Simultaneous Spectrophotometric Determination of Group B Vitamins Using Parallel Factor Analysis: PARAFAC

Jahanbakhsh Ghasemi* and Bahman Abbasi
Department of Chemistry, Faculty of Sciences, Razi University, Kermanshah, Iran

In this work, a simple and rapid analytical procedure was applied for simultaneous determination of folic acid (vitamin B₉), thiamin (vitamin B₁), riboflavin (vitamin B₂) and pyridoxal (vitamin B₆) based on the absorbance data in the pH range 2.0-12.0 at 25 °C using parallel factor analysis (PARAFAC). The effect of the pH as the most important factor on the sensitivity of the determination was studied. The spectral data were recorded in 400-650 nm intervals and a 2-12 pH range for all four vitamins. The calibration set was constructed in the concentration ranges of 4-22, 1-20, 6-26, and 4-20 μg mL⁻¹ for B₆, B₂, B₁ and B₉, respectively. The root mean squares errors of prediction for the prediction set, (RMSEP), are 0.65, 0.63, 1.13 and 0.34 for B₆, B₂, B₁ and B₉, respectively. The recovery percent for the validation set are in the range of 90.6 to 107.0%. The effect of the experimental conditions and diverse species were discussed. The optimum values of these factors were searched according to the relative standard deviation of the prediction set of mixtures solutions.

Keywords: N-way model; PARAFAC; Simultaneous spectrophotometric determination; Vitamin B group.

INTRODUCTION

Vitamins are organic compounds that perform specific and necessary functions in humans. The mixture of vitamins such as vitamin B complex and multivitamins in tablets and other pharmaceutical preparations are used in the treatment of several diseases. Therefore, the simultaneous determination of mixtures of vitamins is very useful in the pharmaceutical industry. There are many common methods for simultaneous determination of these vitamins such as Kalman-filtering, UV-Vis spectrophotometry,¹⁻³ variable-angle synchronous spectrofluorimetry,⁴ flow injection fluorimetry,⁵ derivative spectrophotometry, ratio spectra derivative⁶⁻⁹ and the three-dimensional fluorescence spectrometry by detecting total fluorescence intensity.¹⁰

The three-way methods such as the parallel factor analysis (PARAFAC) model¹¹⁻¹⁴ have recently gained widespread applications in the field of chemometrics. PARAFAC is one of several decomposition methods for multi-way data blocks that originated from psychometrics and conceptually can be compared to bilinear PCA, or rather it is one generalization of bilinear PCA.¹¹ The model was independently proposed by Harshman,¹⁵ and by Carroll and Chang¹⁶ who named the model canonical decomposition, CANDECOMP. Multi-way data are characterized by several sets of variables that are measured in a crossed fashion. Chemical examples could be fluorescence emission spectra measured at several excitation wavelengths, fluorescence lifetime measured at several excitation and emission wavelengths, time-resolved UV-Vis spectra for kinetic systems, and for pH-dependent compounds at different pH values, or any kind of spectrum measured chromatographically for several samples.¹¹ The tables obtained under various conditions can be stacked providing a cubic arrangement of data (a parallelepiped whose lines can be objects, whose columns are variables and slices conditions).

In continuation of our recent work to develop the simultaneous determination of the vitamin B group in the synthetics and real expenditures by chemometrics methods,¹⁷,¹⁸ in the present work, we have studied the simultaneous determination of the vitamin B group using a three-way model according to pH dependence of their spectral absorption data by PARAFAC method.

THEORY

Decomposition of higher order data with respect to a...
bilinear model has several advantages. One of the most advantageous of the three or higher way methods is the uniqueness of the decomposition process and this is the so-called ambiguity free solution. PARAFAC is a decomposition method that can be considered as one possible generalization of PCA to higher order arrays.\textsuperscript{19} It can be considered as a version of the more general method tucker\textsuperscript{30-22} with an identity core matrix. It is less flexible, uses fewer degrees of freedom and provides a unique solution independent of spectroscopic data. For three-way data, PARAFAC decomposes the original data into tri-linear components, each component consisting of one score vector and two loading vectors. Ideally then, PARAFAC resolves the true underlying spectra of each compound in a mixture, when the correct number of factors is chosen for the PARAFAC model.\textsuperscript{11} A PARAFAC model for \textit{F} components is defined by a score matrix \textbf{A}, and loading matrices \textbf{B} and \textbf{C} (also sometimes called first, second and third mode loadings, respectively) with respective elements \(a_{ij}, b_{ij}\) and \(c_{ij}\) (Eq. (1)). The tri-linear model is found to minimize the sum of squares of the residues, \(e_{ijk}\) in the model,\textsuperscript{12} which is presented as follows:

\[
X_{ijk} = \sum a_{ij} b_{ij} c_{ij} + e_{ijk} \tag{1}
\]

where \(X_{ijk}\) is the absorption intensity measured in sample \(i\) at absorbance wavelength \(j\) and pH \(k\), \(e_{ijk}\) holds the residuals of the model. The algorithms used are most often based on alternating least squares (ALS) initialized by either random values or calculated by a direct trilinear decomposition based on a generalized eigenvalue problem.\textsuperscript{11}

**EXPERIMENTAL**

**Reagents**

Vitamins B group (B\textsubscript{0}, B\textsubscript{1}, B\textsubscript{2}, B\textsubscript{6}) were purchased from Merck; all remaining reagents were of analytical reagent grade and used as received. Stock solution of each vitamin was prepared by dissolving 50 mg of vitamin in a 50 mL volumetric flask and diluted to the mark with water. Working mixtures were made from stock solutions at random in a concentration range of 4-22, 1-20, 6-26, and 4-20 g.mL\textsuperscript{-1} for B\textsubscript{6}, B\textsubscript{2}, B\textsubscript{1} and B\textsubscript{0}, respectively. Titration of solutions of mixtures of four vitamins were performed in 0.5 pH increments from pH 2 to 12, the high concentrated solution of NaOH and HCl used for pH adjustment, avoiding thereby dilution of working solutions.

**Instrumentation and software**

The pH values were measured on a HANA pH-meter model 300, calibrated with at least two standard buffer solutions (4.0 and 9.0). All measurements were carried out at room temperature ~25 °C. An Agilent UV-Visible diode array spectrophotometer was utilized, and Agilent UV-Visible ChemStation software was used for data acquisition. A quartz cell of 1.00 cm optical pathlength was used for all measurements.

All the data handling was performed in the MATLAB, (6.5 ver, The Math works, Natick, USA).

**Parallel factor analysis**

The obtained data can be arranged as a cube (with absorption wavelength, samples and pH variations as the axes) constituting a third-order array (three-way data). PARAFAC decomposes higher order arrays (nth order or nth way) into multi-linear (tri-linear in this case) components. This is analogous to the decomposition of bilinear data (two-way data) by PCA. Instead of obtaining one loading matrix for each model, as in PCA, PARAFAC produces two loading matrices corresponding to absorption spectra and concentration profiles of the different species and a score matrix for samples. PARAFAC is thoroughly described by Bro,\textsuperscript{11,23} both from a theoretical viewpoint and with respect to model evaluation and validation. The MATLAB codes to calculate PARAFAC model download from Ref. 24. PARAFAC modeling was carried out by using “The N-way Toolbox for MATLAB”, 2.1 version (R. Bro, Food technology, Copenhagen, Denmark).\textsuperscript{25}

**RESULTS AND DISCUSSION**

Fig. 1 shows the spectra of pure solutions of vitamin B\textsubscript{0}, B\textsubscript{1}, B\textsubscript{2}, and B\textsubscript{6} at different pH. As can be observed, B\textsubscript{6} presents \(\lambda_{\text{max}}\) at 292 nm in acid solutions (pH < 2.5), and the vitamins B\textsubscript{2}, B\textsubscript{1} and B\textsubscript{0} present \(\lambda_{\text{max}}\) at 372, 246 and 285 nm, respectively. The increase of pH, causing a considerable shift in \(\lambda_{\text{max}}\), is clear evidence of the presence of the other deprotonated forms of these vitamins. The presence of isosbestic points show the variously protonated forms and they relate to each other with protonated/deprotonated equilibria. The number of released protons in the indicated pH intervals are 1, 2, 3 and 3 for riboflavin, pyridoxal, thiamine and folic acid, respectively.\textsuperscript{26} All the molecules show the distinct isosbestic points in a regular changing in pH. As is clear, the system shows a high degree of spectral overlapping of the constitut-
ents. To overcome such problems there are several chemometrics approaches. Here, PARAFAC was utilized to study the simultaneous determination of vitamins B group, through three way spectral deconvolution.

The data was arranged in a three-way array $22 \times 276 \times 8$ composed of 22 solutions, with different $B_0$, $B_1$, $B_2$ and $B_6$ concentrations, in the rows (as listed in Table 1), 276 wavelengths in the columns and 8 pH values (2.50, 3.00, 4.50, 5.50, 7.00, 8.50, 10.00, 11.00) in the slices. The selection of the pH values was according to the lower RMSEP for the prediction step of all four constituents. Therefore, the indicated pH values give the lowest RMSEP values. No preprocessing (centering or auto scaling) and no constraints were applied to the data matrix. SVD vectors were used for PARAFAC initialization, and the convergence criterion was $1 \times 10^{-6}$.

In addition to the data set, PARAFAC requires the definition of the number of factors to be included in the model. To determine the proper number of components, the core consistency diagnostic method was used. Six component PARAFAC decomposition yielded a stable solution. Fig. 2 shows the loadings of second and third modes of the resulting model.

The first mode loadings of the PARAFAC model, $C$, corresponds to the sample mode. The $C$-loadings are the relative concentrations of the B group vitamins in the mixtures. In the calibration step, these loadings are regressed against the real concentrations of each vitamin in the mixture to get a linear calibration line. In the prediction step, this regression line can then be used to predict (if no new interferent is present) the concentrations of B group vitamins in future test samples, $X_{\text{in}}$, that are not in the initial calibration dataset, by interpolating their loadings of relative concentration, $C^T$. These loadings can be calculated in advance from the following eqs.:

$$X_i = (B_0 \otimes A)C^T = ZC^T$$

$$X_i : (I \times K)$$

$$B : (J \times F)$$

$$A : (I \times F)$$

$$C^T : (K \times F)$$

$$Z : (I \times K)$$

multiplying the pseudoinverse of the $Z$ matrix by the test sample data, as shown in Eq. (3):

$$C^T = (Z^T Z)^{-1} Z^T X_{\text{in}}$$
Another way to predict future test samples is to include them in the initial PARAFAC model. In this way, the loading matrices for both the calibration and prediction sets are recovered. All the samples are considered to calculate the model parameters, although the regression fit is only performed with the calibration samples. Finally, the loadings of the PARAFAC model for the prediction samples are interpolated into the corresponding regression line to obtain the predicted concentration of each analyte.

Table 2 shows the standard and predicted concentration values of validation set. The predictive capability of a regression model can be defined in various ways. The most general expression is the root mean squares errors of prediction (RMSEP) of the validation set employed for comparison that is given by:

$$RMSEP = \sqrt{\frac{\sum (C_r - C_f)^2}{n}}$$

where $C_r$ is the added concentration of analyte, $C_f$ is the found concentrations of analyte and $n$ is the total number of synthetic mixture. The values of RMSEP for vitamins $B_6$, $B_1$, $B_2$ and $B_0$ are summarized in Table 2.

<table>
<thead>
<tr>
<th>Sample No.</th>
<th>$B_6$ ($\mu$g mL$^{-1}$)</th>
<th>$B_2$ ($\mu$g mL$^{-1}$)</th>
<th>$B_1$ ($\mu$g mL$^{-1}$)</th>
<th>$B_0$ ($\mu$g mL$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.00</td>
<td>14.00</td>
<td>10.00</td>
<td>4.00</td>
</tr>
<tr>
<td>2</td>
<td>20.00</td>
<td>14.00</td>
<td>6.00</td>
<td>10.00</td>
</tr>
<tr>
<td>3</td>
<td>20.00</td>
<td>1.00</td>
<td>18.00</td>
<td>10.00</td>
</tr>
<tr>
<td>4</td>
<td>4.00</td>
<td>16.00</td>
<td>18.00</td>
<td>10.00</td>
</tr>
<tr>
<td>5</td>
<td>4.00</td>
<td>16.00</td>
<td>10.00</td>
<td>14.00</td>
</tr>
<tr>
<td>6</td>
<td>4.00</td>
<td>8.00</td>
<td>26.00</td>
<td>14.00</td>
</tr>
<tr>
<td>7</td>
<td>10.00</td>
<td>1.00</td>
<td>26.00</td>
<td>14.00</td>
</tr>
<tr>
<td>8</td>
<td>10.00</td>
<td>14.00</td>
<td>6.00</td>
<td>20.00</td>
</tr>
<tr>
<td>9</td>
<td>10.00</td>
<td>1.00</td>
<td>18.00</td>
<td>20.00</td>
</tr>
<tr>
<td>10</td>
<td>14.00</td>
<td>8.00</td>
<td>26.00</td>
<td>4.00</td>
</tr>
<tr>
<td>11</td>
<td>14.00</td>
<td>16.00</td>
<td>10.00</td>
<td>4.00</td>
</tr>
<tr>
<td>12</td>
<td>14.00</td>
<td>8.00</td>
<td>6.00</td>
<td>20.00</td>
</tr>
<tr>
<td>13</td>
<td>20.00</td>
<td>8.00</td>
<td>18.00</td>
<td>4.00</td>
</tr>
<tr>
<td>14</td>
<td>18.00</td>
<td>12.00</td>
<td>14.00</td>
<td>8.00</td>
</tr>
<tr>
<td>15</td>
<td>18.00</td>
<td>10.00</td>
<td>8.00</td>
<td>17.00</td>
</tr>
<tr>
<td>16</td>
<td>8.00</td>
<td>10.00</td>
<td>13.00</td>
<td>18.00</td>
</tr>
<tr>
<td>17</td>
<td>16.00</td>
<td>4.00</td>
<td>22.00</td>
<td>18.00</td>
</tr>
<tr>
<td>18</td>
<td>22.00</td>
<td>2.00</td>
<td>16.00</td>
<td>16.00</td>
</tr>
<tr>
<td>19</td>
<td>16.00</td>
<td>6.00</td>
<td>14.00</td>
<td>12.00</td>
</tr>
<tr>
<td>20</td>
<td>6.00</td>
<td>20.00</td>
<td>8.00</td>
<td>12.00</td>
</tr>
<tr>
<td>21</td>
<td>6.00</td>
<td>18.00</td>
<td>12.00</td>
<td>12.00</td>
</tr>
<tr>
<td>22</td>
<td>8.00</td>
<td>18.00</td>
<td>16.00</td>
<td>8.00</td>
</tr>
</tbody>
</table>

Table 2. Added and found results of the synthetic mixtures of $B_6$, $B_1$, $B_2$ and $B_0$ by PARAFAC model

<table>
<thead>
<tr>
<th>Mixture No.</th>
<th>$B_6$</th>
<th>$B_2$</th>
<th>$B_1$</th>
<th>$B_0$</th>
<th>$B_6$</th>
<th>$B_2$</th>
<th>$B_1$</th>
<th>$B_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.00</td>
<td>13.00</td>
<td>20.00</td>
<td>12.00</td>
<td>10.00</td>
<td>12.56</td>
<td>19.33</td>
<td>12.08</td>
</tr>
<tr>
<td>2</td>
<td>9.00</td>
<td>13.00</td>
<td>15.00</td>
<td>19.00</td>
<td>8.16</td>
<td>13.91</td>
<td>14.05</td>
<td>19.25</td>
</tr>
<tr>
<td>3</td>
<td>7.00</td>
<td>17.00</td>
<td>15.00</td>
<td>15.00</td>
<td>7.39</td>
<td>16.68</td>
<td>15.46</td>
<td>14.81</td>
</tr>
<tr>
<td>4</td>
<td>19.00</td>
<td>13.00</td>
<td>13.00</td>
<td>9.00</td>
<td>18.14</td>
<td>12.83</td>
<td>11.81</td>
<td>9.42</td>
</tr>
<tr>
<td>5</td>
<td>18.00</td>
<td>11.00</td>
<td>16.00</td>
<td>10.00</td>
<td>17.82</td>
<td>10.44</td>
<td>15.16</td>
<td>9.95</td>
</tr>
<tr>
<td>6</td>
<td>13.00</td>
<td>15.00</td>
<td>10.00</td>
<td>20.00</td>
<td>13.02</td>
<td>14.58</td>
<td>10.26</td>
<td>19.63</td>
</tr>
<tr>
<td>7</td>
<td>13.00</td>
<td>10.00</td>
<td>20.00</td>
<td>10.00</td>
<td>12.79</td>
<td>9.63</td>
<td>18.66</td>
<td>9.77</td>
</tr>
</tbody>
</table>

RMSEP 0.65 0.63 1.12 0.33
This method was applied to the determination of these vitamins in the synthetic mixture in the presence of diverse interfering species, and the obtained result of predicted values are listed in Table 3. The mean recovery values for these samples are also shown in Table 3. The mean recovery shows that this method is versatile for the prediction of vitamins in the presence of a large number of ingredients real fluids like urine and blood and in the pharmaceutical formulations. The large error in the case of L-cysteine is likely to result from the chemical reactivity of these molecules, and it may alter the chemical identity of the vitamins.

The linear equations and squared correlation coefficients for four vitamins are indicated along with the plots of the found concentration against the added values for the validation set in Fig. 3. As is clear, there is a fair linear relationship between two sets of the concentrations. As is usual in statistical analysis, the randomly distributed data points around the best-fitted lines are a good indication of the linear dependence between two sets of variables.

**CONCLUSION**

According to obtained results, especially RMSEP and squared correlation coefficients $R^2$, of the analysis of the validation and spiked matrix, it can be concluded the PARAFAC method is versatile for the simple and rapid simultaneous determination of the vitamin B groups in a wide range of concentration. It is worth noting that this method
used the pH as independent variables to deconvolute the spectral data and regressed with the known concentration to build an efficient model to predict the concentrations of all vitamins in synthetic and real matrix samples. Finally, combination of spectral profiles and pH dependence of spectral changes with changing pH reveals that three way methods such as PARAFAC can be applied for the determinations of similar compounds in real and spiked samples successfully.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Rasmus Bro, not only for making the algorithms available but also for sharing knowledge concerning PARAFAC.

REFERENCES

7. Surmeian, M.; Ciocodaru, G.; Ionescu, M. S.; Cosofret, V. V. Revue Roumaine De Chmie 1995, 40, 111.
PARAFAC Determination of Group B Vitamins
