give a summary of the different types of variation at different levels in the data. A two-level MSCA model of multiple individuals contains a between-individual (the first level) model and a within-individual (the second level) model. According to the degree of the differences between individuals, MSCA-P, MSCA-PF2, MSCA-IND and MSCA-ECP models have been developed [16]. Among the models, MSCA-P model is computationally simpler, and thus it has been successfully applied in different fields, such as psychology [21,22], metabolomics [23–25], and process data analysis [26–28], etc. These applications proved that the model can give a view on different types of variation in the data and increase interpretability of the data.

In this work, MSCA was used to investigate temperature-induced and concentration-induced spectral variations in NIR spectra. QSTR was established by the temperature coefficients in the between-temperature model, and quantitative analysis was achieved by the concentration coefficients in the within-temperature model. Furthermore, the difference between both the temperature and concentration coefficients of different mixtures was used to study the composition of

* Corresponding author at: College of Chemistry, Nankai University, Tianjin 300071, China. Tel.: +86 22 23503430; fax: +86 22 23502458.
E-mail address: xshao@nankai.edu.cn (X. Shao).

http://dx.doi.org/10.1016/j.talanta.2014.07.081
0039-9140/

solvent. The mixture of isopropanol–ethanol–water was used in this paper because different hydrogen bonds exists in the system.

2. Experimental

2.1. Reagents and sample preparation

Ethanol and isopropanol (Concord Technology Co., Ltd. Tianjin, China) are of analytical grade. Double distilled water was used. Five groups of mixture solutions were prepared. In each group, water–ethanol of different volume ratio was used as the solvent and isopropanol of different volume fraction was used as the solute. The volume ratios of water–ethanol are 8:2, 7:3, 5:5, 3:7 and 2:8 for the five groups, respectively, and volume fraction of isopropanol for each group of samples changes from 10% to 90% with a step of 10%.

2.2. Temperature control and spectral measurement

The temperature in the experiment was controlled by a Premium Heated Immersion Circulator (Thermo Fisher Scientific, New Hampshire, USA). The precision of the equipment for temperature control is ± 0.01 °C. In each experiment, the temperature changes from 60 to 30 °C with a step of -5 °C. The spectrum at each temperature was measured when the temperature does not change for 30 min.

All NIR spectra were measured from 4000 to 12000 cm^{-1} by a Vertex 70 spectrometer (Bruker Optics Inc, Ettlingen, Germany). The spectrometer was furnished with a transmittance optical fiber probe (Bruker Optics Inc, Ettlingen, Germany) of optical path 2 mm. A tungsten–halogen light source and InGaAs detector were used. The spectra are digitalized with ca. 4 cm^{-1} interval in Fourier transform. In the calculations, the variables from 5500 to 7803 cm^{-1} (598 data points) were used. To increase signal to noise ratio, both air reference and the spectra were measured with scan number 64.

As examples, Fig. 1(a) shows the measured NIR spectra of pure water, ethanol, isopropanol and their mixture under three different temperatures, and Fig. 1(b) shows the spectra of five water–ethanol–isopropanol mixtures with different water–ethanol volume ratio and 40% (v/v) isopropanol under seven temperatures. Clearly, the difference between the spectra measured at different temperatures can be seen from Fig. 1(a). Moreover, an isosbestic point can be found in the spectra of each sample and the isosbestic points of ethanol, isopropanol and the mixture appear at lower wavenumbers compared to that of water. In Fig. 1(b), a regular change can be found and the isosbestic points of the five mixtures shift to low wavenumber as the water content decreases. Therefore, both temperature- and concentration-induced spectral variation are included in the spectra. Method for extracting the two kinds of variations is needed to analyze the samples.

3. Calculations

3.1. Organization of the data

In MSCA, the structure of data determines the information obtained from the model. In this study, for investigating the effect of temperature and concentration on the spectra, the spectra of a group samples measured at each temperature were taken as an individual, including the spectra of the samples with nine isopropanol concentrations from 10% to 90%. Therefore, the data of each group contains seven individuals, and each individual includes nine spectra. Five such data were obtained for the five groups of samples. In the following description, $\mathbf{X}_{\mathrm{raw}}$ denotes the spectra of a group, and $\mathbf{X}_{\mathrm{raw},\ i}$ represents the spectra of an individual, where $i=1\ldots I$, I is the number of individual. The individual includes $K_i \times J$ spectra, where K_i is the number of spectra and J is the number of variables in the spectra.

3.2. Multilevel simultaneous component analysis

MSCA can be used if the variation in the data occurs on different levels simultaneously. A two-level MSCA model contains two models to explain the between-individual and within-individual difference in the data. The model for an individual i can be represented by [16,22]

$$\mathbf{X}_{raw,i} = \mathbf{1}_{K_i}\mathbf{m}^T + \mathbf{1}_{K_i}\mathbf{t}_{b,i}^T\mathbf{P}_b^T + \mathbf{T}_{w,i}\mathbf{P}_w^T + \mathbf{E}_{MSCA,i} \quad (1)$$

subjecting to the following constraints:

$$\begin{cases} \sum_{i=1}^{I} K_i \mathbf{t}_{b,i}^T = \mathbf{0} \\ \mathbf{1}_{k_i}^T \mathbf{T}_{w,i} = \mathbf{0} \end{cases} \quad (2)$$

where $\mathbf{1}_{K_i}$ denotes a column vector of ones with size K_i, $\mathbf{m}^{\mathrm{T}}$ is the overall mean of $\mathbf{X}_{\mathrm{raw}}$, $\mathbf{t}_{b,i}^T$ and $\mathbf{T}_{w,i}$ represent the between-individual and within-individual scores, $\mathbf{P}_b$ and $\mathbf{P}_w$ are the between-individual and within-individual loadings, respectively, $\mathbf{E}_{MSCA,i}$ denotes the matrix of residuals, and $\mathbf{0}$ represents a zero vector. By imposing the constraints on the scores, the three parts of the

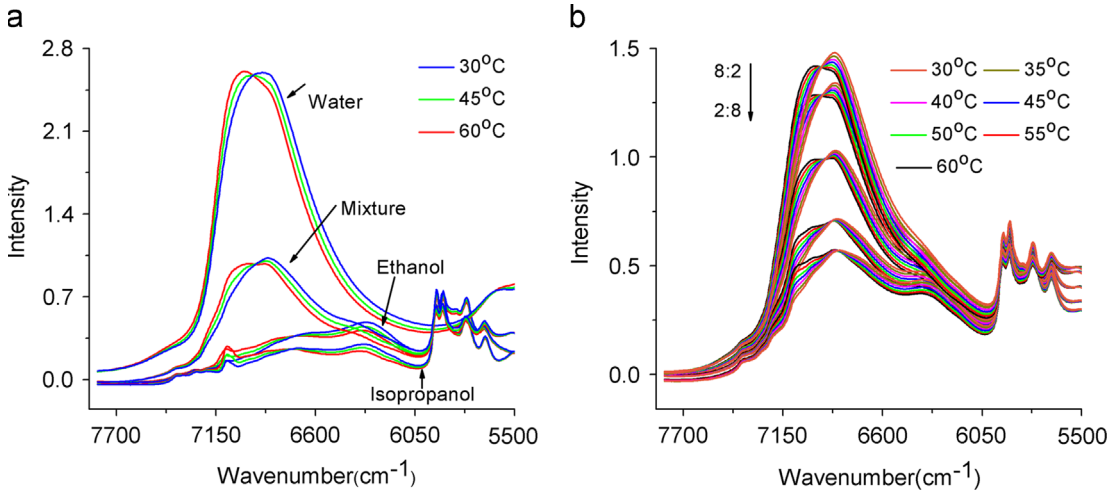

Fig. 1. Measured NIR spectra of pure water, ethanol, isopropanol and their mixture at three temperatures (a) and five groups of mixtures at seven temperatures (b).