C191-E042

SHIMADZU
Excellence in Science

Technical Report

# New Data Processing Method for Photodiode Array Detectors

## Principle and Overview of Intelligent Peak Deconvolution Analysis (i-PDeA II)

Toshinobu Yanagisawa[1]


## Abstract:

An i-PDeA II (Intelligent Peak Deconvolution Analysis II) data analysis technique was developed for extracting target peaks from unseparated peaks by analyzing photodiode array (PDA) detector data using the chemometrics multivariate curve resolution alternating least squares (MCR-ALS) technique. The i-PDeA II function can separate peaks for multiple components in absorption spectra and chromatograms by simply specifying the wavelength and time ranges. The i-PDeA II function can be used to identify spectra and quantitate peaks after separation of individual components, even for difficult-to-separate peaks for which a standard sample cannot be prepared. Furthermore, because i-PDeA II separates peaks based only on differences in spectral shape, it can also be used to separate and quantitate peaks for co-eluted isomers. This report explains the principle used by the i-PDeA II technique to separate peaks, describes an example of using i-PDeA II to analyze a sample with isomers of three components, and evaluates the spectral identification and quantitation performance.


**Keywords: Photodiode array detector, chemometrics, MCR-ALS, and LabSolutions**

## 1. Fundamental Theoretical Basis for Peak Deconvolution Algorithm

### 1-1. Modeling PDA Detector Data

Given peak profiles and spectra for each component in a three-component mixture, $c_1(t)$, $s_1(\lambda)$, $c_2(t)$, $s_2(\lambda)$, $c_3(t)$, and $s_3(\lambda)$, then measurement data in an ideal system $d(t,\lambda)$ can be described by the following expression.

$$d(t,\lambda) = c_1(t)s_1(\lambda) + c_2(t)s_2(\lambda) + c_3(t)s_3(\lambda)$$

Then spectra $d(t_i,\lambda)$ measured as a function of time $t_i$ can be expressed as follows:

$$d(t_i,\lambda) = c_1(t_i)s_1(\lambda) + c_2(t_i)s_2(\lambda) + c_3(t_i)s_3(\lambda)$$

Assuming spectral components are vectors with discrete values $\lambda_j$ (where $j$ = 1 to m), then spectra can be described as follows:

$$\boldsymbol{d}_i^T = \alpha_i\boldsymbol{s}_1^T + \beta_i\boldsymbol{s}_2^T + \gamma_i\boldsymbol{s}_3^T = \begin{pmatrix}\alpha_i & \beta_i & \gamma_i\end{pmatrix}\begin{pmatrix}\boldsymbol{s}_1^T\\ \boldsymbol{s}_2^T\\ \boldsymbol{s}_3^T\end{pmatrix}$$

where,

$$\boldsymbol{d}_i^T = \begin{pmatrix}d(t_i,\lambda_1) & \cdots & d(t_i,\lambda_m)\end{pmatrix}$$

$$\alpha_i = c_1(t_i),\ \beta_i = c_2(t_i),\ \gamma_i = c_3(t_i)$$

$$\boldsymbol{s}_1^T = \begin{pmatrix}s_1(\lambda_1) & \cdots & s_1(\lambda_m)\end{pmatrix},\ \boldsymbol{s}_2^T = \begin{pmatrix}s_2(\lambda_1) & \cdots & s_2(\lambda_m)\end{pmatrix},\ \boldsymbol{s}_3^T = \begin{pmatrix}s_3(\lambda_1) & \cdots & s_3(\lambda_m)\end{pmatrix}$$

By summarizing each spectrum measurement at time $t_i$ (where i = 1 to n), measurements can be expressed in matrix form, as follows:

$$\begin{pmatrix}\boldsymbol{d}_1^T\\ \vdots\\ \boldsymbol{d}_n^T\end{pmatrix} = \begin{pmatrix}\alpha_1 & \beta_1 & \gamma_1\\ \vdots & \vdots & \vdots\\ \alpha_n & \beta_n & \gamma_n\end{pmatrix}\begin{pmatrix}\boldsymbol{s}_1^T\\ \boldsymbol{s}_2^T\\ \boldsymbol{s}_3^T\end{pmatrix}$$

or by direct product (outer product), as follows:

$$D = \boldsymbol{c}_1\boldsymbol{s}_1^T + \boldsymbol{c}_2\boldsymbol{s}_2^T + \boldsymbol{c}_3\boldsymbol{s}_3^T \qquad \text{Eq. (1)}$$

or alternatively

$$D = CS^T \qquad \text{Eq. (2)}$$

where,

$$D = \begin{pmatrix}d(t_1,\lambda_1) & \cdots & d(t_1,\lambda_m)\\ \vdots & \cdots & \vdots\\ d(t_n,\lambda_1) & \cdots & d(t_n,\lambda_m)\end{pmatrix},\ \boldsymbol{c}_1 = \begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix},\ \boldsymbol{c}_2 = \begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix},\ \boldsymbol{c}_3 = \begin{pmatrix}\gamma_1\\ \vdots\\ \gamma_n\end{pmatrix}$$

$$C = \begin{pmatrix}\alpha_1 & \beta_1 & \gamma_1\\ \vdots & \vdots & \vdots\\ \alpha_n & \beta_n & \gamma_n\end{pmatrix},\quad S^T = \begin{pmatrix}s_1(\lambda_1) & \cdots & s_1(\lambda_m)\\ s_2(\lambda_1) & \cdots & s_2(\lambda_m)\\ s_3(\lambda_1) & \cdots & s_3(\lambda_m)\end{pmatrix}$$

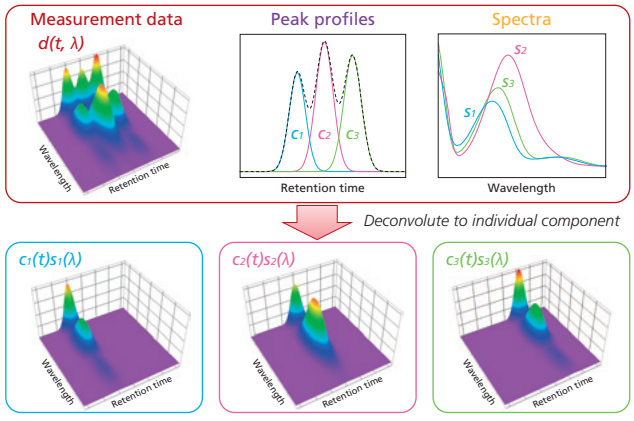



Fig. 1 Measurement Data from Three-Component Mixture Sample

The data can be expressed schematically as follows:

$$\begin{pmatrix}\text{Meaurement spectrum 1}\\ \vdots\\ \text{Meaurement spectrum n}\end{pmatrix} = \begin{pmatrix}\text{Component 1 profile} & \text{Component 2 profile} & \text{Component 3 profile}\end{pmatrix}\begin{pmatrix}\text{Component 1 pure spectrum}\\ \text{Component 2 pure spectrum}\\ \text{Component 3 pure spectrum}\end{pmatrix}$$

Considering measurement error, noise, and unpredictable factors, and given a remainder $R$, the measurement data can be modeled as follows:[1), 2)]

$$D = CS^T + R$$

This relational expression is valid for any number of components.

1 Analytical & Measuring Instruments Division

## 1-2. Solutions Using MCR-ALS Technique

The MCR-ALS technique estimates the peak profile or the spectrum with the closest fit to measurement data by repeatedly approximating *C* (peak profiles) or *S* (spectra) in equation (2) using least squares approximation. The following is the typical method for determining solutions by the MCR-ALS technique.[3), 4), 5), 6)]

Step 1 Specify the number of components in measurement data *D*.
Step 2 Calculate initial estimate (for example, by specifying the initial value for *C*).
Step 3 Using the estimate of *C*, calculate the $S^T$ matrix under appropriately chosen constraints.
Step 4 Using the estimate of $S^T$, calculate the *C* matrix under appropriately chosen constraints.
Step 5 From the product of *C* and $S^T$ found in the above steps of an iterative cycle, calculate an estimate of the original data matrix, *D*.
Step 6 Repeat steps 3, 4, and 5 until convergence is achieved.

Equation (2) generally does not give a unique solution. Therefore, to determine the optimal solution, constraints must be specified based on problem characteristics. Consequently, by specifying appropriate constraints, MCR-ALS can provide valid solutions even without prior information.

## 1-3. i-PDeA II Peak Separation Algorithm

If equation (1) is expanded for N components, the measurement signal *D* can be expressed by the following equation.

$$D = \boldsymbol{c}_1\boldsymbol{s}_1^T + \boldsymbol{c}_2\boldsymbol{s}_2^T + \cdots + \boldsymbol{c}_N\boldsymbol{s}_N^T$$

This algorithm determines a solution by minimizing the following squared errors, with the chromatogram vector $\boldsymbol{c}_k$ substituted by the chromatogram model function $\boldsymbol{f}_k$.[7)]

$$E = \left|D - \Sigma \boldsymbol{f}_k \boldsymbol{s}_k^T\right|^2 \qquad (k = 1, 2, \ldots, N)$$

In this case, a bidirectional exponentially modified Gaussian (BEMG) function is used as the chromatogram model function. BEMG is the reciprocal of the delay time component of the exponentially modified Gaussian (EMG) function, as defined by the following equations.

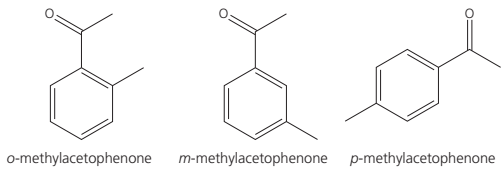
$$bemg(t,a,b) = \int_{-\infty}^{0} e^{ax} \cdot emg(t-x,b)dx$$

$$emg(t,b) = \int_{0}^{\infty} e^{-bx} \cdot exp(-(t-x)^2)dx$$

This algorithm applies the MCR-ALS technique by using an estimated value as the initial value and the BEMG model function as the chromatogram constraint. Since the number of components after separation is unknown, the initial condition starts with a single component and then successively adds components as the presence of unseparated peaks are determined in the residual signal to determine the optimal solution.

## 2. Example of Using the Algorithm for a Three-Component Mixture Sample

The following describes an example of using the algorithm for a mixture of the positional isomers *o*-methyl acetophenone (*o*-MAP), *m*-methyl acetophenone (*m*-MAP), and *p*-methyl acetophenone (*p*-MAP), shown in Fig. 2.




*o*-methylacetophenone *m*-methylacetophenone *p*-methylacetophenone

Fig. 2 Structure of Target Substances

400 µg/mL concentration standard samples of the pure isomers were prepared by dissolving pure *o*-MAP, *m*-MAP, and *p*-MAP in 30 vol% methanol-water solution and then data was acquired using the following analytical conditions.

| Analytical Conditions | |
|---|---|
| HPLC System | : Shimadzu LC-2030C 3D |
| Mobile Phase | : Methanol 30 %/water 70 % |
| Column Type | : Shimadzu Shim-pack XR-ODS III C18 (3.0 × 50 mm, 2.2 µm) |
| Mobile Phase Flowrate | : 1.0 mL/min |
| Oven Temperature | : 40 °C |
| Sampling | : 240 msec |
| Slit Width | : 1.2 nm |
| Time Constant | : 480 msec |
| Sampling Wavelength Range | : 190 nm to 400 nm |
| Sample Injection Volume | : 1.5 µL |

The methylacetophenone isomers eluted in the order *o*-MAP, *p*-MAP, and *m*-MAP respectively, where the similarity between respective components in spectra obtained from peak tops for each isomer in measurement data was 0.8410 for *o*-MAP/*p*-MAP, 0.9123 for *p*-MAP/*m*-MAP, and 0.9809 for *o*-MAP/*m*-MAP (Fig. 3).

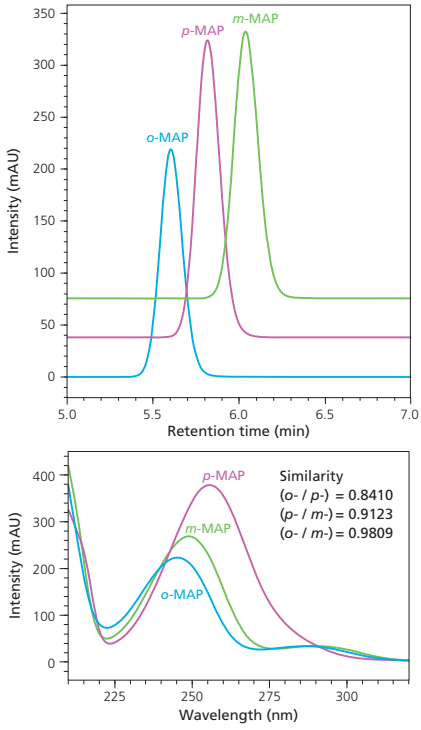



Fig. 3 Measurement Results for Standard Samples of *o*-MAP, *m*-MAP, and *p*-MAP

A mixture solution prepared by mixing *o*-MAP, *m*-MAP, and *p*-MAP standard samples to 400 µg/mL, 400 µg/mL, and 4 µg/mL concentrations, respectively, and then data was acquired using the same analytical conditions. Given the relative concentrations in the order of peak elution (*o*-MAP/*p*-MAP/*m*-MAP =100/1/100), the peak for *p*-MAP (relative concentration of 1) was obscured by the peaks for *o*-MAP and *m*-MAP (relative concentration of 100), which eluted before and after the *p*-MAP peak. Consequently, the presence of *p*-MAP could not be confirmed visually. However, when i-PDeA II was used for measurement data from the time range from 5.0 to 7.0 minutes and wavelength range from 210 to 320 nm, *o*-MAP, *m*-MAP, and *p*-MAP could be separated into independent peaks, as shown in Fig. 4.

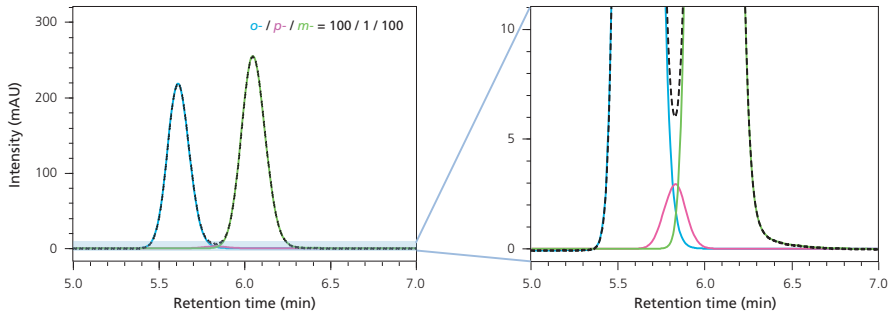



Fig. 4 Separation Results for Mixture Sample of *o*-MAP, *m*-MAP, and *p*-MAP

# 3. Using i-PDeA II for Spectral Analysis and Quantitative Analysis

## 3-1. Using i-PDeA II to Measure Purity

When i-PDeA II was applied to measurement data to confirm the purity of respective standard samples for *o*-MAP, *m*-MAP, and *p*-MAP, an impurity was detected in the *p*-MAP standard sample (Fig. 5).

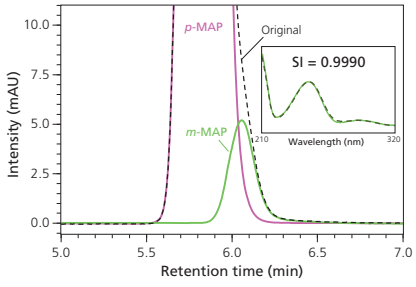



Fig. 5 Impurity Contained in *p*-MAP Standard Sample

Based on the elution time and spectral similarity, the impurity is presumably *m*-MAP.

Therefore, the *p*-MAP standard sample was measured using a Shimadzu Shim-pack XR-Phenyl reversed phase ultra fast analysis column (3.0 × 75 mm, 2.2 µm packing) to separate the impurity (Fig. 6).

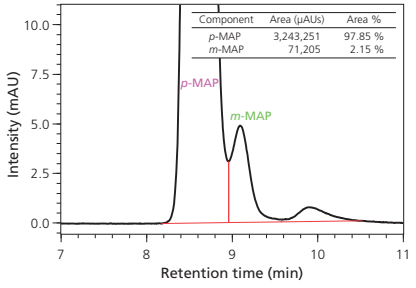



Fig. 6 Measurement of *m*-MAP Content in *p*-MAP Standard Sample

As a result, the *m*-MAP peak was separated and it was confirmed that with the area ratio in results averaged from three measurements the standard sample contained 2.15 % *m*-MAP.

## 3-2. Using i-PDeA II for Quantitative Analysis

To evaluate how well the algorithm performs, 400 µg/mL, 400 µg/mL, and 4 µg/mL standard samples of *o*-MAP, *m*-MAP, and *p*-MAP, respectively, were measured individually using the same analytical conditions as used to analyze the three-component mixture sample. Then the area values from measurement results were compared to the area values of separated peaks. In addition to identifying the separated peaks using the spectra in Fig. 3, their similarity was calculated as well.[8)]

A comparison of area values and spectral similarity from averaged results for three analyses is shown in Table 1. (The true *p*-MAP and *m*-MAP values were calculated by correcting the area value measured from the *p*-MAP standard sample to compensate for the 2.15 % *m*-MAP content.)

Table 1 Evaluation of i-PDeA II Performance for Quantitative Analysis

| Component | Area (µAUs) | | Error % | Similarity |
|---|---|---|---|---|
| | True Value | Mixture Sample (Deconvoluted) | | |
| *o*-MAP | 2,090,806 | 2,080,405 | -0.50 % | 1.0000 |
| *p*-MAP | 27,666 | 26,639 | -3.71 % | 0.9996 |
| *m*-MAP | 2,658,837 | 2,656,836 | -0.08 % | 1.0000 |

Fig. 7 shows the normalized *p*-MAP spectrum from the standard sample compared to the spectrum estimated based on the separated peak from the mixture.

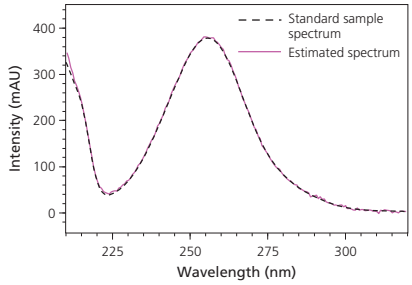



Fig. 7 Shape Comparison of *p*-MAP Spectra

In the case of the three-component mixture sample with relative *o*-MAP, *p*-MAP, and *m*-MAP concentrations of 100, 1, and 100, respectively, there was less than ±1 % error and over 0.9999 similarity between the area values of separated peaks and the corresponding peaks measured from standard samples with relative *o*-MAP and *m*-MAP concentrations of 100 and less than ±4 % error and over 0.9996 similarity between the area values of peaks for the relative *p*-MAP concentration of 1.

## 4. Data Analysis Using LabSolutions

The i-PDeA II peak separation algorithm is included in LabSolutions data analysis functionality. Data for separated peaks can be displayed as chromatograms for individual peaks, and also as separated spectra, in the LabSolutions PDA data analysis window (Fig. 8).

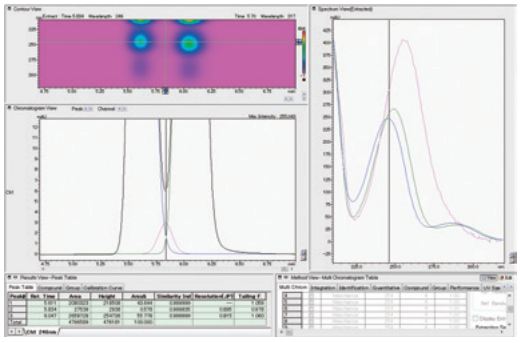


Fig. 8 PDA Data Analysis Window in LabSolutions

The window for i-PDeA II settings is shown in Fig. 9.

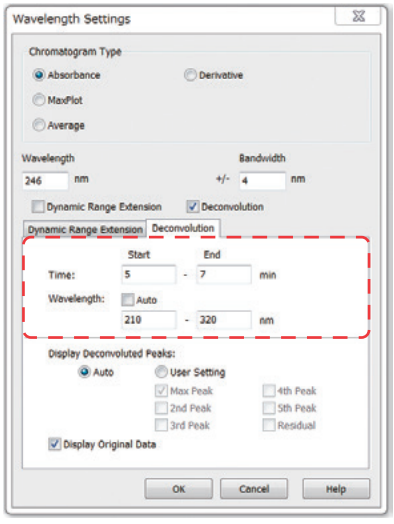



Fig. 9 Window for i-PDeA II Settings

Peaks can be separated using the i-PDeA II function by simply specifying the wavelength and time ranges.

By using the data analysis functionality in LabSolutions, the entire process of separating peaks, integrating the areas under separated peaks, and calculating quantitative values can be performed seamlessly without any data conversion and spectra can be identified and libraries searched based on peak-top spectra.

## 5. Conclusion

A new analytical technique was developed for separating peaks that is based on improved basic photodiode array detector performance, superior HPLC/UHPLC system reproducibility, and incorporation of chemometrics technology, in addition to column technologies even for difficult-to-separate-peaks.

The i-PDeA II function can help analyze samples more quickly and improve laboratory productivity. Key points are summarized below.

- An algorithm for separating co-eluted peaks was developed by applying the MCR-ALS technique from chemometrics to photodiode array detector data.
- Fast and accurate quantitative analysis is possible even if components are not fully separated in the column.
- i-PDeA II can even be used to analyze isomers with identical molecular weights.
- Spectral data can be analyzed even after peak separation.

i-PDeA II provides a unique solution for peak separation or quantitative analysis of isomer samples that was not possible with previously available techniques. These features can be expected to provide further improvements in analytical efficiency and data reliability.

## Acknowledgments

i-PDeA II was developed based on results obtained from joint development work with Eisai Co., Ltd. We are especially grateful to Takashi Kato, Kanta Horie, Shuntaro Arase, Hideki Kumobayashi, and the many others involved for their generous cooperation during development. In particular, we are grateful to Naoki Asakawa for providing the development opportunity and generously sharing his extensive knowledge and valuable suggestions during routine discussions.

### References

1) Takeshi Hasegawa, *Quantitative Spectral Analysis,* Kodansha (2005)
2) Takeshi Hasegawa, *Bunseki* 2014(9), pp. 460-467, Japan Society for Analytical Chemistry (2014)
3) Gemperline, P. (Ed.), *Practical guide to chemometrics 2nd Ed.,* CRC Press (2006)
4) R. Tauler, D. Barceló, Multivariate curve resolution applied to liquid chromatography - diode array detection, *TrAC Trends Anal. Chem.* 12 (1993) 319-327
5) R. Tauler, Multivariate curve resolution applied to second order data, *Chemometr. Intell. Lab.* 30 (1995) 133-146
6) H. Parastar, R. Tauler, Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements:A New Insight to Address Current Chromatographic Challenges, *Anal. Chem.* 86 (2014) 286-297
7) S. Arase et al., Intelligent peak deconvolution through in-depth study of the data matrix from liquid chromatography coupled with a photo-diode array detector applied to pharmaceutical analysis, *J. Chromatogr. A* 1469 (2016) 35–47
8) I. Sakuma et al., Resolution of unresolved peaks containing unknown components by high-performance liquid chromatography with multi-wavelength detection., *J. Chromatogr. A* 506 (1990) 223-243

First Edition: May, 2017

Shimadzu Corporation
www.shimadzu.com/an/


Printed in Japan