# Quantification of Overlapped Peaks with Partial Least Squares Regression: Open Tubular Ion Chromatography for Sodium and Ammonium Ions

Yuta Otsuka[1,*], Hiroki Watanabe[2], Hideji Tanaka[3], and Masaki Takeuchi[3]

[1] Faculty of Pharmaceutical Sciences, Tokyo University of Science, 2641 Yamazaki, Noda, Chiba 278–8510, Japan

[2] Graduate School of Pharmaceutical Sciences, Tokushima University, 1-78-1 Shomachi, Tokushima 770-8505, Japan

[3] Institute of Biomedical Sciences, Tokushima University, 1-78-1 Shomachi, Tokushima 770-8505, Japan

**Abstract**

We propose an approach for the simultaneous determination of multiple component (*i.e.,* $Na^+$ and $NH_4^+$) solutions based on open tubular ion chromatography (OTIC) coupled with chemometrics. The sample solutions containing $Na^+$ and $NH_4^+$ (0 – 500 μM) were prepared at various concentration ratios. The samples were analyzed by OTIC within the time range of $t$ = 0 – 18.0 min. The partial least squares (PLS) regression method was applied to the acquired chromatograms in order to predict ionic concentrations. The constructed PLS regression models were able to predict $Na^+$ and $NH_4^+$ concentrations. The regression vectors of the $Na^+$ and $NH_4^+$ prediction models yielded positive peaks at 11.8 min and 13.3 min, respectively. The determination coefficient between predicted and measured cation concentration were respectively 0.9187 for $Na^+$ and 0.9314 for $NH_4^+$. The combination of the OTIC and PLS regression methods was useful for the determination of ionic concentrations under unresolved chromatogram peak conditions.

**Keywords** Flow analysis, simultaneous quantification, peak resolution, partial least squares regression

## 1. Introduction

Portable flow injection systems enable the minimization of the weight of the measurement system. Combination of the appropriate separation column allows increased temporal resolution and easy measurements. An open tubular capillary column (OTCC) is made of a silica capillary, the inner surface of which is coated with layers of stationary phase. An open tubular ion chromatographic system that needs no high-pressure pump can be expected to be applied to on-site analysis. However, the peak resolution of OTCC is generally lower than that of packed capillary column.

Maeder *et al*. reported a singular value decomposition approach, which belongs to multivariate approaches for chromatographic analyses [1]. The predictions by multi-variational analyses and/or machine learning are suitable for the simultaneous determination of ion concentrations from obtained chemical data. In previous studies, we investigated a pseudo-polymorphic transformation [2] and a metastable calcium phosphate phase transformation [3] based on vibrational spectra and a multivariate regression method. The effect of active pharmaceutical ingredients on the transformation of drugs [4,5] and transformation of bioceramics [6] were studied by a machine learning regression method.

In this study, we explore the combination of open tubular ion chromatography (OTIC) and multivariate regression. The concept is applied to the simultaneous measurements of $Na^+$ and $NH_4^+$ in sample solutions. The OTIC in combination with partial least squares (PLS) regression can accurately predict the concentrations of cations even if the resolution of their chromatographic peaks is low.

## 2. Experimental

### 2.1. Reagents and sample

Poly(butadiene-maleic acid) (PBMA) and azobisisobutyronitrile (AIBN) were purchased from Polysciences Inc., Germany, and Kishida Chemical Co. Ltd., Japan, respectively. Methanol, hydrofluoric acid, nitric acid, hydrochloric acid, tartaric acid, sodium chloride, and ammonium chloride were of analytical grade and were purchased from Kanto Chemical Co. Ltd., Japan. Sartorius Arium 611 DI grade deionized water (>18 MΩ cm) was used throughout.

Table 1 shows the concentrations of sodium ion ($Na^+$) and ammonium ion ($NH_4^+$) in sample solutions prepared. In the quantitative experimental design, the relative variance ($R^2$) of ion concentrations in standard samples should be as low as possible. In the present study, the $R^2$ of the respective concentrations was 0.065.

### 2.2. Open tubular ion chromatography

The OTCC was prepared according to the Kuban method [7]. The fused silica capillary (75 μm i.d., 363 μm o.d., 2 m long, BGB Analytik, USA) was pretreated with a mixture of 2.5% hydrofluoric acid and 2.5% nitric acid for 30 min followed by 1% hydrochloric acid for 30 min using a peristaltic pump (flow rate: 0.5 μL $min^{-1}$). The inner part of the capillary was washed

* To whom correspondence should be addressed.
E-mail: y.otsuka36156803@gmail.com

Table 1 Concentration of cations in sample solutions for simultaneous determination of ionic concentrations.

| | Concentration / μM | |
|---|---|---|
| Sample No. | $Na^+$ | $NH_4^+$ |
| 1 | 0 | 0 |
| 2 | 250 | 250 |
| 3 | 500 | 0 |
| 4 | 0 | 500 |
| 5 | 250 | 0 |
| 6 | 0 | 250 |
| 7 | 400 | 200 |
| 8 | 200 | 400 |
| 9 | 500 | 250 |
| 10 | 250 | 500 |
| 11 | 500 | 200 |
| 12 | 200 | 500 |

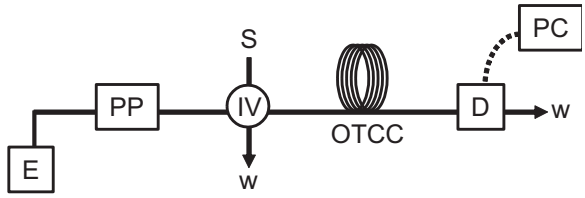


Fig. 1 Schematic of the simultaneous determination of the $Na^+$ and $NH_4^+$ using the OTIC. E, eluent; PP, peristaltic pump; IV, injection valve; OTCC, open tubular capillary column; D, detector; PC, personal computer; S, sample; w, waste.

thoroughly with water and then dried with nitrogen gas ($N_2$, >99.99%, Shikoku Taiyo Nippon Sanso Co. Ltd., Japan). The coating solution, which contained 2.4 g PBMA and 0.12 g AIBN in 10 mL methanol, used for layer construction was filled in the capillary for 5 min and then drained using $N_2$ (pressure: 200 kPa). The thin film of the coating solution was left on the capillary inner wall. The treated capillary was quenched for 15 min at 160°C to allow bonding of its surface. These procedures were repeated nine times to create a stationary phase with ten layers.

Figure 1 shows the schematic for the cation detection system. The eluent of 0.5 mM tartaric acid was aspirated from the eluent tank (E) by a peristaltic pump (PP, OG-RP-400, OGAWA & Co. Ltd., Japan) with a flow rate of 1.05 μL $min^{-1}$. The eluent solution flowed through an injection valve (IV, CN2-4346EH, Valco Instruments Co. Inc., USA), where 250 nL of cation sample was injected. The sample containing cations was then delivered to 2 m OTCC. The separated cation solution flowed through a lab-made capacitance detector (D) [8] that was equipped with a personal computer (PC) for recordings at a sampling frequency of 4.6 Hz. EVAL-AD7746EB-ND (Analog Devices Inc., USA) was applied to the detector as a capacitance to digital converter.

### 2.3. PLS regression

The concentrations of the cations in the mixed sample solutions were estimated by the constructed PLS models with the nonlinear iterative partial least-squares NIPALS algorithm [9]. The details of the NIPALS algorithm were explained elsewhere [10]. The chromatograms acquired in the period of 10.0 – 18.0 min were baseline corrected. Because the chromatographic separations were repeated four times for each sample (Table 1), 48 chromatograms could totally be used for the PLS regression. The regression allowed the estimation of a correlation between the independent ($X$: signal of detector in fF) and dependent variable(s) ($Y$: cation concentration in μM). The function of the regression is the matrix transformation that defines the original $X$ and $Y$ data matrices as the products of the loading and score matrices:

$$X = TP' + E \qquad (1)$$
$$Y = UQ' + F \qquad (2)$$

Herein, $T$ and $U$ are the score matrices, and $P$ and $Q$ are the loading matrices for $X$ and $Y$, respectively, and $E$ and $F$ are the residual matrices. There is a linear relationship between the $X$ and the $Y$ matrices, $Y = Xb$, where $b$ is the regression vector coefficient owing to the relationship of the $T$ and $U$ matrices [11]. The PLS calibration models were evaluated by cross-validation using the leave-one-out method [12]. The optimum number of factors was used so that a minimum was present in the predicted residual sum-of-squares (*PRESS*) *vs.* the PLS components graph. The former variable is defined as,

$$PRESS = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \qquad (3)$$

where $\hat{y}_i$ and $y_i$ correspond to the concentration values in each sample predicted by the chromatograms and reference, respectively. The calibration dataset of the cation samples was used only to construct the PLS regression model. The validation chromatograms were used to assess the prediction values. The root-mean-square error was calculated for the prediction or validation samples (RMSEP) and for the calibration samples (RMSEC) [13]. The regression model construction and prediction were applied using the Unscrambler software (version 10.4, CAMO Software Co. Ltd., Computer Aided Modelling, Trondheim, Norway).

## 3. Results and Discussion

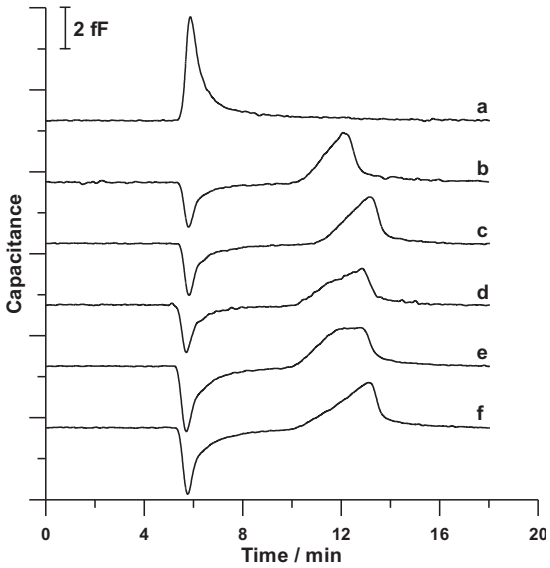


Fig. 2 Chromatograms of sample solutions. a, water; b, 500 μM $Na^+$; c, 500 μM $NH_4^+$; d, 250 μM $Na^+$ and 250 μM $NH_4^+$; e, 400 μM $Na^+$ and 200 μM $NH_4^+$; f, 200 μM $Na^+$ and 400 μM $NH_4^+$.

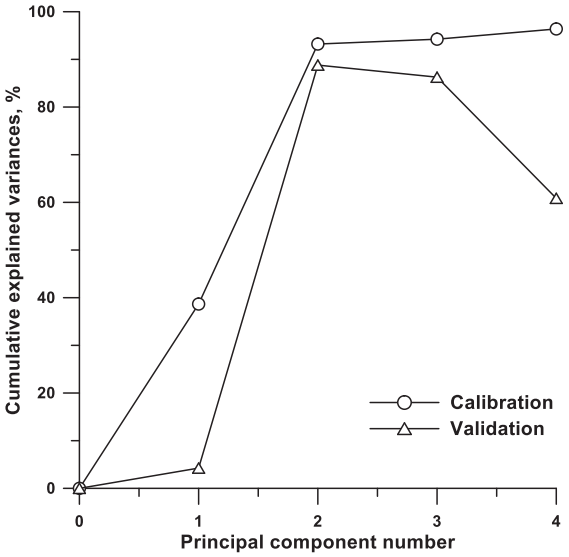


Fig. 3 Cumulative percentage variances of calibration and validation models on response variable.

3.1. Chromatograms of sample solutions

Figure 2 shows the chromatograms of sample solutions and pure bulk water solution. The chromatograms of the samples show distinct capacitance peaks at 6.0 min and between 10.0 – 15.0 min. It is indicated that the peak at 6.0 min was generated owing to the water dips. The capacitance peaks in the range of 10.0 – 15.0 min contain the $Na^+$ and/or $NH_4^+$ peaks. Based on the chromatograms of 500 μM $Na^+$ and 500 μM $NH_4^+$ samples, the resolution and separation factor were calculated to be 0.49 and 1.20, respectively. The number of theoretical plates for $Na^+$ and $NH_4^+$ were 449 and 486, respectively. These values suggest that the respective concentrations cannot be directly determined from the integration of the capacitance area. Therefore, the respective concentrations cannot be directly determined from the integration of the capacitance area.

3.2. Calibrations and validations of constructed PLS regression models based on chromatograms

The PLS regression models were constructed in the range spanning 10.0 – 18.0 min. Figure 3 shows the cumulative percentage variances of the calibration and validation models obtained from the constructed PLS regression model. The cumulative percentage variance at PC1 for the variation model implies that there are not enough values to predict cation concentrations. The factor contribution value from PC0 to PC1 was lower than its from PC1 to PC2 in the models. The chromatogram data set suggested highly collinear data. Uysal and Boyaci reported a qualitative approach for authenticity confirmation of the composition of liquid egg products [14]. Constructed PLS based on NIR spectra for drying sample, shows smaller variance value PC1 than PC2. Otsuka *et al*. also reported smaller variance PC1 than PC2 based on Powder X-ray diffractograms and their predictive quantifications [15]. Cumulative percentage values for the validation decreased in the range from PC2 to PC4. The optimum number of the principal component was, therefore, determined to be PC2. The principal component number for prediction should be selected based on the cumulative percentage variance in order to prevent overfitting risk. The decreases of cumulative percentage variance at PC3 and PC4 for the validation model were considered to be overfitting examples.

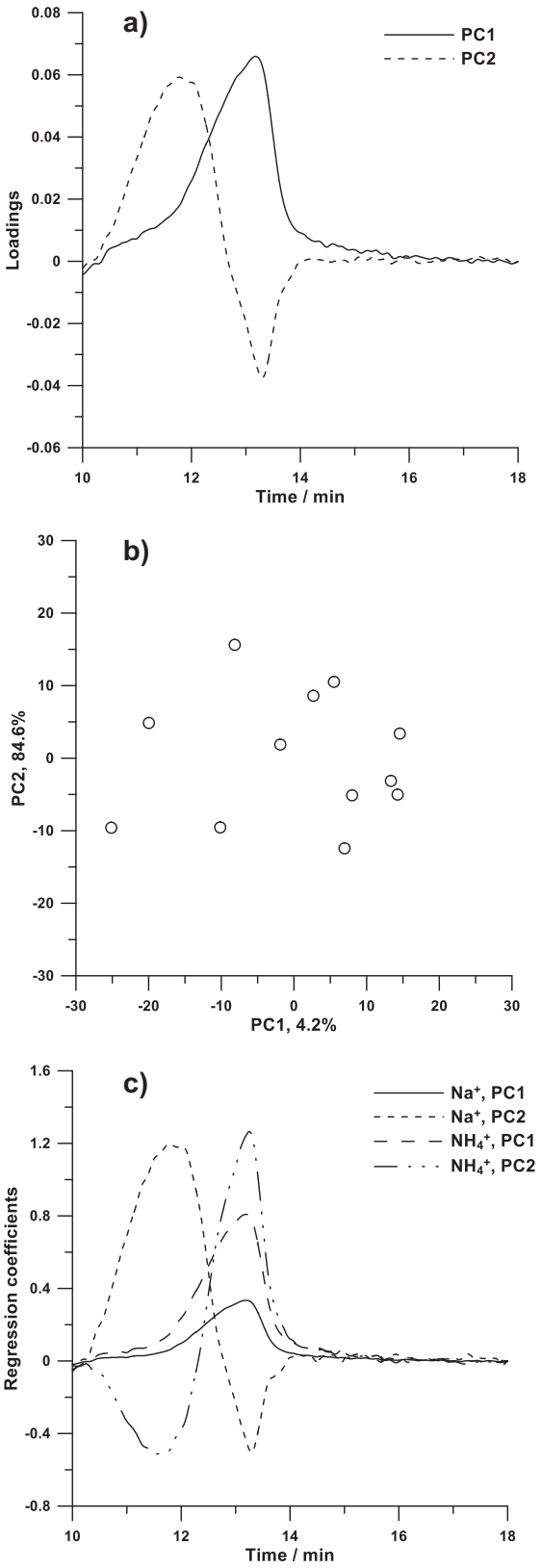


Fig. 4 Constructed PLS regression models at PC2. Loadings on chromatogram dataset (a), scores on chromatogram dataset (b), and regression vectors (c).

Table 2 Properties of constructed PLS regression models. Root-mean-squared error of prediction (RMSEP), standard error of prediction (SEP), tendency (*a*), intercept (*b*) and the correlation coefficient (*r*) of calibration and validation parameters for the prediction of $Na^+$ and $NH_4^+$ concentrations.

| | PC | RMSEP | SEP | *a* | *b* | *r* |
|---|---|---|---|---|---|---|
| $Na^+$ | 2 | 85.09 | 88.47 | 0.8877 | 20.43 | 0.892 |
| $NH_4^+$ | 2 | 64.90 | 67.51 | 0.8321 | 48.53 | 0.937 |

Table 2 lists the RMSEP, SEP, tendency, intercept and correlation coefficient value (*r*) at PC2 of the constructed PLS models. The RMSEP of the constructed PLS regression models for $Na^+$ and $NH_4^+$ at PC2 were respectively 85.09 and 64.90. The *r* values of the model for $Na^+$ and $NH_4^+$ at PC2 were 0.892 and 0.937, respectively. These values suggest that adequate predictive accuracy was obtained for the $Na^+$ and $NH_4^+$ concentrations in the validation data. Santos *et al*. reported the constructed PLS models based on Raman, mid-infrared, and near-infrared spectra to predict the density viscosity and total surface area [16]. Based on the NIR analysis, the *r* values of their constructed model for the prediction of the total surface area were lower than 0.8.

Figures 4(a–c) shows the loadings, scores, and regression vectors (RVs) of the constructed PLS models. In Fig. 4(a), the PC1 shows positive contributions on the peak obtained at 13.3 min. PC2 has a positive peak at 11.8 min and a negative peak at 13.3 min. PC1 loading shows $NH_4^+$ peak. PC2 of loading shows $Na^+$ and $NH_4^+$ peaks. This regression considered that quantitative prediction was performed as a composite function of waveforms. The loading similarities between PC1 and PC2 were calculated with the use of a least-squares method. The correlation value between PC1 and PC2 was 0.0179. In Fig. 4(b), the score plot from the PLS model was constructed with the use of the chromatograms of the calibration samples. The pure water chromatogram is plotted in the range of -25.1 to -9.59 on the score plot. The nearest sample in the center of the score plot was the sample of $Na^+$: $NH_4^+$ = 250 µM: 250 µM. In the PLS regression (Fig. 4(c)), the RV shows the correlation between the predicted data and the datasets, such as the chromatograms. It is known that the RV evaluation is one of the important factors for the construction of robust prediction models. The RVs of the constructed models show negative and positive peaks from 10.0 to 14.5 min. The peak time similarities between the RVs of the models and their chromatograms suggest that the model excluded the tails of the chromatograms.

Figures 5(a) and (b) show the predicted concentration data of the $Na^+$ and $NH_4^+$ from the constructed PLS models at PC2 (*n* = 3). The determination coefficient ($r^2$) for the $Na^+$ and $NH_4^+$ were respectively 0.9187 and 0.9314. The cation peaks have low resolutions in the chromatogram, as shown in Fig. 2. The peak area integral method cannot, therefore, estimate their concentrations. On the other hand, the PLS method has adequate accuracy in determining each cationic concentration.

In summary, we have constructed the PLS regression model for the simultaneous determination of $Na^+$ and $NH_4^+$ concentrations under undecomposed chromatogram peak conditions in the portable OTIC system. The model estimated the correlation between the chromatogram data set and the prediction of cation retention times and was successful in determining both cationic concentrations.

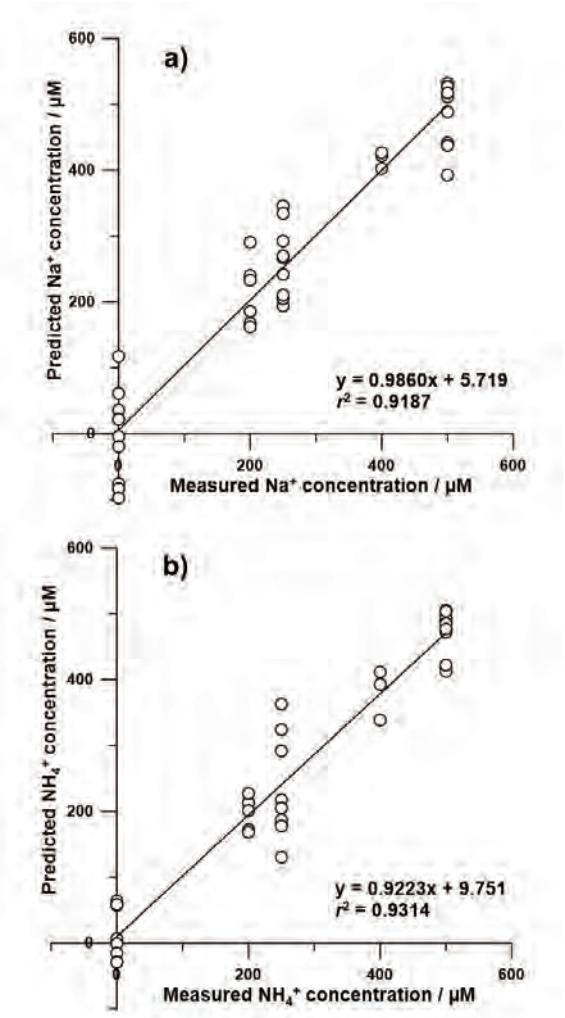


Fig. 5 Predicted concentrations of $Na^+$ (a) and $NH_4^+$ (b) from constructed PLS models at PC2.

## Acknowledgments

The authors would like to thank Prof. Satoru Goto, Faculty of Pharmaceutical Sciences, Tokyo University of Science. This study was partly supported by the research program for development of intelligent Tokushima Artificial Exosome (iTEX) from Tokushima University.

## References

[1] M. Maeder, *Anal. Chem.*, **1987**, *59*, 527.
[2] Y. Otsuka, M. Yamamoto, H. Tanaka, M. Otsuka, *Biomed. Mater. Eng.*, **2015**, *25*, 223.
[3] Y. Otsuka, M. Takeuchi, M. Otsuka, B. Ben-Nissan, D. Grossin, H. Tanaka, *Colloid Polym. Sci.*, **2015**, *293*, 2781.
[4] Y. Otsuka, A. Ito, S. Matsumura, M. Takeuchi, H. Tanaka, *Colloid Polym. Sci.*, **2015**, *293*, 3471.
[5] Y. Otsuka, A. Ito, M. Takeuchi, H. Tanaka, *J. Pharm. Sci.*, **2017**, *106*, 3458.
[6] A. Ito, Y. Otsuka, M., Takeuchi, H., Tanaka, *Colloid Polym. Sci.*, **2017**, *295*, 2011.
[7] P. Kubáň, P. Pelcová, V. Kubáň, L. Klakurková, P. K. Dasgupta, *J. Sep. Sci.*, **2008**, *31*, 2745.
[8] M. Takeuchi, Q. Li, B. Yang, P. K. Dasgupta, V. E. Wilde, *Talanta* **2008**, *76*, 617.

[9] H. Abdi, *WIREs Comp. Stat.* **2010**, *2*, 97.
[10] G. P. S. Smith, K. C. Gordon, S. E. Holroyd, *Vib. Spectrosc.*, **2013**, *67*, 87.
[11] D. M. Haaland, E. V. Thomas, *Anal. Chem.*, **1988**, *60*, 1193.
[12] M. Zhao, Y. Nian, P. Allen, G. Downey, J. P. Kerry, C. P. O'Donnell, *Data in Brief*, **2018**, *19*, 1355.
[13] K. Kachrimanis, D. E. Braun, U. J. Griesser, *J. Pharm. Biomed. Anal.*, **2007**, *43*, 407.
[14] R. S. Uysal, I. H. Boyaci, *J. Sci. Food Agric.,* **2020,** *100*, 855.
[15] Y. Otsuka, A. Ito, S. Matsumura, M. Takeuchi, S. Pal, H. Tanaka, *Chem. Pharm. Bull.*, **2016**, *64*, 1129.
[16] V. O. Santos, F. C. C. Oliveira, D. G. Lima, A. C. Petry, E. Garcia, P. A. Z. Suarez, J. C. Rubim, *Anal. Chim. Acta*, **2005**, *547*, 188.

(Received July 29, 2020)
(Accepted December 28, 2020)