การพัฒนาโปรแกรมการแยกชัดข้อมูลหลายตัวแปรโดยเทคนิคการจำลองตัวเอง สำหรับการประยุกต์ทางเคมี

นายคเณศ วงษ์ระวี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเคมี ภาควิชาเคมี
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2548

ISBN 974-17-4805-1

DEVELOPMENT OF MULTIVARIATE CURVE RESOLUTION PROGRAM BASED ON SELF-MODELING TECHNIQUES FOR CHEMICAL APPLICATIONS

Mr. Kanet Wongravee

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Chemistry
Department of Chemistry
Faculty of Science
Chulalongkorn University
Academic year 2005

ISBN 974-17-4805-1

Thesis Title	DEVELOPMENT OF MULTIVARIATE CURVE RESOLUTION PROGRAM BASED ON SELF-MODELING TECHNIQUES FOR CHEMICAL APPLICATIONS
By	Mr. Kanet Wongravee
Field of Study	Chemistry
Thesis Advisor	Associate Professor Vudhichai Parasuk, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree.

T. Vitidsant

..... Deputy Dean for Administrative Affairs,
Acting Dean, The Faculty of Science

(Associate Professor Tharapong Vitidsant, Ph.D.)

THESIS COMMITTEE

Sirirat Kokpol

..... Chairman

(Associated Professor Sirirat Kokpol, Ph.D.)

Vudhichai Parasuk

..... Thesis Advisor

(Associate Professor Vudhichai Parasuk, Ph.D.)

..... Member

(Associate Professor Sanong Ekasit, Ph.D.)

Amarawan Intasiri

..... Member

(Amarawan Intasiri, Ph.D.)

คเณศ วงษ์ระวี : การพัฒนาโปรแกรมการแยกชัดข้อมูลหลายตัวแปรโดยเทคนิคการจำลองตัวเอง สำหรับการประยุกต์ทางเคมี. (DEVELOPMENT OF MULTIVARIATE CURVE RESOLUTION PROGRAM BASED ON SELF-MODELING TECHNIQUES FOR CHEMICAL APPLICATIONS) อ.ที่ปรึกษา : รศ.ดร. วุฒิชัย พาราสุข , 90 หน้า. ISBN 974-17-4805-1.

ได้นำเทคนิคสำหรับการหาองค์ประกอบและปริมาณสารในระบบที่มีหลายองค์ประกอบมาใช้ โดยพัฒนาโปรแกรม SMCR รุ่น 1.0 ขึ้น เพื่อใช้ในการแยกสเปกตรัมรังสียูวีวิสิเบิลและโครมาโตแกรม โปรแกรมนี้แยกสเปกตรัมให้เป็นข้อมูลทางความเข้มข้นและค่าคงที่การดูดกลืนแสงขององค์ประกอบที่สำคัญ โดยใช้วิธีออร์โตโกนอลโปรเจกชัน (OPA), การวิเคราะห์โดยการจำลองตัวเองแบบอินเตอร์แรกตีฟ (SIMPLISMA) และการวิเคราะห์ตัวประกอบ (EFA) ซึ่งเป็นส่วนหนึ่งของเทคนิคการจำลองตัวเองในการแยกชัดข้อมูล โปรแกรมที่พัฒนาขึ้นนี้เขียนอยู่ในภาษาของโปรแกรม MATLAB v6.5 ของบริษัท Math Works ได้ทำการประเมินประสิทธิภาพและพิสูจน์ความถูกต้องของโปรแกรมด้วยสเปกตรัมที่จำลองขึ้นของการแตกตัวของกรด การเกิดสารประกอบเชิงซ้อนโลหะ-ลิแกนด์ และระบบโครมาโทกราฟี ข้อมูลทางความเข้มข้นและค่าคงที่การดูดกลืนแสงที่ได้จากการแยกชัดด้วยวิธีการทั้งสามวิธีสอดคล้องกับค่าที่ตั้งไว้เป็นอย่างดี โดยวิธีการ EFA เป็นวิธีการที่ให้ผลดีที่สุด โปรแกรมนี้ยังใช้ในการวิเคราะห์สเปกตรัมจากการทดลองปฏิกิริยา 2 ขั้นตอนของสาร 3-คลอโรฟีนิลไฮดราโซนโพรเพน กับ 2-เมอริคอซิโตเอทานอล และสเปกตรัมจากการทดลองปฏิกิริยาการเกิดสารประกอบเชิงซ้อนคอปเปอร์-ไกลซีน และคอปเปอร์-อะลานีน ข้อมูลทางความเข้มข้นของระบบเหล่านี้ให้ผลสอดคล้องกับการศึกษาก่อนหน้านี้เป็นอย่างดี

ภาควิชา เคมี ลายมือชื่อนิสิต .....
สาขาวิชา เคมี ลายมือชื่ออาจารย์ที่ปรึกษา .....
ปีการศึกษา 2548

4672216823 : MAJOR CHEMISTRY
KEY WORD: CHEMOMETRICS, SELF-MODELING CURVE RESOLUTION, OPA, SIMPLISMA, EFA

KANET WONGRAVEE: DEVELOPMENT OF MULTIVARIATE CURVE RESOLUTION PROGRAM BASED ON SELF-MODELING TECHNIQUES FOR CHEMICAL APPLICATIONS. THESIS ADVISOR: ASSOC. PROF. VUDHICHAI PARASUK, Ph. D. 90 pp., ISBN 974-17-4805-1

A technique for characterization of multi-component system has been implemented. The program SMCR version 1.0 was developed to resolve the UV-visible spectra and chromatogram. The program resolves spectra into concentration and absorptivity profiles of significant components using self-modeling curve resolution techniques based on Orthogonal Projection Approach (OPA), Simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) and Evolving Factor Analysis (EFA) methods. The program was developed in MATLAB version 6.5 (MathWorks, Inc.). The efficiency and validation of the program were performed with simulated spectra in acid-dissociation, metal-ligand complexation and chromatographic systems. The resolved concentration and absorptivity profiles from three methods are in good agreement with the presetting values. The EFA method gives the best results of three methods. The program was later used to resolve the experimental spectra of two-step reaction between 3-chlorophenyl-hydarzonopropane dinitril and 2-meracaptoethanol and copper-glycine and copper-alanine complexation systems. The concentration profiles of these systems are in good agreement with previous studies.

Department Chemistry Student's signature Kanet Wongravee
Field of study Chemistry Advisor's signature Vudhichai Parasuk
Academic year 2005

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to a number of people who give me the guidance, help and support to accomplish my study. First of all, I would like to give all gratitude affectionately to my parents for all support and encouragement during the whole period of my study.

I also would like to express my sincere gratitude to Associate Professor Dr. Vudhichai Parasuk, my advisor, for his multifarious understanding, useful guidance, kind suggestion and encouragement throughout my study, especially patience during the time of this thesis preparation. Grateful thanks to Dr. Arunchai Tangcharoenbumrungsuk, for good advice. Again, I am express special thanks to Associate Professor Dr. Sirirat Kokpol, Associate Professor Dr. Sanong Ekgasit and Dr. Amarawan Intasiri who are the thesis committee.

Finally, I would like to acknowledge the Computational Chemistry Unit Cell (CCUC) and Computer Center for Advance Research at Faculty of Science, Chulalongkorn University for computer resources and other facilities. I also would like acknowledge the financial support by the development and promotion of science and technology talents project of Thailand (DPST) during my graduate studies at Chulalongkorn University.

CONTENTS

Page

Page

LIST OF FIGURES

Figure **Page**

Figure **Page**

LIST OF TABLES

Table **Page**

LIST OF ABBREVIATIONS

SMCR	Self-modeling curve resolution
GUI	Graphical user interface
EFA	Evolving factor analysis
OPA	Orthogonal projection approach
SIMPLISMA	Simple-to-use interactive self-modeling mixture analysis
WFA	Window factor analysis
SFA	Subwindow factor analysis
ALS	Alternating least squares
SVD	Singular value decomposition
rms	Root mean square error
nrms	Normalized root mean square error
nnls	Non-negative least square
nm	Nanometer
GlyH	Glycine
AlaH	Alanine
α	Mole ratio of each component
β	Cumulative stability constant

CHAPTER I

INTRODUCTION

The field of chemistry is currently facing major changes. In recent year, computer power has increased dramatically. This development, together with other factors, provides a new opportunity but also challenge to chemists in research and development. The research and development of the complex chemical system with the application of statistical and mathematical techniques have been confined mainly to analytical studies as *Modern analytical chemistry*. This modern analytical chemistry has long been recognized mainly as a measurement science. In its development, new methodologies developed in mathematical, computer, and biological sciences as well as other fields are also employed to provide in-depth and broad-range analysis.

Previously, the main problem confronting analytical scientists was how to obtain data. At the time, measurements were labor-intensive, tedious, time-consuming, and expensive, with low-sensitivity, and manual recording. There were also problems of preparing materials, lack of proper techniques, as well as inefficient equipment and technical support. Chemists also had to extract as much information as possible about the structure, composition, and other properties of the system under investigation, which was an insurmountable task in many cases. After an analytical measurement, the data collected are often treated by different signal processing techniques. The aim is to obtain higher quality or “true” data and to extract maximum amount of meaningful information, although this is not easy to accomplish. For instance, in spectroscopic and chromatographic study, two experimental data were carried out on the same sample mixture. The two peaks acquired usually differed from each other to a certain extent because of the variations in instrumentation, experimental conditions, and other factors. In this condition, if the pure peak areas or peak positions are available, the concentrations of these compounds can also be determined with high level of accuracy. If the overlapping peaks that arise from different component mixtures, the determined concentrations of these components will be erroneous. Statistical methods can also be applied to help evaluating the results and to calculate the level of confidence or

identifying concentrations of the components. These data manipulation are very important in preparing a reliable report for an analytical test. Data treatment and data interpretation on, for instance, the spectroscopic and chromatographic studies is a part of an interdisciplinary known as *chemometrics*.

The term chemometrics was introduced by Svate Wold and Bruce R. Kowalski in the early 1970s. Since then, Chemometrics is one of those subjects that use mathematical and statistical methods to handle, interpret, and predict chemical data. Chemometrics has been developing and is now widely applied to different fields of chemistry, especially modern analytical chemistry. The powerful methodologies have opened new vision for chemists and provided useful solutions for many chemical problems. Self-modeling curve resolution (SMCR) [1] has proved to be one of the most potent techniques in the chemometric world. There are more than 20 different self-modeling curve resolution methods or variations [2] which have been reported in the chemical literature. Nearly all of these techniques were originally developed and applied to estimate pure spectra and concentration profiles form mixture spectra.

1.1 Definition of Self-Modeling Curve Resolution (SMCR)

The history of SMCR techniques stretches back to 1960. The starting point of SMCR lay in the recognition of the fact that if each component in a mixture has a different spectrum and corresponding concentration profile. The matrix rank of data matrix has a very good range of only one-component correspondence to the number of chemical components in the system. The SMCR technique consists of a family of chemometric methods that utilize a certain mathematical decomposition to resolve the two-way signals from instrumentally unresolved multi-component mixtures into factors for single species. As the terminology “self-modeling” implied, the SMCR, in principle, does not require a priori of any specific information concerning the data to resolve the pure variables. The only premises are a certain bilinear model for the data and some generic knowledge about the pure variable, such as non-negativity, unimodality and closure as natural constraint. In common practice, these premises are naturally satisfied for two-way data obtained from multivariate measurements on mixtures with varying

compositions. The SMCR provides a useful tool for exploring multi-component phenomena in complex chemical systems.

1.2 Principles of SMCR

The SMCR methods basically treat the spectra as a data matrix (**D**) consisting of *r* rows of wavelength (nm) or spectral channels and *c* column of samples.

$$\mathbf{D} = \begin{array}{c} \\ \text{row designee} \end{array} \begin{array}{c} \text{column designee} \\ \begin{bmatrix} d_{11} & d_{1,2} & \dots & d_{1,nc} \\ d_{21} & d_{2,2} & \dots & d_{2,nc} \\ \vdots & \vdots & & \vdots \\ d_{nr,1} & d_{nr,2} & \dots & d_{nr,nc} \end{bmatrix} \end{array} \qquad (1.1)$$

The row and column headings of the matrix are called designee. The symbol d_{ij} represents the data point associated with the i^{th} row and j^{th} column of the matrix.

In this research, the absorbance data obeys the Beer's law; therefore, the factor can be interpreted chemically. The two-way data matrix obtained from multivariate spectrometric or chromatographic measurements on a set of mixtures of varying compositions can be represented by the bilinear model. It is assumed that each point in the data matrix must be a linear sum of product terms. The number of terms in the sum, *n*, is called the absorbing components. The d_{ij} can be expressed as:

$$d_{ij} = \sum_{k=1}^{n} \varepsilon_{ik} \, c_{kj} + e_{ij} \qquad (1.2)$$

where ε_{ik} and c_{kj} are called row factors as absorptivity and column factors as concentration of each component, e_{ij} is called noise of data matrix respectively.

For bilinear data modeled by equation 1.2, the noises of data matrix are very low compared with significant factors. The noises of data matrix can be eliminated by using static methods. The data matrix can be decomposed into two matrices.

$$\mathbf{D} = \mathbf{A}_{(\text{absorptivity})} \cdot \mathbf{C}_{(\text{concentration})} \quad (1.3)$$

where

$$\mathbf{A} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{1,2} & \cdots & \varepsilon_{1,n} \\ \varepsilon_{21} & \varepsilon_{2,2} & \cdots & \varepsilon_{2,n} \\ \vdots & \vdots & & \vdots \\ \varepsilon_{nr,1} & \varepsilon_{nr,2} & \cdots & \varepsilon_{nr,n} \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} c_{11} & c_{1,2} & \cdots & c_{1,c} \\ c_{21} & c_{2,2} & \cdots & c_{2,c} \\ \vdots & \vdots & & \vdots \\ c_{n,nc} & c_{n,nc} & \cdots & c_{n,nc} \end{bmatrix}$$

In this case, the spectral measurements all have non-negative values. It is known that if the norm of **A** and **C** cannot be uniquely determined, then one can prescribe a certain scale constraint for the pure spectral variables or the pure concentration variables.

The basic principle of SMCR is to seek a bilinear model that gives the best fit, in the sense of least squares or weighted least squares, to the two-way data **D**. In other words, SMCR estimates pure variables, **C** and **A** that minimize the following error criterion :

$$\mathbf{E} = \| \mathbf{D} - \mathbf{AC} \|^2 \quad (1.4)$$

The most commonly used error criterion is the squared difference between **D** and **AC**, though some SMCR methods use weighted [3] or normalized squared errors [4]. One notices that minimization of equation (1.4) over **A** and **C** can not guarantee a unique solution to the pure variables. Fortunately, in chemical practice some generic knowledge concerning pure variables is available and the evolving behavior of pure variables can be effectively exploited via local rank analysis, which, optimistically, may confine the feasible solution to a desirably small region. Under such circumstances, SMCR techniques are expected to generate solutions well qualified for practical use.

1.3 Objective of this research

The objective of this research is to develop the SMCR program based on SMCR techniques for resolving and applying it to solve the real data to obtain the concentrations and pure absorptivity of components in multicomponent mixtures.

1.4 Scope of this research

In this research, the program, namely SMCR, version 1.0 was developed. The program was written in MATLAB language using MATLAB complier program version 6.5 in personal computer. The efficiency and validation of the program were tested by resolving the UV/VIS absorption spectra for the acid-base equilibria and also the formation of metal-ligand complex formation. After validation, the program was used in the chemical experiments such as UV/VIS spectra of metal-ligand complexation system. In addition, the program was also applied to chromatography data to show the generality of the technique and our developed program. The program was designed to be interactive, easy-to-use, and user-friendly. Thus, Graphic user interface (GUI) was implemented and user manual is provided.

CHAPTER II

THEORETICAL CONSIDERATION

This chapter mentions methodologies of the self-modeling curve resolutions (SMCR) techniques with focusing on spectroscopic and chromatographic problems.

2.1 Outline of Self-Modeling Curve resolution

In terms of the uniqueness of the solution, two-way resolution techniques for SMCR can be classified into two groups; unique resolution methodologies and rational resolution methodologies. Evolving factor analysis (EFA) [5], Window factor analysis (WFA) [6], Subwindow factor analysis (SFA) [7] belong to the first group. Whereas Orthogonal projection analysis (OPA) [8], Simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) [9], Alternating least squares (ALS) [10] belong to the second group. The unique resolution methodologies try to find a unique resolution in which the factors for single species are uniquely defined according to the mathematical principles involved. If the mathematical principles are compatible with the chemical or physical model for the data, the unique solution is generally consistent with the true profiles. A characteristic feature of the unique resolution techniques is the exploitation of information in local feature regions such as selective regions or zero-concentration regions. These feature regions can generally be identified with the aid of the local rank analysis. In principle, the motivation of search for feature regions is always a certain theorem concerning the resolution uniqueness. The drawback of the unique resolution techniques is the difficulty to find the feature regions which can provided enough accuracy with local rank analysis and the solutions obtained turn out to be, to some degree, dependent upon the experience of the data analysts.

The rational resolution methodologies aim at finding a rational resolution in which the factors for single species do not violate the generic prior knowledge such as non-negativity, unimodality and so on. Rational resolution may produce a set of feasible solutions and the accuracy of the solutions depends on the correlation or colinearity among the pure profiles underlying the two-way data. Nevertheless, the user-friendly

implementation of the rational resolution methods has made it a common practice in real-world chemical applications. Actually, in situations where the correlation among the pure profiles is not very severe, rational resolution is expected to yield solutions which can approximate the true profiles as well. The above approaches for rational resolution can be distinguished from each other either in the way to obtain the starting estimate of pure variables or in the optimization algorithm to iteratively improve the estimate. A good starting estimate can give a refined resolution while an efficient optimization algorithm may show fast convergence. The methods described above are grouped as given in Figure 2.1

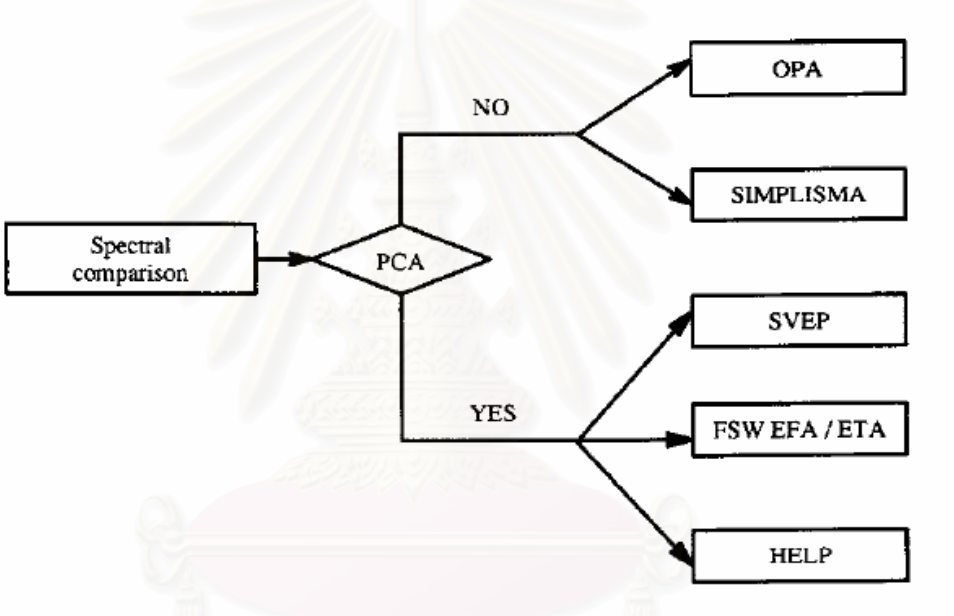


Figure 2.1 SMCR techniques for the determination of the number of compounds presented in the multicomponent system : Orthogonal Projection Approach (OPA), SIMPLe-to-use Interactive Self-modeling Mixture Analysis approach (SIMPLISMA), Singular Value Evolving Profile (SVEP), Fixed Size Window Evolving Factor Analysis (FSW EFA), Eigenstructure Tracking Analysis (ETA), Heuristic Evolving Latent Projection approach (HELP), Evolving Principled components innovation Analysis (EPCIA), and Iterative Target Transformation Factor Analysis Approach (ITTFA).

2.2 Self-modeling curve resolution programming

Self-modeling curve resolution (SMCR) techniques involves four main steps; preparation, selectivity, optimization and prediction. Figure 2.2 shows the flow chart of these steps. Details of each step are provided in the subsection.

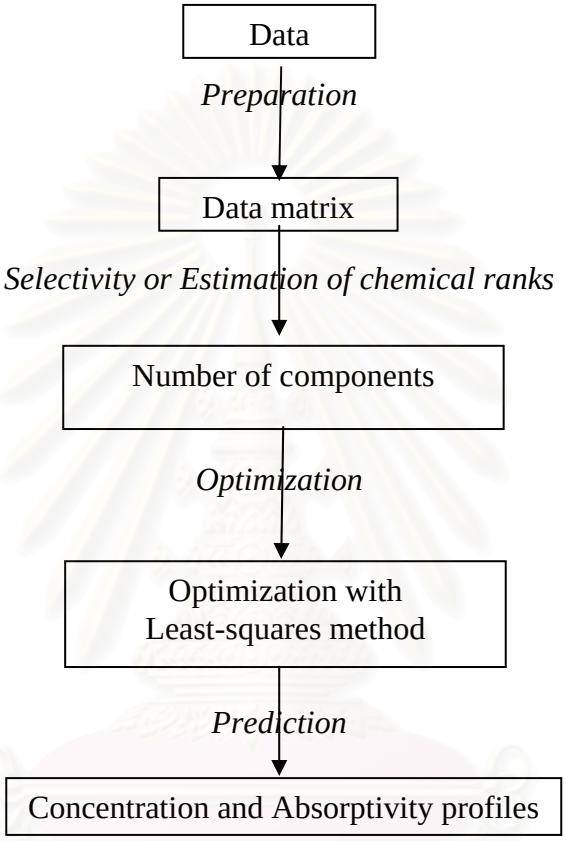


Figure 2.2 Block diagram of the main steps in the SMCR technique.

2.2.1 Preparation

The objective of the preparation step is to obtain a data matrix into which yields best situation matrix for self-modeling curve resolution analysis. The kinds of information sought from a data analysis should carefully be designed before applying self-modeling curve resolution analysis. The design of data matrix is the necessary step because it makes experimental data easily or suitable to analyte with the program. The design matrix may consist of a series of experiments performed under different conditions, e.g. a reaction at differing pHs, temperatures, and concentrations.

A data matrix **D** with the rank **(*nr x nc*)** where the ***nr*** refers to number of row elements, and ***nc*** refers to number of column element. This matrix could be represent as

$$\mathbf{D} = \begin{array}{c} \\ \text{row designee} \end{array} \begin{array}{c} \text{column designee} \\ \begin{bmatrix} d_{11}\ d_{1,2}\ \ldots & d_{1,nc} \\ d_{21}\ d_{2,2}\ \ldots & d_{2,nc} \\ \vdots\ \vdots & \vdots \\ d_{nr,1}\ d_{nr,2}\ \ldots & d_{nr,nc} \end{bmatrix} \end{array} \qquad (2.1)$$

Where d_{ij} is the element of data matrix. In many cases, the d_{ij} is spectroscopic signal such as absorbance where row designee are different wavelength and column designees are other variables such as pH, time, etc. To be able to resolve the data matrix, it must possess the bilinear property, which means that it can be decomposed into the product of the concentration profiles (**C**) and pure component spectra (**A**). The two-way data matrix obtained from multivariate spectroscopic measurements on a set of mixtures of compositions can be represented by the bilinear model and thus can be resolved.

2.2.2 Estimation of chemical rank

Estimation of chemical rank is a key term in quantitative analysis. The measured intensity or level of a selective variable/measurement is due to one single analytical variable. The variable represents the reliable information on the analyze. In traditional chemistry, multicomponent samples are often purified chemically prior to further quantitative or qualitative analysis and selective measurements that provide reliable results for each analyze would then be obtained. Here, we attempt to show how selective information of an analyze can be recognized without the need of purification in certain instrument profiles using the following methods:

- Orthogonal Projection Approach (OPA) [11]
- SIMPLe-to-use Interactive Self-modeling Mixture Analysis approach (SIMPLISMA) [11]
- Evolving Factor Analysis (EFA) [12]

2.2.2.1 Orthogonal Projection Approach (OPA)

OPA is based on Gram-Schmidt orthogonalisation, and on assumption that the purest spectra in the data matrix are commonly more dissimilar than the corresponding mixture spectra. The objective of OPA method is to determine the number of components present in the mixture and their corresponding spectra. The dissimilarity, $\mathbf{X_{ij}}$, is defined as the determinant of the dispersion matrix $\mathbf{Y_j}$. Each $\mathbf{Y_j}$ contains spectrum d_j plus reference spectra. Initially, the mean spectrum of $\mathbf{D}$ is considered as the reference spectrum in $\mathbf{Y_j}$ referred to as d_{ref}. Therefore, initially $\mathbf{Y_j}$ is a matrix (*2 x n*) containing d_{ref} and d_j.

$$\text{Mean spectrum } (d_{ref}) = \frac{\sum_{j=1}^{nc} d_{ij}}{nc} \qquad i = 1,\ldots,nr \qquad (2.2)$$

$$\text{Dispersion matrix } (\mathbf{Y_j}) = [\, d_{ref} \;\; d_j \,] \qquad j = 1,\ldots,nc \qquad (2.3)$$

The dissimilarity, $\mathbf{X_j}$, between d_j and d_{ref} *is* given by

$$\mathbf{X_j} = \det(\mathbf{Y_j}\mathbf{Y_j}^T) = (\|d_{ref}\| \,.\, \|d_j\| \sin\alpha_j)^2 \quad \text{for j} = 1, \ldots, nc \qquad (2.4)$$

where α_j is the angle determined between d_j and d_{ref}. T denoted transpose, and the double bars represent the Euclidean norm.

For simplification, the mean spectrum, and in general all reference spectra in $\mathbf{Y_j}$, are normalized to length equal to 1. The spectra d_j could be also normalized to constant length, but it has been shown that the method is more sensitive [13] if both factors, i.e., the angle α_j and length of the spectra, are considered in the dissimilarity measurement. The determinant of the matrix of $\mathbf{Y_j}\mathbf{Y_j}^T$ in eq. 2.4 measures the area of the parallelogram determined by each spectrum d_j and the mean spectrum d_{ref}. The higher the area, the higher the dissimilarity. The first spectrum selected, d_{s1}, is the one most dissimilar with respect to the mean spectrum.

In the next step, the spectrum selected, d_{s1}, is taken as reference in $\mathbf{Y_j}$. Thus $\mathbf{Y_j}$ is a matrix containing d_{s1} and d_j.

$$\text{Dispersion matrix } (\mathbf{Y_j}) = [\, d_{s1} \quad d_j \,] \qquad j = 1,\ldots,\text{nc} \qquad (2.5)$$

As before, the determinant of the matrix of $\mathbf{Y_j}\mathbf{Y_j}^T$ ($\mathbf{Y_j}$ is the dispersion matrix in eq. 2.5) is calculated, and the spectrum that yields the highest determinant, i.e., the most dissimilar with respect to d_{s1}, is selected. The process is repeated, i.e. including the second spectrum selected, d_{s2} as a reference spectrum in $\mathbf{Y_j}$. Thus, $\mathbf{Y_j}$ now contains three spectra, i.e., d_{s1}, d_{s2}, and d_j and each d_j is compared with respect to d_{s1} and d_{s2} by determining the determinant of the matrix of $\mathbf{Y_j}\mathbf{Y_j}^T$.

$$\text{Dispersion matrix } (\mathbf{Y_j}) = [\, d_{s1} \;\; ds_2 \;\; d_j \,] \qquad j = 1,\ldots,\text{nc} \qquad (2.6)$$

At each iteration, the dissimilarity from determinant of the matrix $\mathbf{Y_j}\mathbf{Y_j}^T$ is plotted as a function of variable. The procedure continues adding new reference spectra to $\mathbf{Y_j}$, until the dissimilarity plot represents only noise, see Figure 2.3.

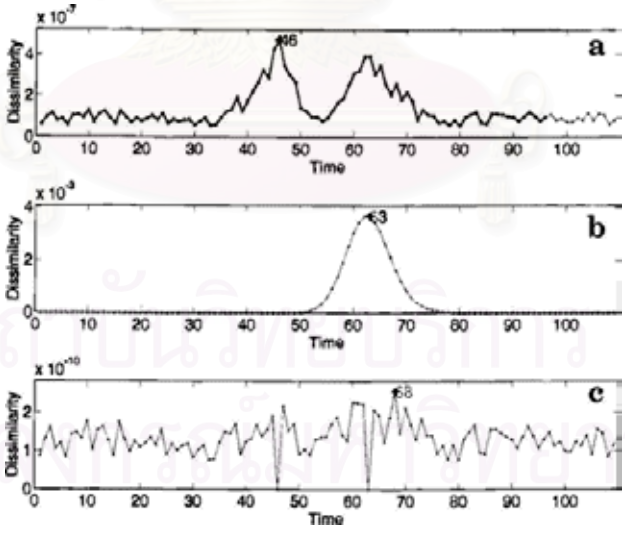


Figure 2.3 Dissimilarity plot of each spectrum with respect to **(a)** the mean spectrum, **(b)** the spectrum at variable 46, and **(c)** the spectra at variables 46 and 63 , this dissimilarity plot in (c) represents only noise.

A random profile of the dissimilarity plot indicates that the number of component and corresponding spectra have been selected in the iteration before. In the ideal

situation, the number of spectra selected is equal to the number of components presenting in the mixture.

2.2.2.2 SIMPLe-to-use Interaction Self-modeling Mixture Analysis approach (SIMPLISMA)

SIMPLISMA is based on the selection of what are called pure variables or objects. A pure variable is a wave-length at which only one of the components in the system is absorbing. When the spectra of both minor and main compounds are very similar, normally pure wavelengths or pure variable could not be established. However in most cases, one can always find, at least the main or region belong to any pure components. SIMPLISMA consists of the following steps.

In the first step, the standard deviation (σ_j) and mean (μ_j) of each spectrum are determined. The ratio between deviation and mean of each spectrum (*p0j*) is calculated as correlation factor.

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{NR}\left(d_{ij} - u_j\right)^2}{NR}} \qquad j = 1,\ldots,nc \qquad (2.7)$$

and

$$\mu_j = \frac{\sum_{i=1}^{NR} d_{ij}}{NR} \qquad j = 1,\ldots,nc \qquad (2.8)$$

when

$$p0\,j = \frac{\sigma_j}{u_j} \qquad j = 1,\ldots,nc \qquad (2.9)$$

In the second step, spectrum d_j (for j = 1,…, nc) is normalized by dividing each row element (d_{ij}) in data matrix by the length of the row $\| d_j \|$.

$$z_{ij} = \frac{d_{ij}}{\left\| d_j \right\|} \qquad i = 1, \ldots, nr\ ;\ j = 1,\ldots,nc \qquad (2.10)$$

where $\|d_j\| = \sqrt{\sum_{i=1}^{NR} d_{ij}^2} = \sqrt{NR.(\sigma_j^2 + u_j^2)}$ j = 1,…, nc (2.11)

and z_{ij} is the element of the normalized data matrix.

The third step is the determination of the weight of each spectrum, W_j. The weight is defined as the determinant of the matrix of $\mathbf{Y_j}^T\mathbf{Y_j}$ containing the normalized spectra and each individual normalized spectrum z_j.

$$W_j = \det(\mathbf{Y_j}^T.\mathbf{Y_j}) \qquad j = 1,\ldots,nc \tag{2.12}$$

Initially, when no spectrum has been selected, each $\mathbf{Y_j}$ contains only one column, z_j, and the weight of each spectrum is equal to the square of the length of the normalized spectrum:

$$W_j = \det(\mathbf{Y_j}^T.\mathbf{Y_j}) = \| z_j^2 \| \qquad j = 1, \ldots, nc \tag{2.13}$$

When the first spectrum has been selected, z_{s1}, each matrix $\mathbf{Y_j}$ consists of two columns: z_{s1} and each individual spectrum z_j, and the weight are equal to:

$$W_j = \det(\mathbf{Y_j}^T.\mathbf{Y_j}) = (\| z_{s1} \| . \| z_j \| . \sin\alpha_j)^2 \qquad j = 1, \ldots, nc \tag{2.14}$$

When two spectra have been selected, z_{s1} and z_{s2}, each $\mathbf{Y_j}$ consists of those two selected spectra and each individual spectrum z_j, and so on. The determination of the weight is very similar to the determination of the dissimilarity in OPA.

At the last step, the purity p_j of each spectrum is determined and plotted as a function of variable:

$$p_j = p0j.\ W_j \qquad j = 1,\ldots,nc \tag{2.15}$$

SIMPLISMA is an interactive approach, i.e. the user can decide whether a spectrum ought to be selected or not. If the spectrum with the highest purity value corresponds to noise, which can be seen by looking at the standard deviation, then the

offset should be used in the denominator of eq. (2.16) and the procedure is repeated from the first step.

$$p0\,j = \frac{\sigma_j}{u'_j} \qquad j = 1, \ldots, nc \tag{2.16}$$

with $u'_j = u_j + (\text{offset}/100) * \max(u_j)$ $\qquad j = 1, \ldots, nc$

Otherwise, the spectrum with the highest purity value is selected and steps 3 and 4 are repeated. At each iteration, the weight (W_j) from determinant of the dispersion matrix is plotted as a function of variable shown in Figure 2.4.

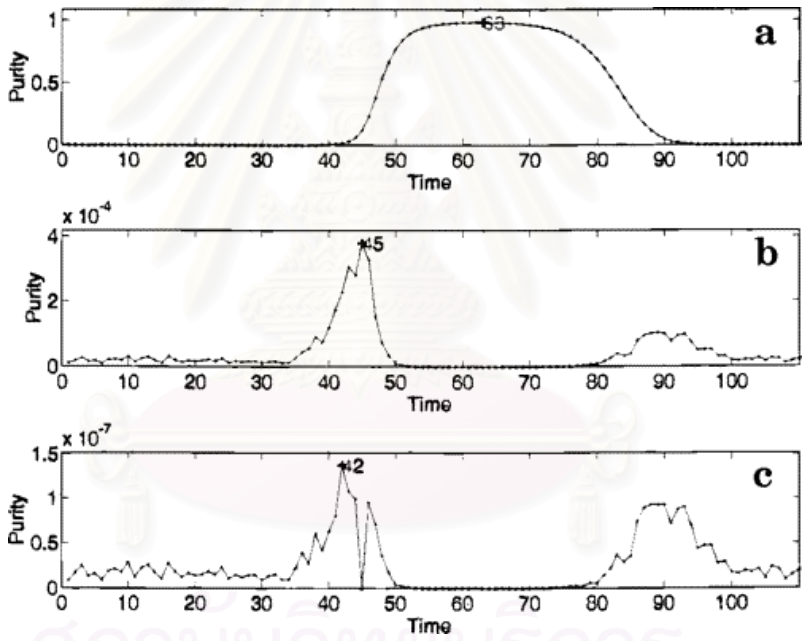


Figure 2.4 Purity plot when **(a)** no spectrum has been selected, **(b)** the spectrum at variable 63 has been selected, and **(c)** the spectra at variables 63 and 45 have been selected. The purity plot in (c) represent only noises.

A random profile of the purity plot and significantly decreasing of purity value indicate the number of component and corresponding spectra. The initial estimation in this algorithm is a number of components in the system and corresponding spectra.

2.2.2.3 Evolving Factor Analysis (EFA)

The core of EFA method is the determination of the chemical ranks of the system matrix under investigation. The analysis is performed in the forward that is the size of sub-matrices ($\mathbf{D_j}$) linearly increases from the start (the first column of $\mathbf{D}$) to the end (the last column of $\mathbf{D}$) and backward that is the size of sub-matrices linearly increases from the end to the start. In general, the series of chemical rank analyses are performed using singular value decomposition (SVD):

$$\mathbf{D} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V} \qquad (2.17)$$

(nr x nc) (nr x **n**) (n x **n**) (**n** x nc)

Where **U** is formed by the significant eigenvectors of $\mathbf{DD}^T$.

V is formed by the significant eigenvectors of $\mathbf{D}^T\mathbf{D}$.

S is a diagonal matrix, its element are the positive square root of significant eigenvalue of $\mathbf{DD}^T$ or $\mathbf{D}^T\mathbf{D}$.

n is the number of components.

The number of component is investigated by the change or the evolution of the rank of sub matrices $\mathbf{D_j}$, formed by the *first 1,2,…j,…,nc* spectra of data matrix $\mathbf{D}$. The appearance of each new component is associated with the increase of the rank by one. When a new absorbing species begins to appear, an eigenvalue evolves from the pool of error eigenvalues, increasing in value in relation to its contribution to the enlarged data set. This procedure is called "*forward evolving factor analysis*". The increasing of the rank is detected by plotting between the log of the singular values ($\mathbf{S_j}$) of the submatrices $\mathbf{D_j}$ and the variable (pH or time).

$$\mathbf{d_j} = [\mathbf{d_j} \ \ldots \ \mathbf{d_{j+1}}], \quad j = 1, 2, \ldots, nc\text{-}1 \ \text{(forward direction)} \qquad (2.18)$$

Here, $\mathbf{D_j}$ is the j^{th} column of data matrix $\mathbf{D}$. The EFA forward plots between a series of diagonal matrices $\mathbf{S_j}$ and the retention time corresponding to the end rows of the sub-matrix $\mathbf{D_j}$., where the appearance of every new factor is easily detected and thus

the total number of components is an obvious result of such an analysis is shown in Figure 2.5

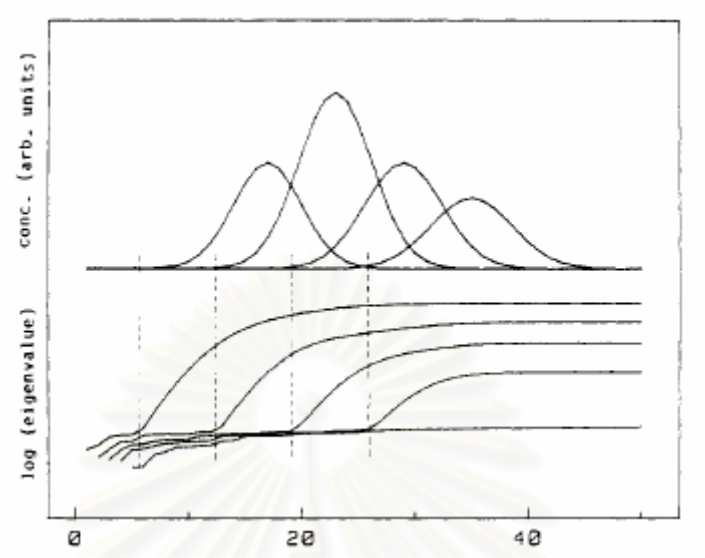


Figure 2.5 *Upper part* : the model concentration profiles used to generate the data.
Lower part : the EFA forward plot indicating the elution of a new component around the variables 6, 12, 19 and 26.

It is straightforward to carry out the backward calculation by repeating the EFA calculation from the opposite end. The backward calculation is performing a rank analysis of the sub matrices $\mathbf{D_j}$ formed by the *last 1,2,…,j,…nc* spectra. In this way, the information about the disappearance of the component is obtained. The procedure is called "*Backward evolving factor analysis*" The disappearance of the component is detected by plotting between the log of the singular values ($\mathbf{S_j}$) of the sub matrices $\mathbf{D}_i$ and the variable (pH or time).

$$\mathbf{D_j} = [\mathbf{D}_{nc} \ \ldots \ \mathbf{D}_{nc-j}], \quad j = 1, 2, \ldots, nc\text{-}1 \quad \text{(backward direction)} \tag{2.19}$$

Here, $\mathbf{D_i}$ is the j^{th} column of data matrix $\mathbf{D}$

After the decomposition of all the sub-matrices in both the forward and backward directions, a rank map is obtained by plotting the logarithms of the singular values, contained in a series of diagonal matrices $\mathbf{S_j}$, versus the retention time corresponding to

the end rows of the sub-matrix $\mathbf{D_i}$. The number of significant components is selected from the shape of plotting by increasing of logarithms of singular values in forward direction and decreasing of logarithms of singular values in backward direction when the dimension of $\mathbf{D_j}$ increase, whereas the noise level is not changing. The noise level plot is located in the baseline region. For example, this EFA plot is shown in Figure 2.4. This plot is expressed that there are four significant components in this system.

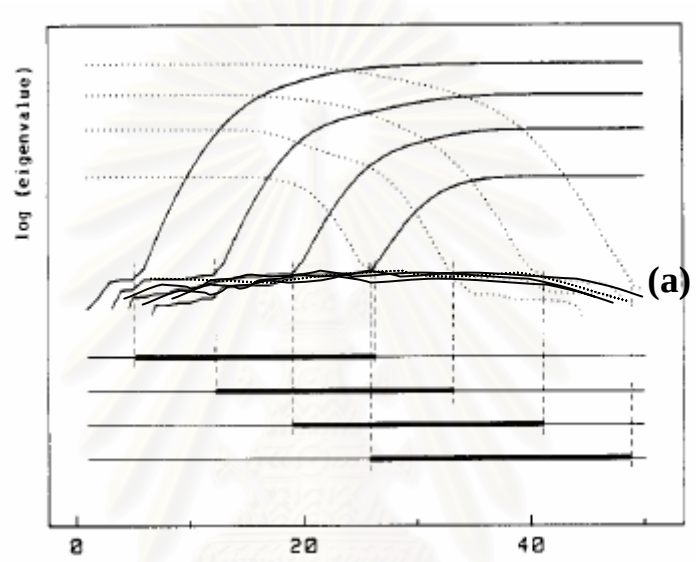


Figure 2.4 Concentration window (**-**) for the j^{th} component is defined by the rise of the j^{th} eigenvalue in the forward EFA plot (-) and the $(nc+1-j)^{th}$ eigenvalue in the backward EFA plot (**…**) . The noise level is plotted in the baseline region (**a**)

In the next part a noniterative calculation of the concentration profiles is proposed. **D** can be decomposed either according to eq. 1.3 or according to the singular value decomposition eq 2.16 Combination of eq 1.3 and 2.16 ($\mathbf{AC = USV}$) and multiplication from the right with the nonsingular matrix $\mathbf{V^T(AV^T)^{-1}}$ gives

$$\mathbf{C = U S (AV^T)^{-1} = UR} \qquad (2.20)$$

$\mathbf{R = S(AV)^{-1}}$ is a nonorthogonal rotation or projection matrix using the zero – concentration region for the $\mathbf{n}^{th}$ component to transform **U** matrix into initial concentration.

2.2.3 Optimization

After the selection of the pure components and their corresponding spectra, the optimization method called Multivariate Curve resolution is performed. The multivariate curve resolution is an effective tool in self-modeling technique. The task of multivariate curve resolution is to determine the matrices of absorptivity (**A**) and concentration (**C**). Let's begin with the selective pure spectra in the selection parts. Then **A** can be approximated by these row matrix and pure concentration of components can be obtained by the iterative procedure.

$$\mathbf{C} = \mathbf{A}^{+}.\mathbf{D} \tag{2.21}$$

where $\mathbf{A}^{+}$ is the pseudoinverse of **A** matrix. Then the new estimation of absorptivity is obtained by calculation as

$$\mathbf{A} = \mathbf{D}.\mathbf{C}^{+} \tag{2.22}$$

where $\mathbf{C}^{+}$ is the pseudoinverse of **C** matrix.

These procedures are repeated iteratively until the relative difference in the lack of fit values of two consecutive iterations is lower than a pre-defined convergence limit. To reduce the ambiguity of optimization three constraints are applied during optimization : (1) non-negativity- all concentration and absorptivity values must be positive, and (2) constant total concentration (closure) – the total concentrations of each solution must be equal or normalized, and (3) unimodality – the concentration of some species would be unimodality (have one maximum peak).

$$\text{Lack of fit} = \sqrt{\frac{\sum_i \sum_j \left(d_{i,j} - \hat{d}_{i,j}\right)^2}{\sum_i \sum_j \left(d_{i,j}\right)^2}} \tag{2.23}$$

$d_{i,j}$ the data point in the experiment data

$\hat{d}_{i.j}$ the data point in the calculated data

2.2.4 Prediction

At the end of the algorithm, concentration profiles (**C**) and pure spectra (**A**) are stable, and the data matrix was resolved. The prediction provides the number of significant species in the system and corresponding concentration and absorptivity of each species.

2.3 Fundamental Concept of Chemical Equilibria

This section concerned the fundamental concept of chemical equilibria that underlying the problem of complex formation and acid-dissociation system. There are two particular equilibrium constants that are commonly given special names. When the Lewis acid is a proton ionization, the equilibrium constant for the reaction is known as the acid dissociation constant (K_a^θ) of the acid HA.

$$HA \overset{K_a}{\rightleftharpoons} H^+ + A^-$$

$$K_a^\theta = \frac{f_{H+}[H^+]f_{A-}[A^-]}{f_{HA}[HA]} \qquad (2.24)$$

The second special case is when the Lewis acid is a metal ion (M) and the Lewis base is a ligand (L).

$$M^+ + L^- \overset{K_{ML}}{\rightleftharpoons} ML$$

$$K_{ML}^\theta = \frac{f_{ML}[ML]}{f_M[M^+]f_L[L^-]} \qquad (2.25)$$

The equilibrium constant (K_{ML}^θ) is known as the formation constant of the complex (ML). The activity coefficient, f_i , is general dreary and difficult to measure. They also depend very significantly on the nature and concentrations of the other species that present in solution. To avoid this problem, the background of ionic strength in electrolyte is used to maintain the activity coefficients effectively constant. It can incorporate the f_i terms into K_a^θ or K_{ML}^θ and obtain the general forms as

$$K_a = \frac{[H^+][A^-]}{[HA]} \quad (2.26)$$

$$K_{ML} = \frac{[ML]}{[M^+][L^-]} \quad (2.27)$$

Where K is known as stoichiometric equilibrium constants whereas K^θ is known as thermodynamic equilibrium constants.

2.3.1 Acid-Base equilibria

Consider a system of *k* step acid dissociation denoted as H_kA, the equilibria are established as follow.

$$\begin{aligned}
H_kA &\rightleftharpoons H_{k-1}A^- + H^+ ; & Ka_1 &= \frac{[H^+][H_{k-1}A^-]}{[H_kA]} \\
H_{k-1}A &\rightleftharpoons H_{k-2}A^{2-} + H^+ ; & Ka_2 &= \frac{[H^+][H_{k-2}A^{2-}]}{[H_{k-1}A]} \\
&\vdots & &\vdots \\
HA^{(k-1)-} &\rightleftharpoons A^{k-} + H^+ ; & Ka_k &= \frac{[H^+][A^{k-}]}{[HA^{(k-1)-}]}
\end{aligned} \quad (2.28)$$

To express the distribution of each component, we introduce the concept of degree of formation, α, the mole ratio of one component with respect to all components. The mole ratio of component is expressed in the general form as

$$\alpha_{k+1} = \frac{[A^{k-}]}{[H_kA] + [H_{k-1}A^-] + \ldots + [A^{k-}]} \quad (2.29)$$

Substituting in equation (2.29) with (2.28) and then rearranging it, we obtain the new general form of the mole ratio of component.

$$\alpha_{k+1} = \frac{1}{\frac{[H^+]^k}{Ka_1Ka_2\ldots Ka_k} + \frac{[H^+]^{k-1}}{Ka_2Ka_3\ldots Ka_k} + \ldots + \frac{[H^+]}{Ka_k} + 1} \quad (2.30)$$

Concentrations of all species can be determined in terms of α_k, degree of dissociation of order *k*.

$$[HA^{(k-1)-}] = C_O \cdot \alpha_{k+1} \qquad (2.31)$$

where C_O is the initial concentration of acid

2.3.2 Metal-Ligand Complex Equilibria

Equilibria of a system of *p* step formation of metal (M) and protonated ligand (HL) can be expressed by the equation

$$\begin{aligned}
M^+ + HL &\rightleftharpoons ML + H^+ & K_1 &= \frac{[ML][H^+]}{[M^+][HL]} \\
ML + HL &\rightleftharpoons ML_2 + H^+ & K_2 &= \frac{[ML_2][H^+]}{[ML][HL]} \\
&\vdots & &\vdots \\
ML_{p-1} + HL &\rightleftharpoons ML_p + H^+ & K_p &= \frac{[ML_p][H^+]}{[ML_{p-1}][HL]}
\end{aligned} \qquad (2.32)$$

In many literatures, the concept of overall or cumulative stability constants, usually denoted by β_i is employed. This is expressed by

$$\begin{aligned}
(x_1)\,M + (y_1)\,HL &\xrightleftharpoons{\beta_1} ML^{(1)} + (z_1)\,H^+ \\
(x_2)\,M + (y_2)\,HL &\xrightleftharpoons{\beta_2} ML^{(2)} + (z_2)\,H^+ \\
&\vdots \\
(x_p)\,M + (y_p)\,HL &\xrightleftharpoons{\beta_p} ML^{(p)} + (z_p)\,H^+
\end{aligned} \qquad (2.33)$$

The general form of cumulative stability constants from equation (2.32) and (2.33) is

$$\beta_p = \prod_{i=1}^{p} K_i = K_1 K_2 \ldots K_p = \frac{[ML^{(p)}][H^+]^{zp}}{[M]^{xp}[HL]^{yp}} \qquad (2.34)$$

By the principle of mass balance and the assumption of mononuclear complex, we obtain

$$Mtot = [M] + [ML^{(1)}] + [ML^{(2)}] + \ldots + [ML^{(p)}]$$

$$[M] = \frac{M_{tot}}{1 + \beta_1 \frac{[HL]^{y1}}{[H^+]^{z1}} + \beta_2 \frac{[HL]^{y2}}{[H^+]^{z2}} + \ldots + \beta_p \frac{[HL]^{yp}}{[H^+]^{zp}}} \qquad (2.35)$$

where M_{tot} is the total concentration of metal (M). The protonated ligand may exist in dynamic equilibrium with its conjugated acid and/or base.

$$H_2L \rightleftharpoons HL^- + H^+ \qquad Ka_1 = \frac{[HL^-][H^+]}{[H_2L]}$$

$$HL^- \rightleftharpoons L^{2-} + H^+ \qquad Ka_2 = \frac{[L^{2-}][H^+]}{[HL^-]}$$

Thus, the mass balance of the ligand can be expressed as

$$L_{tot} = [H_2L] + [HL^-] + [L^{2-}] + y_1[ML^{(1)}] + y_2[ML^{(2)}] + \ldots + y_p[ML^{(p)}]$$

$$= [HL^-]\left(\frac{[H^+]}{Ka_1} + 1 + \frac{Ka_2}{[H^+]}\right) + [M]\left(y_1\beta_1\frac{[HL^-]^{y1}}{[H^+]^{z1}} + y_2\beta_2\frac{[HL^-]^{y2}}{[H^+]^{z2}} + \ldots + y_p\beta_p\frac{[HL^-]^{yp}}{[H^+]^{zp}}\right) \qquad (2.36)$$

where L_{tot} is the total concentration of ligand. Substitute (2.35) into (2.36), and applying the binary search for determining the concentration of HL. Consequently, evaluate the concentration of M, and substitute successively into (2.34) then obtain the general form of the mole ratio, α, of $ML^{(p)}$

$$\alpha_{p+1} = \frac{[ML^{(p)}]}{[M] + [ML^{(1)}] + [ML^{(2)}] + \ldots + [ML^{(p)}]}$$

$$= \frac{1}{\frac{[H^+]^{zp}}{\beta_p[HL^-]^{yp}} + \frac{\beta_1[H^+]^{zp-z1}}{\beta_p[HL^-]^{yp-y1}} + \frac{\beta_2[HL^-]^{yp-y2}}{\beta_p[H^+]^{zp-z2}} + \ldots + 1} \qquad (2.37)$$

Concentration of all species can be determined in terms of α , degree of complexation of order *p*.

$$[ML^{(p)}] = M_{tot} \cdot \alpha_{p+1} \qquad (2.38)$$

where M_{tot} is the initial concentration of metal solution.

CHAPTER III

PROGRAM IMPLEMENTATION

The SMCR program was developed with the MATLAB language. The main application of the program is to resolve absorbance spectra of multi-component mixtures into pure spectra and concentration profiles of each component. In addition, the program can also be further adapted to resolve data from the other applications. The program manual and example were given in the appendix.

The program can be separated into four parts:

3.1 Main routine

3.2 Chemical rank analysis and spectral resolution

3.3 Optimization.

3.4 Summary of the requirements for input and default values.

3.5 Summary of the output

The SMCR program contains 16 subroutines for calculation and 10 subroutines for performing graphical user interface (GUI). They are listed in Table 3.1 and 3.2 respectively. The main routine controls the program flow. All subroutines in the SMCR program were newly implemented except the *nnls.m* subroutine for non-negative constraint which was taken from the built-in source code of MATLAB. and *ALS.m* subroutine for alternating least square optimization which was modified from the original codes of Prof. Dr. Roma Tauler.

3.1 Main routine

The main subroutine controls program flows and input-output of the program. It was designed to display input and output in the graphical user interface (GUI) manner. Following items are required for the input of the program.

- Spectra filename (*.txt)
- Range of variables (pH or monitoring time)
- The chemical rank and corresponding initial guess using EFA, OPA and SIMPLISMA subroutines
- Convergence criteria
- Output direction

Table 3.1 List of all subroutine files of the program SMCR version 1.0

File	Function
Main subroutine	
SMCR.m	Main subroutine
Mprint_SMCR.m	Subroutine for saving the results
Chemical rank analysis	
OPA.m	Subroutine of OPA method
- *OPAnumber.m*	Subroutine for determining the chemical rank based on OPA method
- *OPAconc.m*	Subroutine for finding corresponding spectra of each components
SIMPLISMA.m	Subroutine of SIMPLISMA method
- *pure.m*	Subroutine for calculating purity value
- *SIMnumber.m*	Subroutine for determining the chemical rank based on SIMPLISMA
- *SIMconc.m*	Subroutine for finding corresponding spectra of each components
EFA.m	Subprogram for determining the chemical rank and initial guess of concentration of each component based on EFA method.

Table 3.1 (continue)

Optimization	
ALS.m	Subroutine for optimization based on multivariate curve resolution (MCR)
- *nonneg.m*	Subroutine for nonnegative constraint
- *nnls.m*	Build-in subroutine for non-negative least square constraint
- *unimodal.m*	Subroutine for unimodal constraint
- *clos.m*	Subroutine for closure constraint
Statistic testing	
res.m	Subroutine for calculating the norm of error
rms.m	Subroutine for calculating the root-mean square error
nrms.m	Subroutine for calculating the normalized root-mean square error

Table 3.2 List of all subroutine files of graphical user interface (GUI) of program SMCR version 1.0

File	Function
GUI	
GUI_SMCR101	Main GUI
GUI_SMCR_01	Subroutine for performing GUI of preparing data part
GUI_SMCR_02	Subroutine for performing GUI of chemical rank analysis
GUI_SMCR_03	Subroutine for performing GUI of convergence criteria
GUI_SMCR_04	Subroutine for performing GUI of optimization results
GUI_SMCR_05	Subroutine for performing GUI of saving results

The sequence step of input in main subroutine was shown in Figure 3.1.

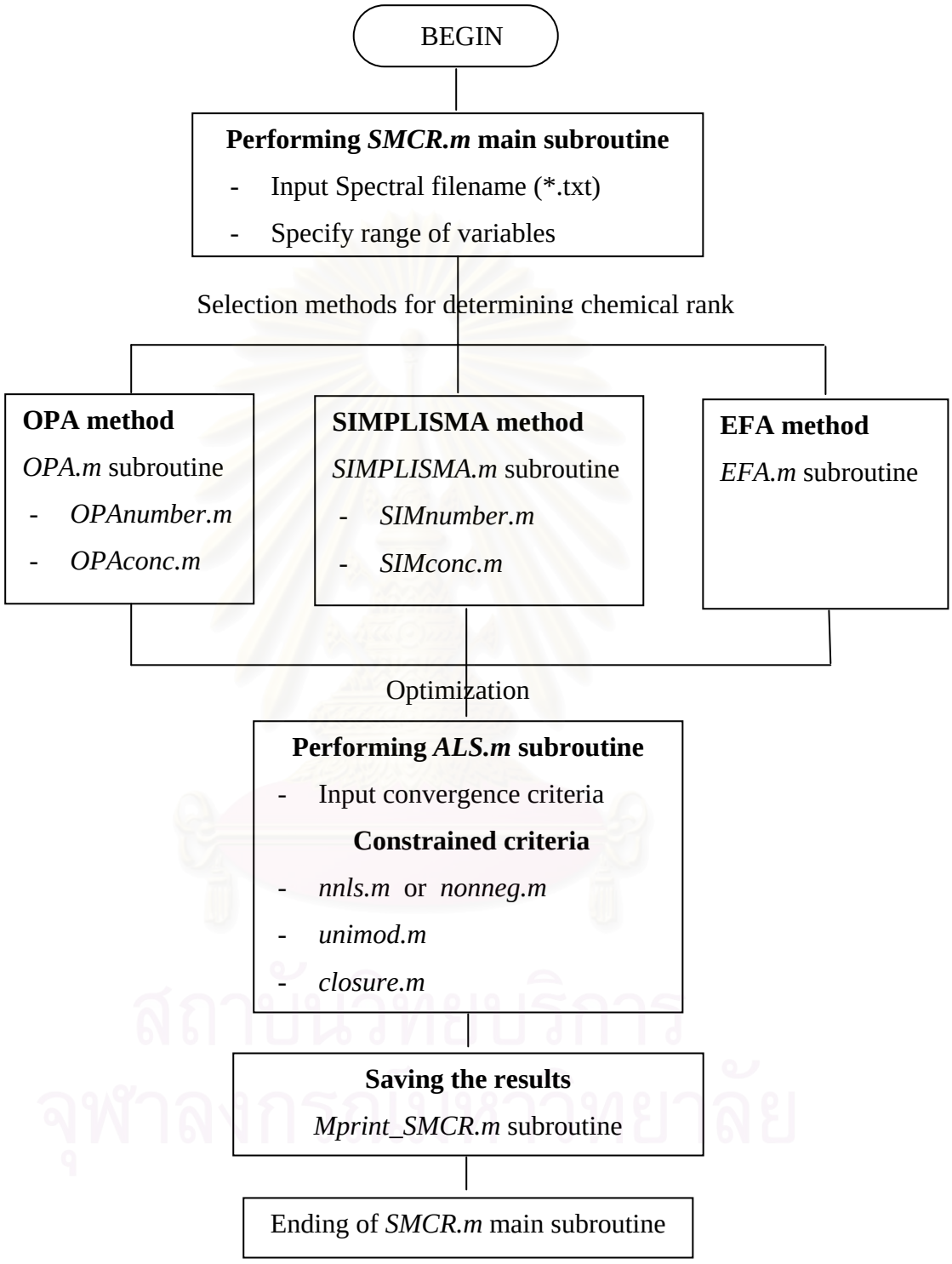


Figure 3.1 Flowchart of the main program of SMCR version 1.0.

3.2 Chemical rank analysis and spectral resolution

The part of chemical rank analysis was designed to perform the Evolving Factor Analysis (EFA), Orthogonal Projection Approach (OPA) and SIMPLe-to-use Interactive Self-modeling Mixture Analysis approach (SIMPLISMA) as describe in the section 2.1.2.

For performing the SMCR program calls *OPA.m.* This subroutine further call *OPAnumber.m* and *OPAconc.m* subroutines to calculate dissimilarity value and select the number of components and corresponding spectra. The flowchart that shows algorithm for *OPA.m* subroutine is given in figure 3.2.

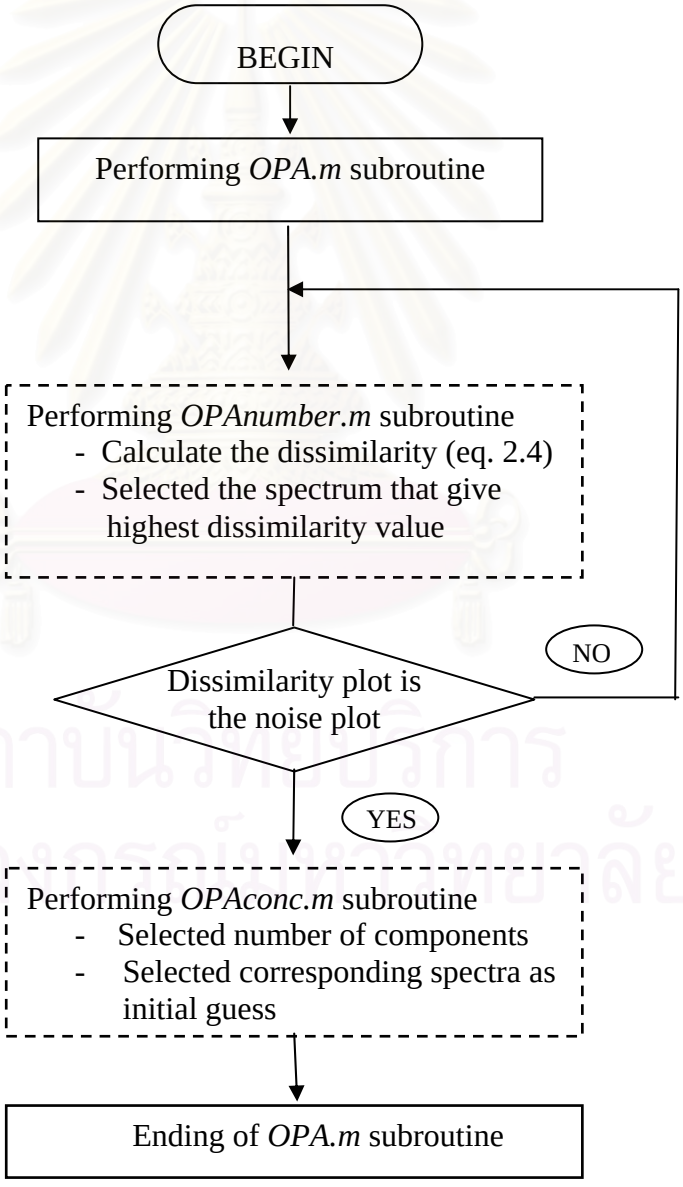


Figure 3.2 Flowchart of the *OPA.m* subroutine

For performing the SIMPLISMA, the SMCR program calls *SIMPLISMA.m* routine which further calls *pure.m, SIMnumber.m* and *SIMconc.m* subroutines to calculate purity value and select the number of components and corresponding spectra. The flowchart that shows algorithm for *SIMPLISMA.m* subroutine is given in figure 3.3.

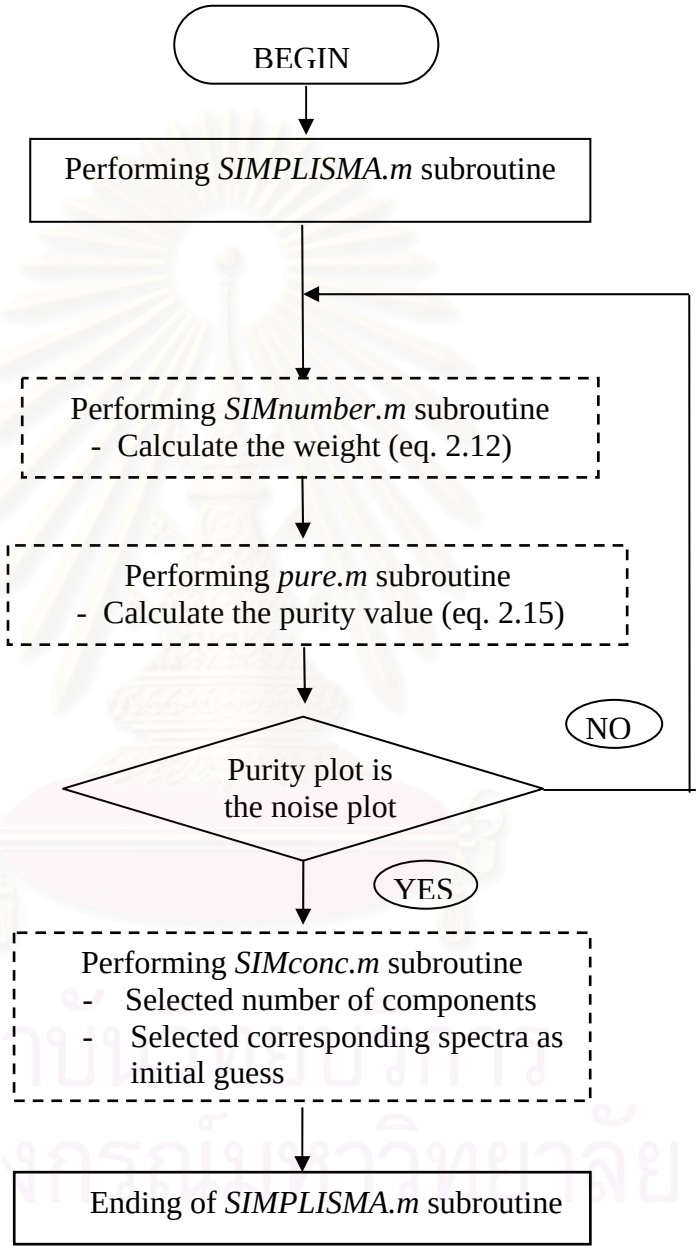


Figure 3.3 Flowchart of the *SIMPLISMA.m* subroutine

For performing EFA, the SMCR program calls *EFA.m* routine to calculate eigenvalue by using singular value decomposition (SVD). The chemical rank is determined by plotting "*forward evolving factor analysis plot*" and "*backward evolving factor analysis plot*". The flowchart that shows algorithm for *EFA.m* subroutine is given in figure 3.4.

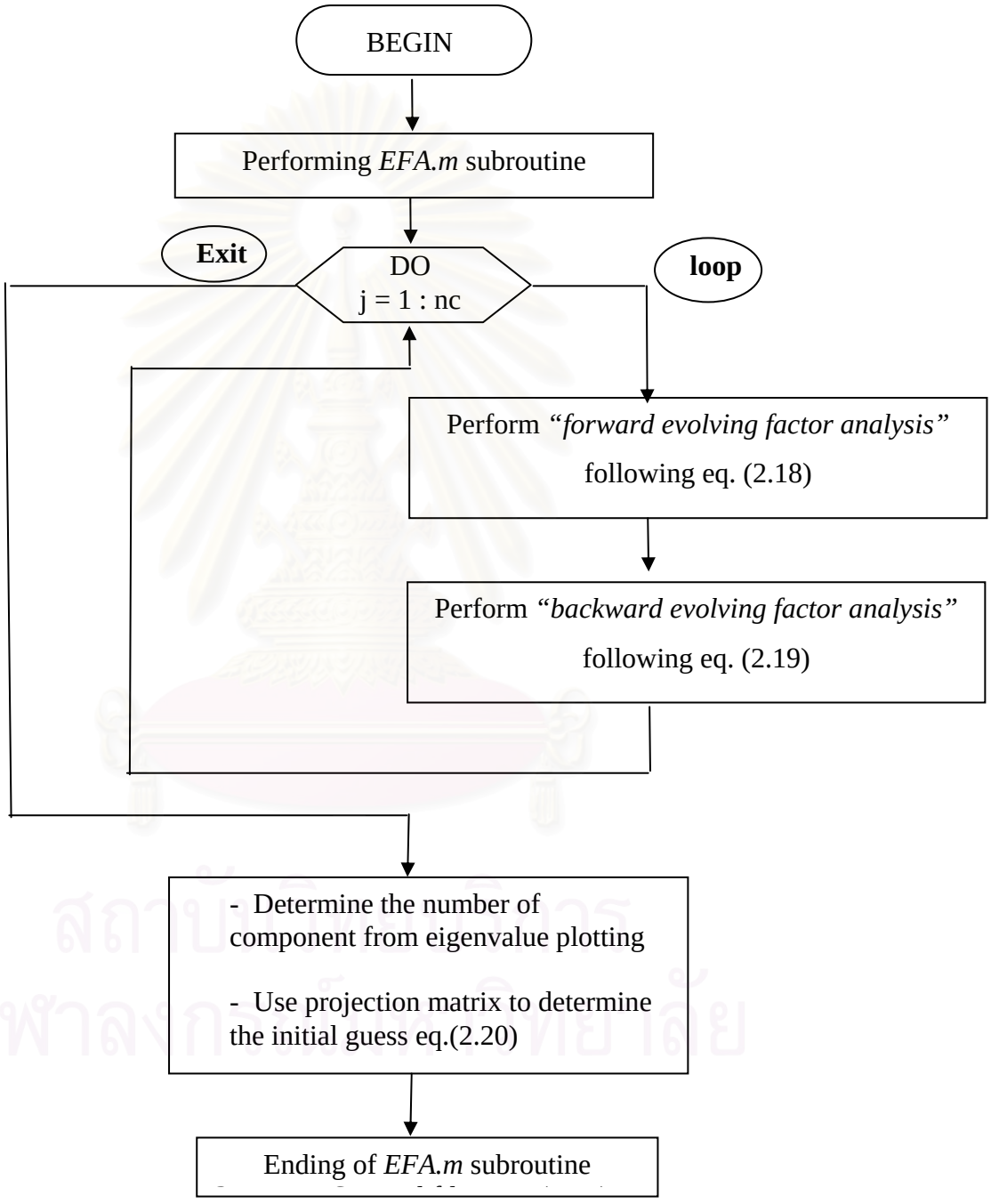


Figure 3.4 Flowchart of the *EFA.m* subroutine.

3.3 Optimization with Multivariate Curve Resolution

The subprogram for optimization was designed to determine exact concentration profiles (C) and absorptivity profiles (A). The procedure of this part and the algorithm of *ALS.m* subroutine were shown in Figure 3.5.

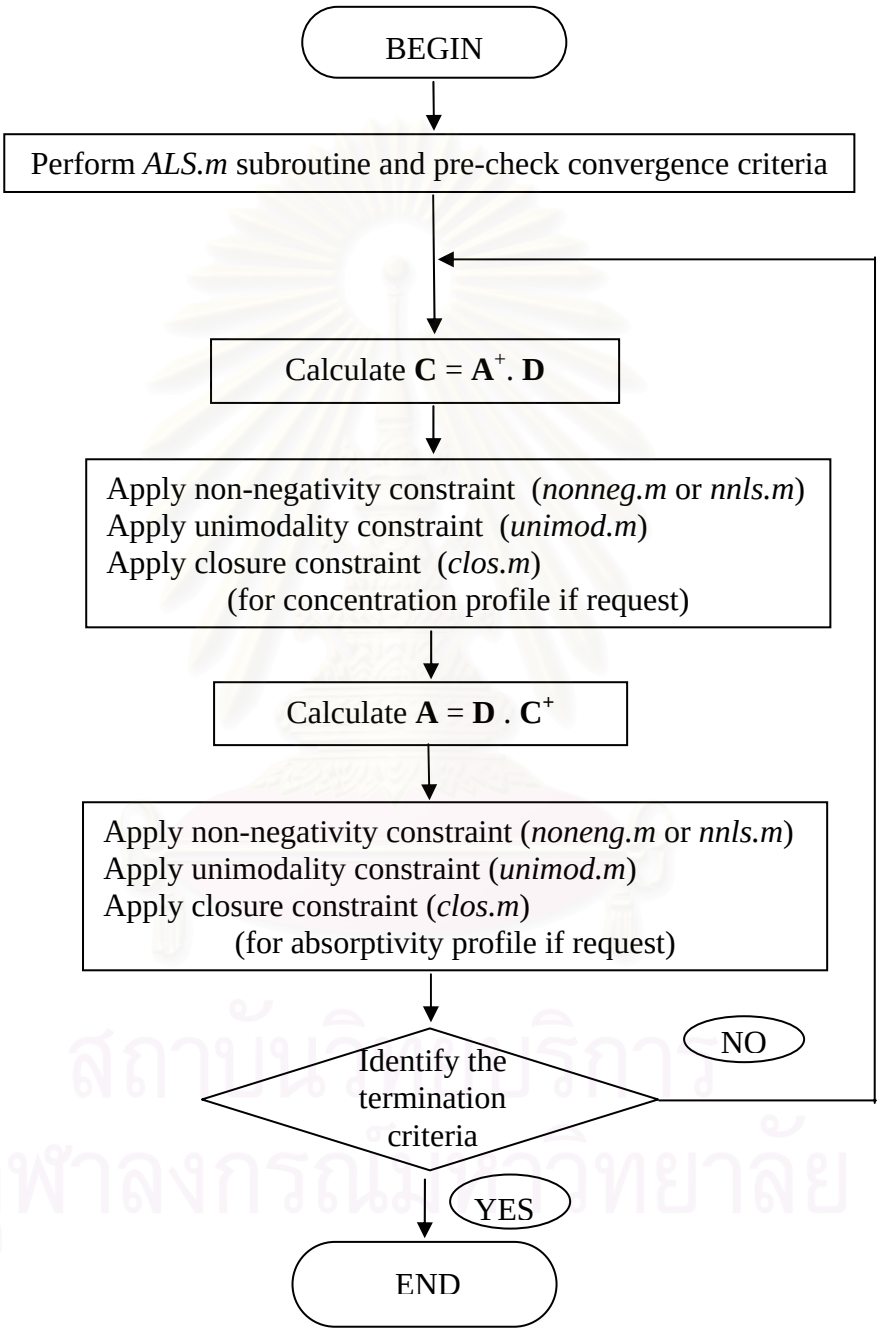


Figure 3.5 Flowchart of the *ALS.m* subroutine.

3.4 Summary of Input Requirements and Default Values

The following are input requirements and default values.

i) Prepare the spectra as a textfile (*.txt) with ASCII format. The example of spectra shown in Appendix part.

ii) Specify the range of variables (Time and pH)

iii) Select an option include OPA, SIMPLISMA and EFA subroutines for setting the automatic or manual for initial guess.

iv) Set convergence criteria for ALS subroutine. The default values are: maximum cycle = 150 and tolerance = 0.01

v) Input the non-negativity, unimodality and closure constraint parameters

vi) Select an option for saving the results.

3.5 Summary of output

The following are output resolved from the SMCR program.

i) Elementary data were shown as spectra filename, dimension of spectra matrix, range of variables, and total concentration value.

ii) Chemical rank analysis were shown as method for estimating chemical rank, and number of significant components.

iii) Resolved concentration and absorptivity profiles would be an ASCEII format.

CHAPTER IV

VALIDATION OF PROGRAM

The mission of this chapter is to validate the SMCR program using spectral simulated data and experimental data. The program validation is an important part of the program development. The purposes of the validation are to test the confident of the program and to determine the efficiency and accuracy of the program. We applied two approaches for the program validation. The first approach is testing with the simulated system where all measured properties are known and the program is applied to find and desired properties. The second approach is testing with known experimental data and results from curve resolution are compared with previous study. If results are in good agreement or correspond to the presetting properties or reference under an acceptable error, then the program is validated.

4.1 Validation Approach

In the program validation, there are two features to be examined, i.e.

(i) validation of program sensible logic.

(ii) validation of the whole program.

The first feature involves the calculations of matrix operation, the orthogonal projection, and the error indicators. Calculations and logic of all subprograms mentioned above were verified and tested individually.

For the second feature, simulated systems and experimental systems were used for the validation. Only whole program testing was reported here. Three simulated chemical application i.e. acid-base equilibria, metal-ligand complex equilibria and chromatographic separation system and two experimental data were used in the validation of the whole program. The detail of which was given in the section 4.2.

4.2 The simulated spectra

The data matrix of simulated spectra for acid-base equilibria, metal-ligand complex equilibria and chromatographic separation system were generate by the multiplication of molar absorptivity matrix, denoted by **A**, and concentration matrix, denoted by **C.**

$$\mathbf{D} = \mathbf{AC} \tag{4.1}$$

Profiles of the simulated absorptivity were generated based on the Guassian distribution function, following equation was used:

$$\varepsilon = \sum_{j=1}^{n} h_{0,j}\, exp\left(-4.ln(2)\left(\frac{x - x_{0,j}}{W_{1/2,j}} \right)^2 \right) \tag{4.2}$$

where n is the number of peak and $h_{0,j}$, $x_{0,j}$ and $W_{1/2,j}$ are the intensity, position and width at half height of peak j respectively. Since we were interested in acid-base equilibria, metal-ligand complex equilibria and chromatographic separation systems, the simulated concentration profiles were generated separately for each problem.

4.2.1) Simulated system of acid-base equilibria.

In this case, a triprotic acid namely H_3A is dissociated as

$$H_3A \underset{}{\overset{K_{a1} = 3.35 \times 10^{-4}}{\rightleftharpoons}} H_2A^- + H^+$$

$$H_2A^- \overset{K_{a2} = 2.20 \times 10^{-6}}{\rightleftharpoons} HA^{2-} + H^+$$

$$HA^{2-} \overset{K_{a3} = 1.35 \times 10^{-7}}{\rightleftharpoons} A^{3-} + H^+$$

where K_{a1}, K_{a2} and K_{a3} are the dissociation constants of the three step dissociation. Four components, represented as H_3A, H_2A^-, HA^{2-} and A^{3-}, were formed as the dissociative species. The concentration matrix, **C**, was formulated by the equation 2.29 and 2.30, where k = 3 i.e.

$$[H_3A] = \alpha_1 C_{tot} = \frac{C_{tot}}{1+\frac{Ka_1}{[H^+]}+\frac{Ka_2Ka_1}{[H^+]^2}+\frac{Ka_3Ka_2Ka_1}{[H^+]^3}}$$

$$[H_2A^-] = \alpha_2 C_{tot} = \frac{C_{tot}}{\frac{[H^+]}{Ka_1}+1+\frac{Ka_2Ka_1}{[H^+]}+\frac{Ka_3Ka_2Ka_1}{[H^+]^2}}$$

$$[HA^{2-}] = \alpha_3 C_{tot} = \frac{C_{tot}}{\frac{[H^+]^2}{Ka_1Ka_2}+\frac{[H^+]}{Ka_2}+1+\frac{Ka_3}{[H^+]}}$$

$$[A^{3-}] = \alpha_4 C_{tot} = \frac{C_{tot}}{\frac{[H^+]^3}{Ka_1Ka_2Ka_3}+\frac{[H^+]^2}{Ka_2Ka_3}+\frac{[H^+]}{Ka_3}+1} \quad (4.3)$$

where $[H_3A]$, $[H_2A^-]$, $[HA^{2-}]$, and $[A^{3-}]$ are row vectors of the matrix **C**.

For the acid-base equilibria, twenty solutions of acid (H_3A) with the constant total concentration of 0.1000 M and pH varying from 2 to 12 at intervals of 0.5 were simulated. The twenty solutions of H_3A were recorded from 200 to 600 nm at intervals of 1 nm. The dimension of matrix is shown in Table 4.1.

Table 4.1 Dimensions of each matrix for acid-base equilibria

Cases of Equilibria	Data matrix (**D**)	Dimensions Matrix of Absorptivities (**A**)	Dimensions Matrix of Concentration (**C**)
Acid-base equilibria	(400 x 21)	(400 x 4)	(4 x 21)

The molar absorptivity profiles, **A**, and the concentration profiles, **C**, were shown in Figure 4.1 and Figure 4.2, respectively.

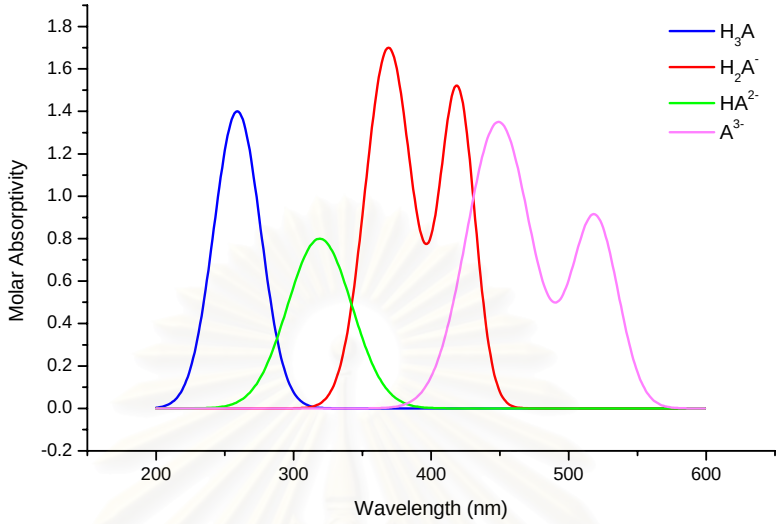


Figure 4.1 Simulated molar absorptivity spectra of 4 components for the acid-base equilibria

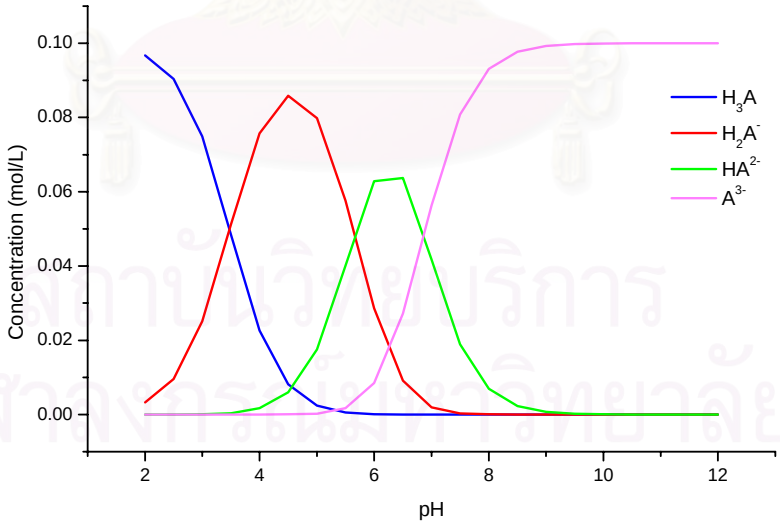


Figure 4.2 Simulated concentration profiles of 4 components computed by equation 4.3

It is well-known that experiment uncertainly does exist and blends into the pure data. Therefore, the error matrix (**E**) was added to the equation (4.1). Thus,

$$\mathbf{D} = \mathbf{AC} + \mathbf{E} \qquad (4.4)$$

The elements in error matrix ($\mathbf{E_{ij}}$) are generated using the pseudo-random number. Since the standard deviation in absorbance was estimated to vary at 0.0005 absorbance unit (approximately 5% of relative noise level of the mean absorbance value). Therefore, level of error of ±0.0005 was introduced and added to the simulated spectra. The simulated spectra of the acid-base system were shown in Figure 4.3.

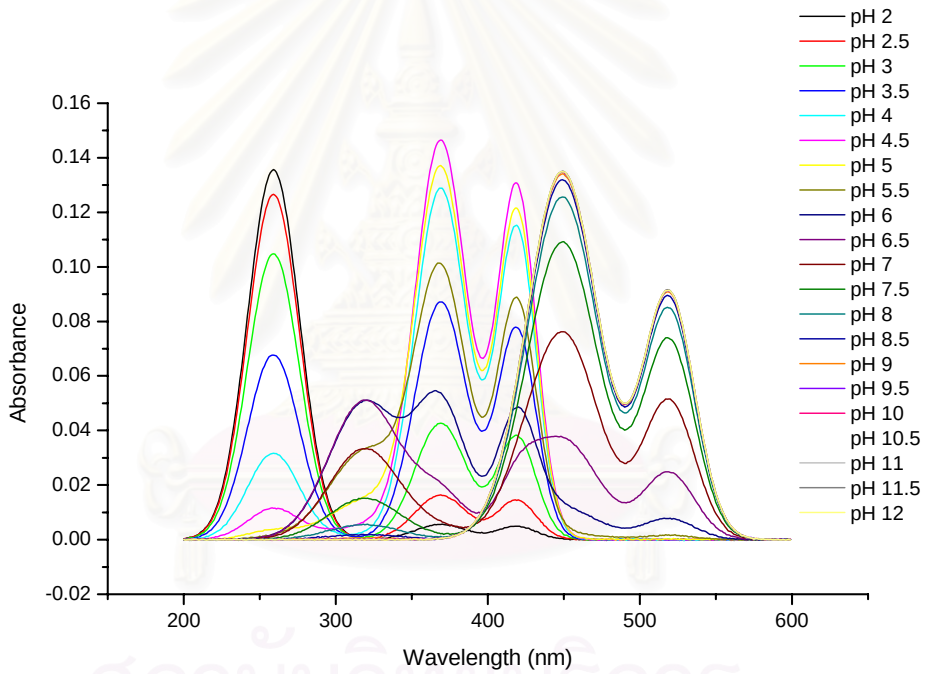


Figure 4.3 Twenty simulated spectra of H_3A dissociation with added ±0.0005 errors.

4.2.2) Simulated spectra of metal-ligand complex formation system.

For the simulated system, a metal (M^{2+}) and protonated ligand (HL^-) are in the equilibria as shown below.

$$H_2L \underset{}{\overset{K_{a1} = 4.467 \times 10^{-4}}{\rightleftharpoons}} HL^- \overset{K_{a2} = 1.667 \times 10^{-10}}{\rightleftharpoons} L^{2-}$$

$$M^{2+} + HL^- \overset{\beta_1 = 7.59}{\rightleftharpoons} M(HL)^+$$

$$M^{2+} + HL^- \overset{\beta_2 = 5.39 \times 10^{-2}}{\rightleftharpoons} ML + H^+$$

$$M^{2+} + 2\,HL^- \overset{\beta_3 = 3.37 \times 10^{-5}}{\rightleftharpoons} ML_2^{2-} + 2\,H^+$$

K_{a1} and K_{a2} are the acid dissociation constants of the ligand, and β_1, β_2, and β_3 are the overall stability constants of complexation. This chemical equilibria can be

expressed in matrix form

$$\begin{array}{c} \\ M(HL)^+ \\ ML \\ ML_2^{2-} \end{array} \begin{array}{c} \begin{array}{ccc} M & HL^- & H^+ \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \end{array}$$

The columns of the matrix represent M, HL^- and H^+ species which are X_p, Y_p and Z_p in eq. 2.36 respectively, *p* refers to the step of metal ligand complexation. Here, we assign *p* = 3, and 4 components; M^{2+}, $M(HL)^+$, ML and ML_2^{2-} were formed. The concentration matrix, **C**, were formulated by the equation 2.36 and 2.37.

$$[M^{2+}] = \alpha_1 . M_{tot} = \frac{M_{tot}}{1 + \beta_1 [HL^-] + \beta_2 \frac{[HL^-]}{[H^+]} + \beta_3 \frac{[HL]^2}{[H^+]^2}}$$

$$[M(HL)^+] = \alpha_2 . M_{tot} = \frac{M_{tot}}{\frac{1}{\beta_1 [HL^-]} + 1 + \frac{\beta_2}{\beta_1}\frac{1}{[H^+]} + \frac{\beta_3}{\beta_1}\frac{[HL]}{[H^+]^2}}$$

$$[ML] = \alpha_3 \cdot M_{tot} = \frac{M_{tot}}{\frac{1}{\beta_1}\frac{[H^+]}{[HL^-]} + \frac{\beta_1}{\beta_2}[H^+] + 1 + \frac{\beta_3}{\beta_2}\frac{[HL]}{[H^+]}}$$

$$[ML_2^{2-}] = \alpha_4 \cdot M_{tot} = \frac{M_{tot}}{\frac{1}{\beta_3}\frac{[H^+]^2}{[HL^-]^2} + \frac{\beta_1}{\beta_3}\frac{[H^+]^2}{[HL^-]} + \frac{\beta_2}{\beta_3}\frac{[HL]}{[H^+]} + 1} \quad (4.5)$$

where c_1, c_2, c_3 and c_4 are row vectors of the concentration matrix **C**.

For the metal-ligand complex, twenty solutions of metal (M^{2+}) and ligand (HL^-) with the constant total concentration of metal of 0.0020 M and ligand of 0.1000 M, and pH varying from 1 to 12 at 0.5 intervals were simulated. The spectra of the twenty solutions were generated from 200 to 600 nm at 1 nm intervals. The dimension of matrix is shown in Table 4.2.

Table 4.2 Dimensions of each matrix for metal-ligand equilibria

Cases of Equilibria	Data matrix (**D**)	Dimensions Matrix of Absorptivities (**A**)	Dimensions Matrix of Concentration (**C**)
Metal – ligand complexation	(400 x 23)	(400 x 4)	(4 x 23)

The molar absorptivity profiles, **A**, and the concentration profiles, **C**, were shown in Figure 4.4 and Figure 4.5, respectively.

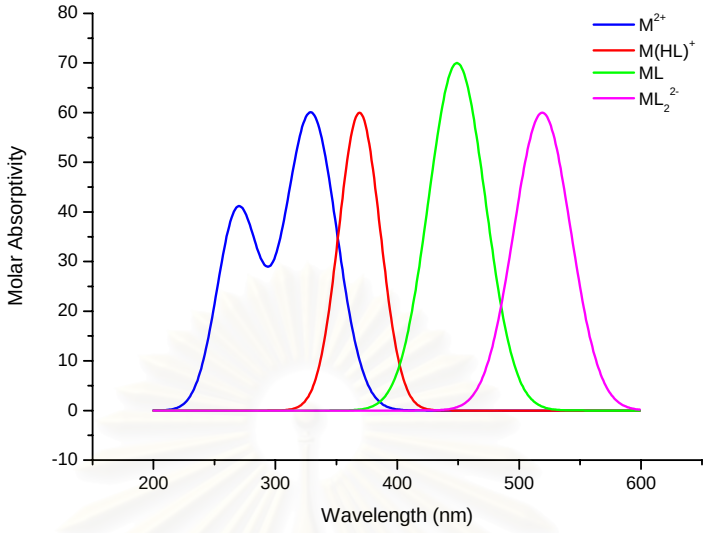


Figure 4.4 Simulated molar absorptivity spectra of 4 components for the metal-ligand complexation equilibria

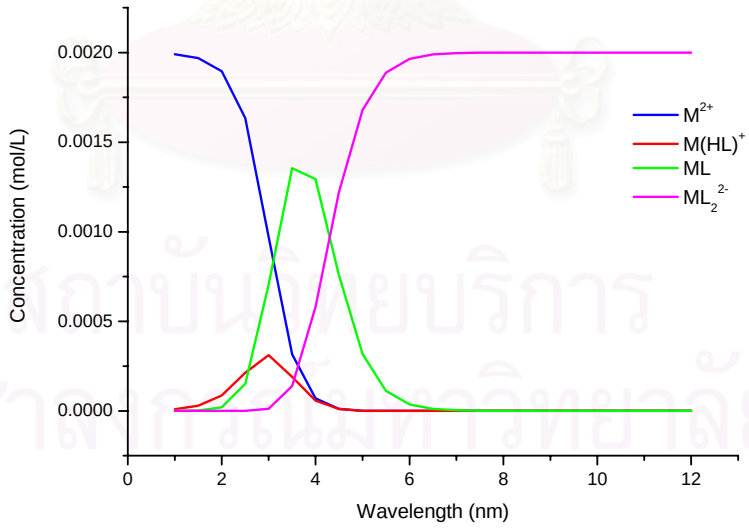


Figure 4.5 Simulated concentration profiles of 4 components of metal-ligand complexation system computed by equation 4.5

According to equation (4.4), the elements of error matrix ($\mathbf{E_{ij}}$) are generated using the pseudo-random number. Similarly, the standard deviation in absorbance was estimated to vary at 0.0005 absorbance unit. Thus, level of error of ±0.0005 was introduced in the simulated spectrum. The simulated spectra of the metal-ligand complexation were shown in Figure 4.6.

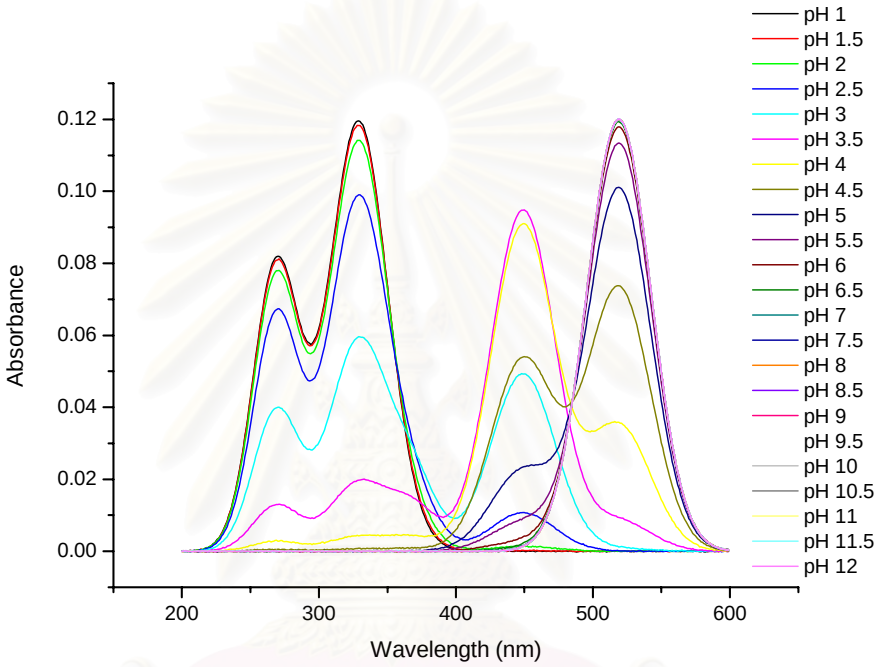


Figure 4.6 Twenty simulated spectra of metal-ligand complexation with added ±0.0005 errors

4.2.3) Simulated system of chromatographic separation.

The chromatograms and spectra were generated with simulated three-component mixtures with overlapping chromatography peaks. In order to simulate the overlapping chromatograms, the following equation was adopted.

$$h = \sum_{j=1}^{n} h_{0,j} \, exp\left(-\frac{\left(t - t_{0,j}\right)^2}{2\sigma_j^2} \right)$$

$$h = \sum_{j=1}^{n} h_{0,j} \, exp\left(-4 ln(2) \left(\frac{t - t_{0,j}}{W_{1/2,j}} \right)^2 \right) \qquad (4.6)$$

Where h and t are, respectively, the intensity and the retention time of the simulated chromatogram. n is the component number and $h_{0,j}$, $t_{0,j}$ and $W_{1/2,j}$ are the intensity, position and width at half height of component j, respectively. The resolution of the adjacent peaks is evaluated by

$$R_s = \frac{t_{0,j+1} - t_{0,j}}{W_{1/2,j+1} + W_{1/2,j}} \qquad (4.7)$$

The denominator in equation (4.7) is the average of the baseline widths, and the numerator is the separation of the peaks. The parameter R_s provides a quantitative measure of how much mixing of materials between two adjacent bands. At R_s = 0.5, the bands are quite mixed. At R_s = 1.0, the points of two triangles that approximate the peaks just touch where they meet at the baseline. At R_s = 1.5, the overlap between the actual peak is about 0.1% [14].

The chromatogram profile was generated at a chromatographic resolution of 0.40 and equal peak heights for all three peaks at the elution time 23, 32, and 41 min., respectively. Elution time is varying from 1 to 61 min. at the intervals 1 min. These peaks were recorded form 220 to 320 nm at 1 nm intervals. The dimension of matrix is shown in Table 4.3

Table 4.3 Dimensions of each matrix for chromatographic separation system

Cases	Data matrix (**D**)	Dimensions Matrix of Absorptivities (**A**)	Dimensions Matrix of Concentration (**C**)
Chromatography	(101 x 60)	(101 x 3)	(3 x 60)

The molar absorptivity profiles, **A**, and the concentration profiles, **C**, were shown in Figure 4.7 and Figure 4.8, respectively.

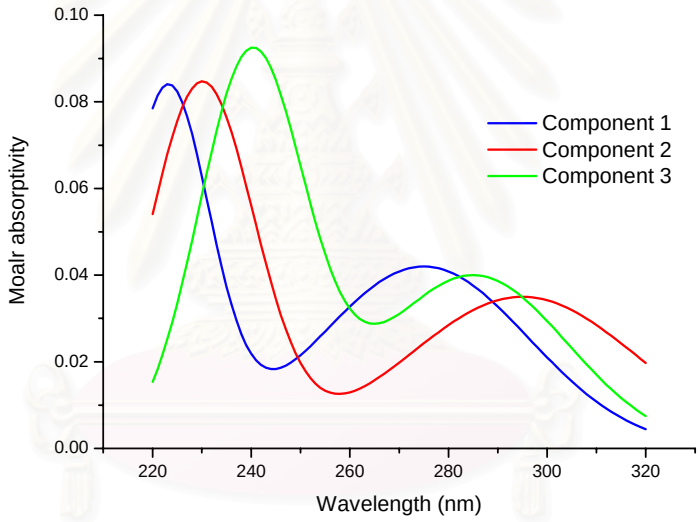


Figure 4.7 Simulated molar absorptivity spectra of 3 components

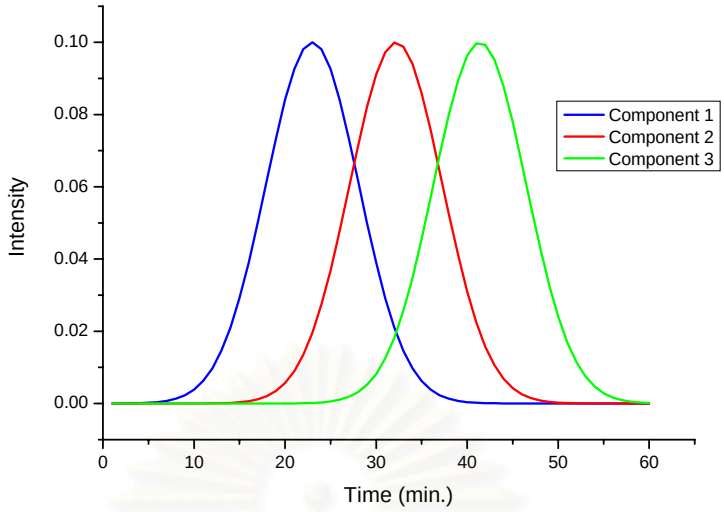


Figure 4.8 Simulated chromatogram of 3 components from chromatographic separation computed by equation (4.6) and (4.7).

According to equation (4.4), the elements of error matrix ($\mathbf{E_{ij}}$) are generated using the pseudo-random number. The standard deviation in absorbance was estimated to vary at 0.0005 absorbance unit. Thus, level of error of ±0.0005 was introduced in the simulated spectra. The simulated spectra of the chromatographic separation system were shown in Figure 4.9.

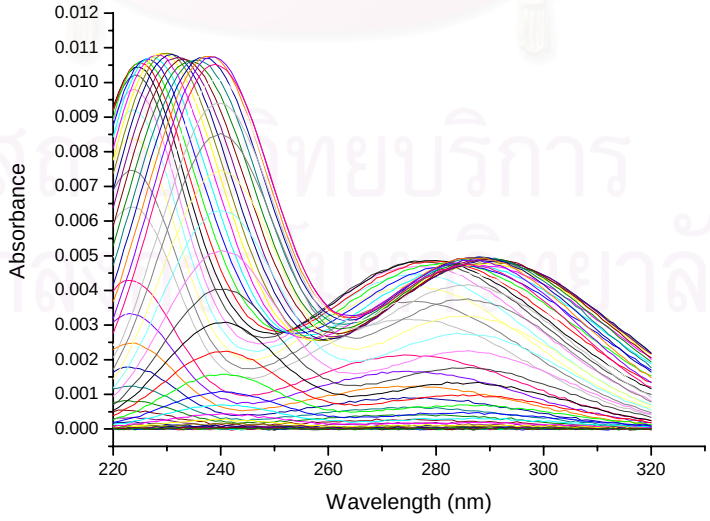


Figure 4.9 The simulated spectra of chromatographic separation with added ±0.0005 errors

4.3 Experimental spectra

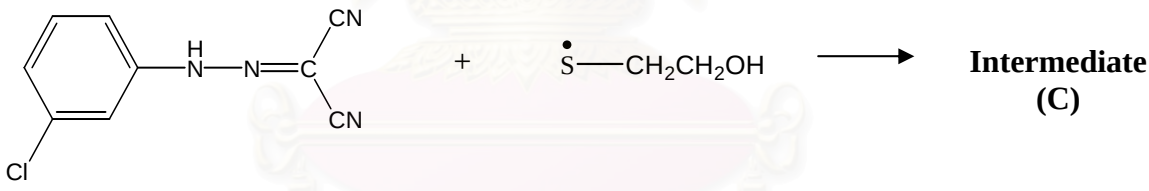
The SMCR program was tested with 2 UV-visible measurements.

1. UV-vis spectra of a two-step reaction.
2. UV-vis spectra of Copper-Glycine and Copper-Alanine Complexation.

4.3.1 Spectra of two-step reaction [15]

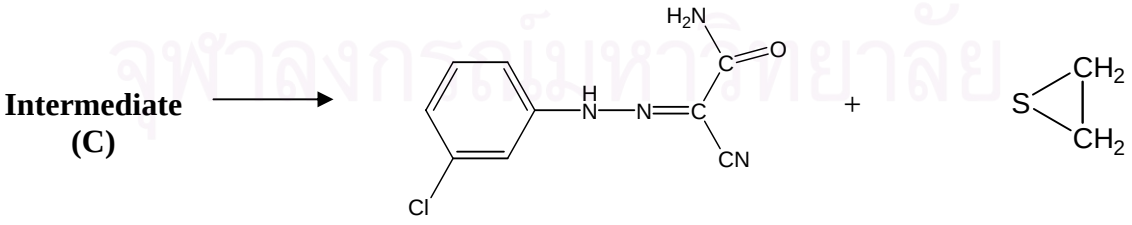
The UV-visible spectra of the two-step consecutive reaction of 3-chlorophenyl-hydarzonopropane dinitril (A), an uncouple of oxidative phosphorylation in cells, with 2-meracaptoethanol (B) described by Bijlsma, Louwerse and Smilde [15] was used. The two chemicals form an intermediate adduct (C) which is then hydrolysed to 3-chlorophenyl hydrazonocyanoacetamide (D) and ethylene sulphide (E) as by product. A proposed reaction mechanism is given below.

Reaction I



(A) **(B)**

Reaction II



(D) **(E)**

The spectra at several time intervals of spectroscopic active species A, C and D are monitored. The most important experimental conditions are given in Table 4.1. The 3-chlorophenyl-hydarzonopropane dinitril (A) is the determinate reactant with the initial concentration of 51.71 μmol/L.

Table 4.1 The important experimental conditions for two-step reaction system

Reaction Temperature	25 ^{0}C
Integration Time	1 second
Sampling Time	10 seconds
Total run time	2700 seconds
Wavelength range	250 – 500 nm 200 – 600 nm
Wavelength interval	1 nm
Number of recorded spectra	271

The UV-vis spectra of this system were recorded using a Hewlett Packard 8453 UV-vis spectrophotometer with diode array detection. A quartz cuvette with 1.00 cm. path length was used for obtaining spectra of the reaction mixture. A Pt-100 and a constant-temperature bath (Neslab) were used for the temperature control. The study was performed at two ranges of wavelength. i.e.250 – 500 nm., and 200 – 600 nm. The spectra were illustrated in the figure 4.10 and 4.11 respectively.

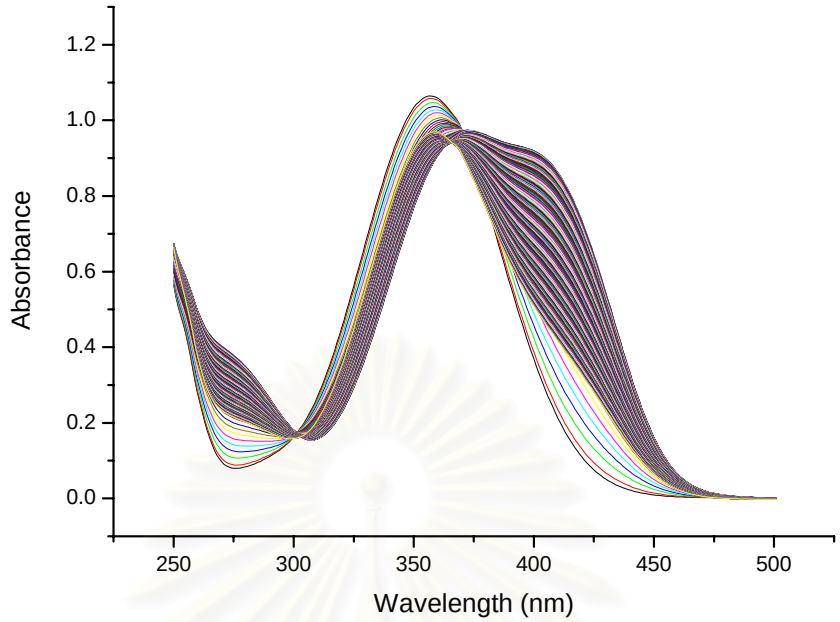


Figure 4.10 The UV-VIS spectra of the two-step reaction in wavelength 250 – 500 nm at intervals 1 nm and time varying from 0 to 2710 seconds at intervals 10 seconds.

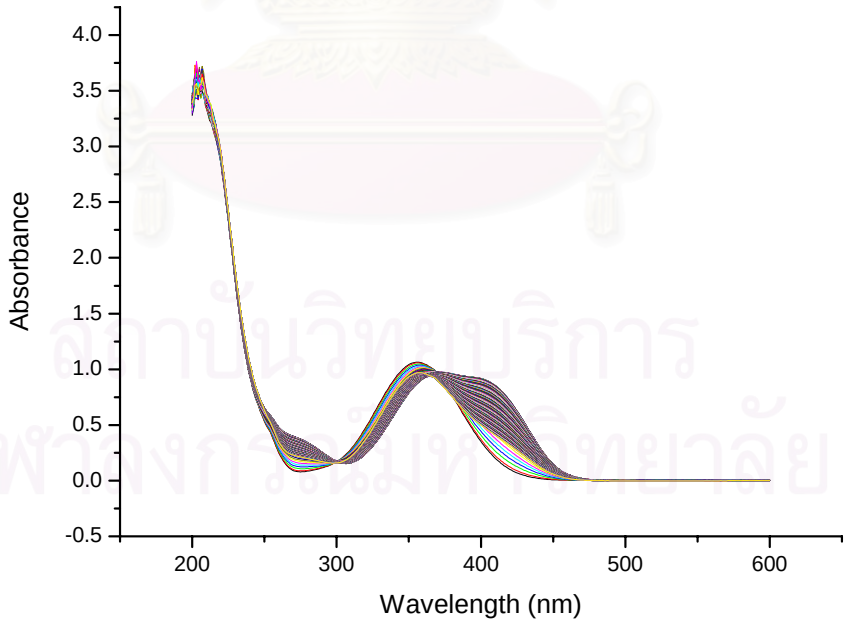


Figure 4.11 The UV-VIS spectra of the two-step reaction in wavelength 200 – 600 nm at intervals 1 nm and time varying from 0 to 2710 seconds at intervals 10 seconds.

4.3.2 Spectra of Copper-Glycine and Copper-Alanine Complexation [16]

Glycine, H_2NCH_2COOH, and Alanine, $H_2NCH(CH_3)COOH$, are amino acids. Thus, they are a zwitter-ions molecules, containing an amino group and a carboxylic acid group, and exhibit properties of both acid and base. Both glycine and alanine are known to complex with copper (II) forming different species. The UV-visible spectra of the copper-glycine and copper-alanine complexation in the different pHs described by Arunchai Tungcharoenbumrungsuk [16] was used. In the literature, Darj and Malinowski [17] used the window factor analysis (WFA) to evaluate the visible spectra of Cu(II) and glycine complexes, and expressed the complex formation as;

$$Cu^{2+} + GlyH \rightleftharpoons Cu(GlyH)^{2+}$$

$$Cu^{2+} + GlyH \rightleftharpoons Cu(GlyH)^{+} + H^{+}$$

$$Cu^{2+} + 2\,(GlyH) \rightleftharpoons Cu(Gly)_2 + 2H^{+}$$

From the experiment, 18 solutions of copper-glycine were prepared with pHs varying from 1 to 7. Each solution contained 0.002 M. copper (II) and 0.10 M glycine. The visible spectra solutions were recorded from 450 to 850 nm at 3-nm interval and shown in Figure 4.12, yielding (134 x 18) absorbance matrix.

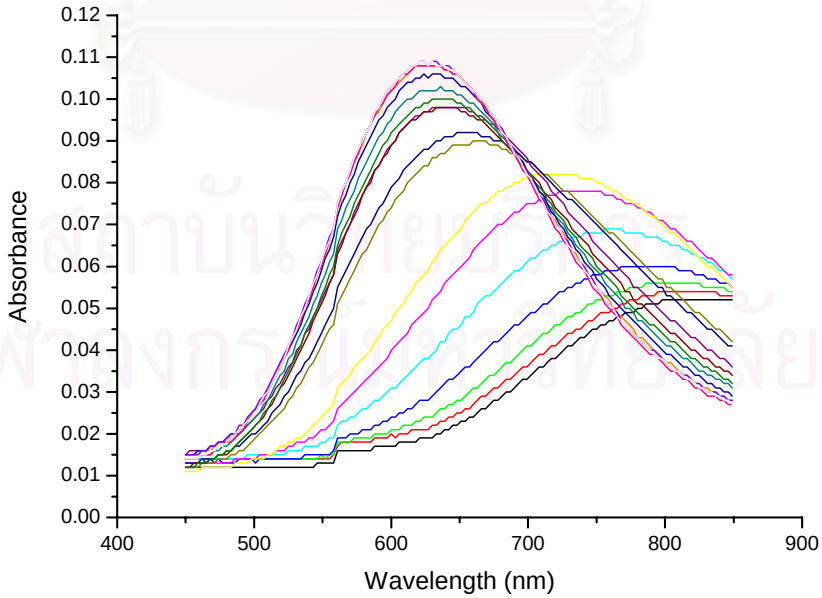


Figure 4.12 Visible spectra of 18 copper-glycine solutions with pH ranging from 1 to 7.

Analogous to the system of glycine, the complexe formation of Cu(II) and alanine ought to be;

$$Cu^{2+} + AlaH \rightleftharpoons Cu(AlaH)^{2+}$$

$$Cu^{2+} + AlaH \rightleftharpoons Cu(AlaH)^{+} + H^{+}$$

$$Cu^{2+} + 2\,(AlaH) \rightleftharpoons Cu(Ala)_2 + 2H^{+}$$

From the experiment, 13 solutions of copper-alanine were prepared with pHs varying from 1 to 7. Each solution contained 0.002 M copper(II) and 0.10 M alanine The visible spectra solutions were recorded from 450 to 850 nm at 2-nm interval and shown in Figure 4.13, yielding (201 x 13) absorbance matrix.

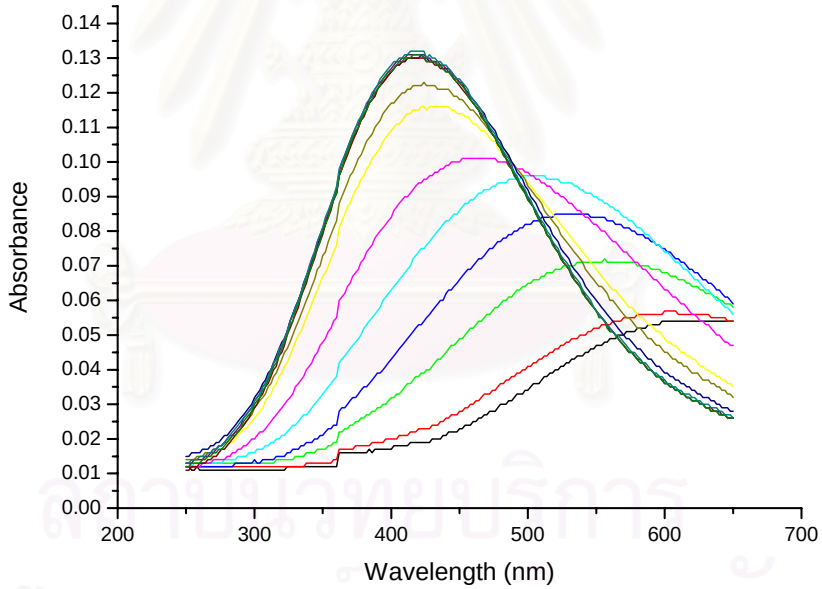


Figure 4.13 Visible spectra of 13 copper-alanine solutions with pH ranging from 1 to 7.

CHAPTER V

RESULTS AND DISCUSSION

5.1 Error estimation

The performance of the program and SMCR algorithm can be assessed by the error estimation. Followings are the error estimations used in the SMCR program.

$$\text{Norm of error} = \sqrt{\sum_{i=1}^{r}\sum_{j=1}^{c}\left(d_{ij}-\hat{d}_{ij}\right)^2} \quad (5.1)$$

$$\text{Root-mean-square (rms) error} = \sqrt{\frac{\sum_{i=1}^{r}\sum_{j=1}^{c}\left(d_{ij}-\hat{d}_{ij}\right)^2}{r\ x\ c}} \quad (5.2)$$

$$\text{Normalized root-mean-square (nrms) error} = 100\ \text{x}\ \sqrt{\sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{d_{ij}-\hat{d}_{ij}}{d_{ij}}\right)^2} \quad (5.3)$$

where d_{ij} and $\hat{d}_{ij}$ are the respective experimental or simulated and predicted data. The r and c are the number of row and column of the data matrix, respectively.

Norm of error expresses the whole error of data. The rms error expresses the average error of the data. The nrms expresses the percentage of relative rms error.

5.2 H_3A simulated spectra.

The simulated absorbance matrix of acid-base equilibria was prepared in text format (*.txt), and then they were input to the program SMCR. The absorbance data was resolving using different chemical rank analysis methods (OPA, SIMPLISMA and EFA) and their error estimation were given in Table 5.1, 5.2, and 5.3, respectively. Comparison between simulated and calculation spectra were illustrated in Figure 5.1. Furthermore, comparison between presetting and resolved of acid dissociation constant (pKa) was given in Table 5.4.

Table 5.1 Prediction of number of components and error estimation in concentration profiles of H_3A dissociation resolved by OPA, SIMPLISMA and EFA methods.

Method	**Number of Components**	**Error of Prediction**		
		Concentration profiles		
		Norm	**rms**	**Nrms**
OPA	**4**	0.0083	9.127×10^{-4}	2.07
SIMPLISMA	**4**	0.0083	9.118×10^{-4}	2.07
EFA	**4**	0.0083	9.117×10^{-4}	2.06

Table 5.2 Prediction of number of components and error estimation in absorptivity profiles of H_3A dissociation resolved by OPA, SIMPLISMA and EFA methods.

Method	**Number of Components**	**Error of Prediction**		
		absorptivity profiles		
		Norm	**rms**	**Nrms**
OPA	**4**	0.6035	0.0151	3.29
SIMPLISMA	**4**	0.6029	0.0151	3.29
EFA	**4**	0.6028	0.0151	3.29

Table 5.3 Prediction of number of components and error estimation in spectra of H_3A dissociation resolved by OPA, SIMPLISMA and EFA methods.

Method	Number of Components	Error of Prediction		
		Spectra profiles		
		Norm	rms	Nrms
OPA	**4**	0.0110	1.206×10^{-4}	0.27
SIMPLISMA	**4**	0.0110	1.205×10^{-4}	0.27
EFA	**4**	0.0110	1.205×10^{-4}	0.27

Table 5.4 Comparison between presetting pKa and resolved pKa of triprotic acid (H_3A)

Acid dissociation constant	Presetting value	Method		
		OPA	SIMPLISMA	EFA
pK_{a1}	**3.47**	3.38	3.38	3.38
pK_{a2}	**5.65**	5.62	5.62	5.62
pK_{a3}	**6.89**	6.98	6.98	6.98

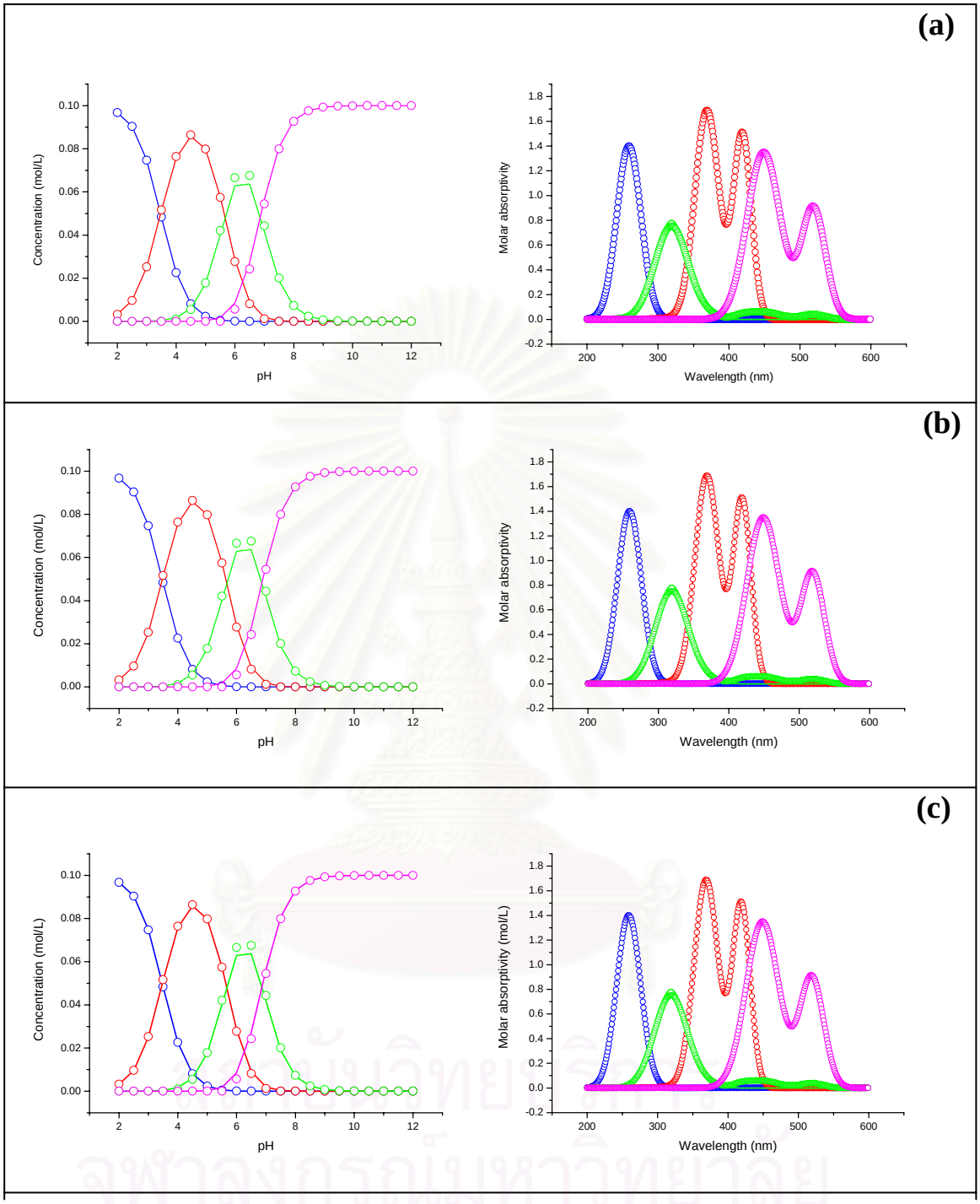


Figure 5.1 Comparison between simulated (solid lines) and predicted (open circles) concentration and absorptivity profiles of a 4 component mixture of H_3A dissociation system using different methods for chemical rank analysis **:(a)** OPA, **(b)** SIMPLISMA and **(c)** EFA methods.

From Table 5.1, 5.2, 5.3 and Figure 5.1, it can be seen that four components were explicitly resolved from the spectra which is consistent to the presetting condition. The simulated spectra were resolved into pure component spectra. It appears that there exists areas where no overlapping between component spectra. From Table 5.4, the resolved pKa from 3 methods is not different. In accordance with this observation, this is concerned that there is no difference in the efficiency of all 3 methods for resolving the spectra. The absorptivity profiles, concentration profiles and also resolve pKa agree very well with the preset values. The percentage of relative rms error (nrms) is 2.06% and 3.29% for concentration and absorptivity profiles, respectively. The program can resolved the simulated four component acid dissociation with high accuracy.

5.3 Metal-Ligand simulated spectra.

The simulated absorbance matrix of metal-ligand equilibria was prepared in text format (*.txt), and then they were input to the program SMCR. The absorbance data were resolved using different chemical rank analysis methods (OPA, SIMPLISMA and EFA), and the error estimations of this system were given in Table 5.5, 5.6, and 5.7, respectively. Comparison between simulated spectra and calculation spectra were illustrated in Figure 5.2. Furthermore, comparison between presetting and resolved of stability constants of complexation (β) was given in Table 5.8.

Table 5.5 Prediction of number of components and error estimation in concentration profiles of M-HL complexation resolved by OPA, SIMPLISMA and EFA methods.

Method	**Number of Components**	**Error of Prediction**		
		Concentration profiles		
		Norm	**rms**	**Nrms**
OPA	**4**	4.03×10^{-4}	4.20×10^{-5}	4.50
SIMPLISMA	**4**	4.30×10^{-4}	4.20×10^{-5}	4.50
EFA	**4**	1.38×10^{-4}	1.45×10^{-5}	**1.55**

Table 5.6 Prediction of number of components and error estimation in absorptivity profiles of M-HL complexation resolved by OPA, SIMPLISMA and EFA methods.

Method	**Number of Components**	**Error of Prediction**		
		Absorptiviy profiles		
		Norm	**rms**	**Nrms**
OPA	**4**	182	4.57	22.54
SIMPLISMA	**4**	182	4.57	22.54
EFA	**4**	88	2.21	**10.90**

Table 5.7 Prediction of number of components and error estimation in spectra of M-HL complexation resolved by OPA, SIMPLISMA and EFA methods.

Method	**Number of Components**	**Error of Prediction**		
		Spectra profiles		
		Norm	**rms**	**Nrms**
OPA	**4**	0.0071	7.48×10^{-5}	0.19
SIMPLISMA	**4**	0.0071	7.48×10^{-5}	0.20
EFA	**4**	0.0071	7.48×10^{-5}	0.20

Table 5.8 Comparison between presetting and resolved stability constants of complexation (β).

Acid dissociation constant	**Presetting value**	**Method**		
		OPA	**SIMPLISMA**	**EFA**
β_1	**7.59**	12.26	12.26	9.05
β_2	**5.39×10^{-2}**	3.85×10^{-2}	3.85×10^{-2}	2.70×10^{-2}
β_3	**3.37×10^{-5}**	1.98×10^{-5}	1.98×10^{-5}	2.41×10^{-5}

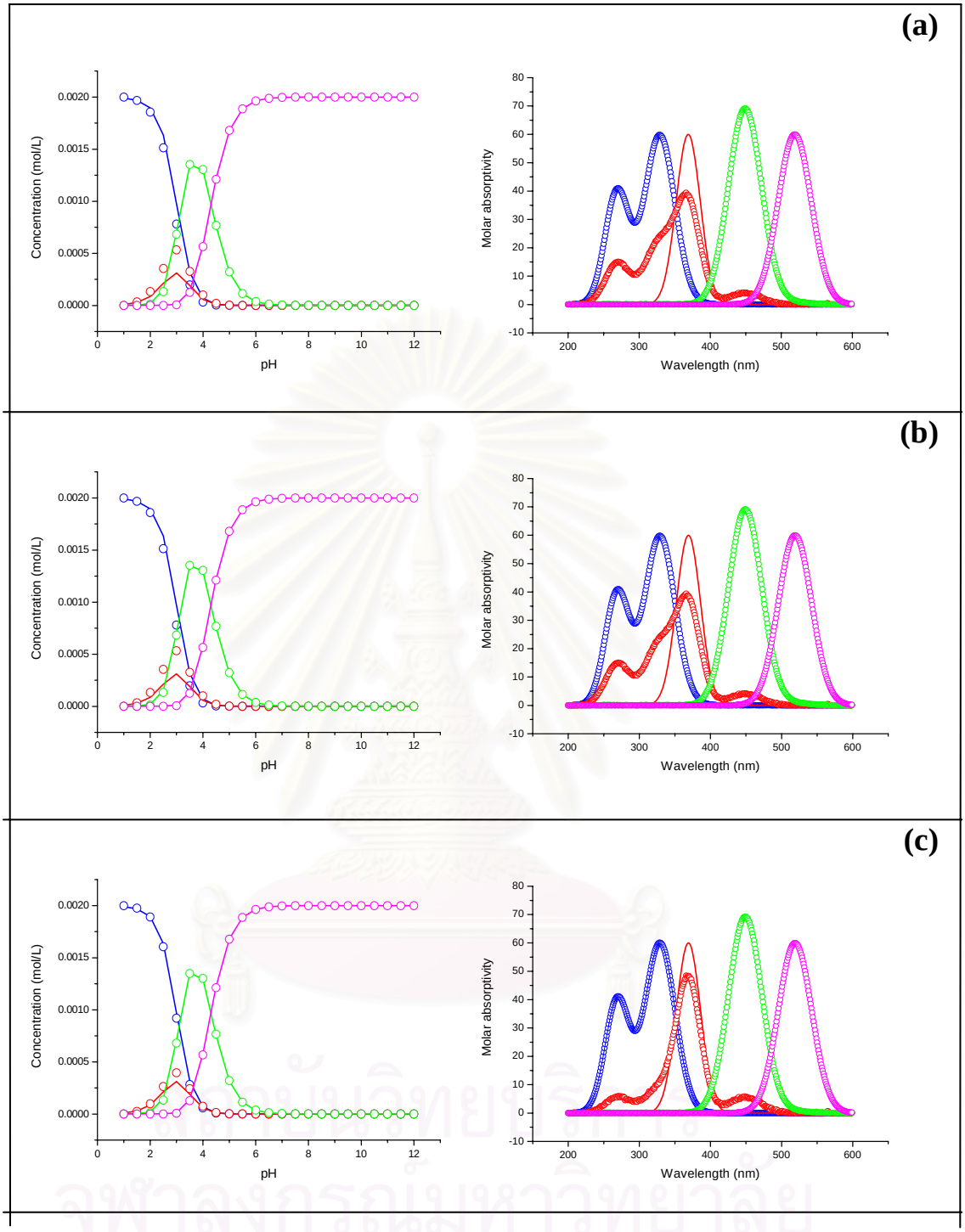


Figure 5.2 Comparison between simulated (solid lines) and predicted (open circles) concentration and absorptivity profiles of a 4 component mixture of M-HL complexation system using different methods for chemical rank analysis **:(a)** OPA, **(b)** SIMPLISMA and **(c)** EFA methods.

From Table 5.5, 5.6, 5.7 and Figure 5.2, it shown that all methods resolved for four components. From absorptivity profiles, pure component spectra were illustrated. Unlike those of the acid-base equilibria of H_3A acid, we could not find non-overlapping area for pure component in the simulated spectra. In accordance with the observation, it was very difficult to obtain resolute number of components and their corresponding absorptivity profile. From the error estimation, that the efficiency of all 3 methods in resolving the spectra is different. In absence of non-overlapping pure component spectra, the program based on EFA method seems to give the best results of 3 methods. From Table 5.8, the resolved stability constants of complexation from EFA method were also most closely to the presetting values. From Table 5.5 and 5.6, the percentage of relative rms error (nrms) from EFA method is 1.55 and 10.90 for concentration and absorptivity profiles, respectively. Interestingly, EFA generates concentration profiles first and then absorptivity, while OPA and SIMPLISMA both search for pure component spectra and absorptivity profiles first. It would imply that for system in absence of non-overlapping pure components the algorithm based on the elucidation of concentration profiles gives higher accuracy than there based on the resolution of pure component spectra.

5.4 Chromatographic simulated spectra.

The simulated absorbance matrix of the chromatographic system was prepared in text format (*.txt), and then they were input to the program SMCR. The spectra were resolved using different chemical rank analysis methods (OPA, SIMPLISMA and EFA), the testing results of this system were given in Table 5.9, 5.10, and 5.11, respectively and fitting results of simulated and calculation spectra were illustrated in Figure 5.3.

Table 5.9 Prediction of number of component and error estimation in concentration profiles of chromatographic system resolved by OPA, SIMPLISMA and EFA methods.

Method	**Number of Components**	**Error of Prediction**		
		Concentration profiles		
		Norm	**rms**	**Nrms**
OPA	**3**	0.0288	0.0021	**5.52**
SIMPLISMA	**3**	0.1067	0.0079	20.48
EFA	**3**	0.0411	0.0031	6.89

Table 5.10 Prediction of number of components and error estimation in absorptivity profiles of chromatographic system resolved by OPA, SIMPLISMA and EFA methods.

Method	Number of Components	Error of Prediction		
		Absorptivity profiles		
		Norm	rms	Nrms
OPA	**3**	0.0485	0.0028	**6.53**
SIMPLISMA	**3**	0.2665	0.0153	35.93
EFA	**3**	0.0766	0.0044	10.33

Table 5.11 Prediction of number of components and error estimation in spectra of chromatographic system resolved by OPA, SIMPLISMA and EFA methods.

Method	Number of Components	Error of Prediction		
		Spectra profiles		
		Norm	rms	Nrms
OPA	**3**	0.0110	1.477×10^{-5}	0.42
SIMPLISMA	**3**	0.0110	1.342×10^{-5}	0.38
EFA	**3**	0.0110	1.444×10^{-5}	0.41

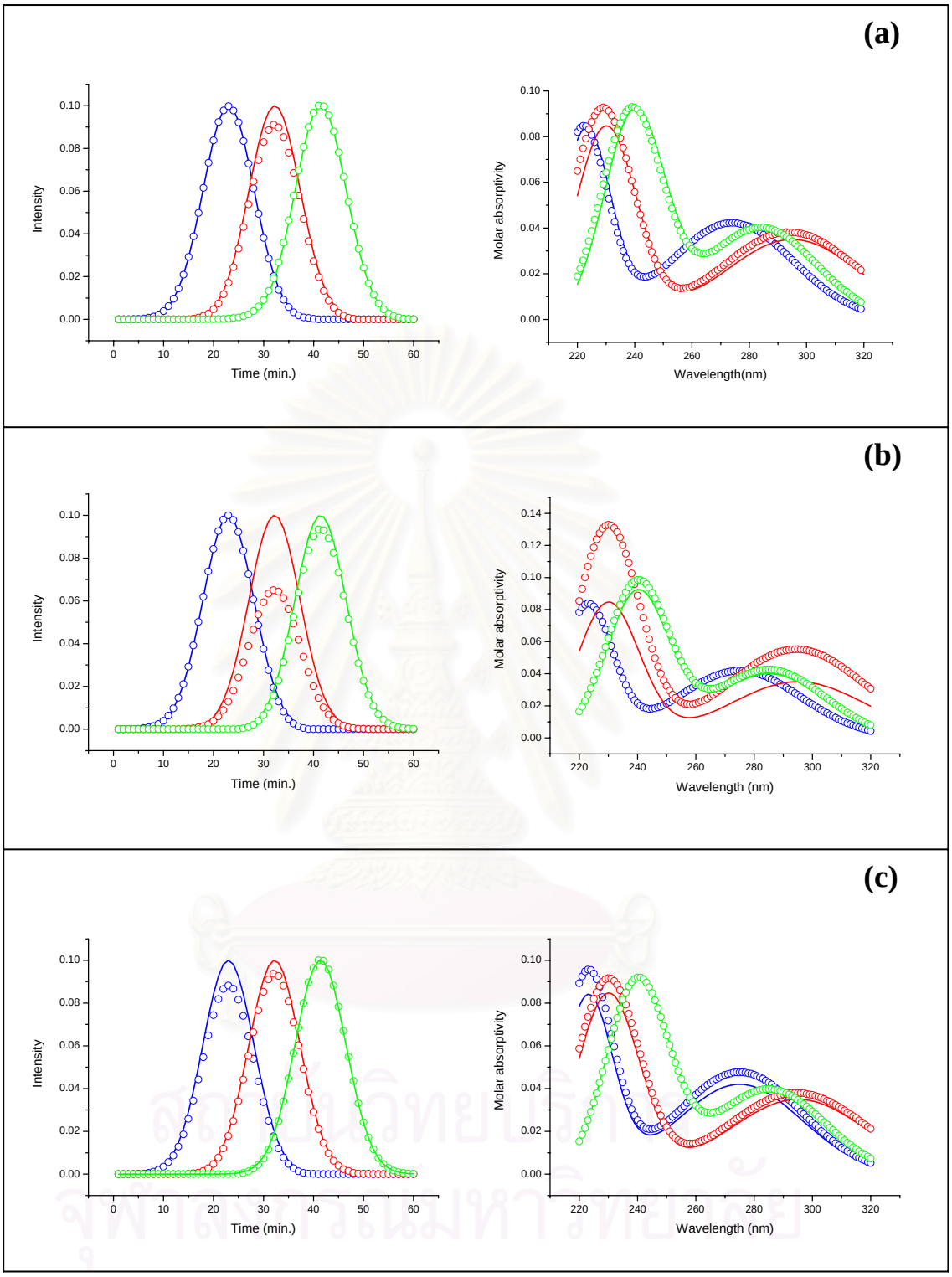


Figure 5.3 Comparison between simulated (solid lines) and predicted (circles) concentration and absorptivity profiles of a 3 component mixture of chromatographic separation system using different methods for chemical rank analysis; **(a)** OPA, **(b)** SIMPLISMA and **(c)** EFA methods

Figure 5.3 illustrates pure component spectra and concentration profiles and three components were explicitly resolved. It can also be seen that there exists where non-overlapping region between component spectrums. In this system, their components were separated by chromatographic techniques prior to absorbance detection. It differs from acid-base and metal-ligand complexation equilibria in which absorbance of all components were detected in the same time.

From the algorithm, both the OPA and EFA methods resolved component independently and therefore are suitable for chromatographic system. The OPA and EFA methods predicted the spectra of purest component at the elution time 24, 31, 41 and 22, 32, 43, respectively, in consistence to the presetting value (23, 31 and 43). The SIMPLISMA method is very sensitive to noise, if the noise level is high as it often selects component from noise. It predicted profiles of component with using correlation factor in the equation 2.9 and 2.15. Thus, this method searches the pure component spectra dependently with other components. The elution time of each purest component spectra was resolved at 19, 32 and 47, not in agreement with the presetting elution time. The effect of noise could be reduced either by reducing the objects dimension or by increasing the offset value following eq. 2.16 [20].

In the case of chromatographic separation technique, the program based on OPA method seems to yield the best result of the three methods. From Table 5.9 and 5.10, the percentages of relative rms errors (nrms) from OPA method are 5.52 and 6.53 for concentration and absorptivity profiles, respectively.

5.5 Spectra of two-step reaction system

The SMCR program was applied to the two-step reaction of 3-chlorophenyl-hydarzonopropane dinitril (A) and with 2-meracaptoethanol (B).

Determination of the number of species

For the spectra recorded between 250 – 500 nm and 200 – 600 nm, the output results were considered together. The significant components were determined by chemical rank analysis methods (OPA, SIMPLISMA and EFA). Three methods resolved "3" primary component which corresponding to 3 absorbing species in the system. Thus, the absorbance spectra was contributed by 3-chlorophenyl-hydarzonopropane dinitril(A), intermediate (C), and 3-chlorophenyl hydrazonocyanoacetamide (D).

After determining number of significant components, the initial guess of concentration and absorptivity profiles were resolved with MCR algorithm. The resolved concentration profiles were constrained to be nonnegative and unimodality. The total concentration of reactant (set to 51.71 μmol/L) remains constant along the whole two-step reaction process. The condition of a closed system was also applied as a closure constraint in the optimization. In contrast, spectra in absorptivity profiles were constrained to be only nonnegative. The spectra recorded between 250 – 500 nm. (251 x 271) and 200 – 600 nm. (401 x 271) for two-step reaction were resolved and their error estimation of concentration profiles were given in Table 5.12 and 5.13. The resolved concentration and absorptivity profiles from spectra recorded between 250 – 500 nm and 200 – 600 nm for two-step reaction are shown respectively in Figure 5.4 and 5.5.

Table 5.12 Prediction of number of components and error estimation in concentration profiles of two-step reaction recorded between 250 – 500 nm. resolved by OPA, SIMPLISMA and EFA methods.

Method	Number of Components	Error of Prediction		
		Concentration profiles		
		Norm	rms	Nrms
OPA	**3**	2.45×10^{-4}	8.60×10^{-6}	38.64
SIMPLISMA	**3**	2.45×10^{-4}	8.60×10^{-6}	38.64
EFA	**3**	2.42×10^{-4}	8.56×10^{-6}	38.60

Table 5.13 Prediction of number of components and error estimation in concentration profiles of two-step reaction recorded between 200 – 600 nm. resolved by OPA, SIMPLISMA and EFA methods.

Method	Number of Components	Error of Prediction		
		Concentration profiles		
		Norm	rms	Nrms
OPA	**3**	2.02×10^{-4}	7.07×10^{-6}	30.78
SIMPLISMA	**3**	2.11×10^{-4}	7.42×10^{-6}	33.37
EFA	**3**	**1.47×10^{-4}**	**4.89×10^{-6}**	**20.37**

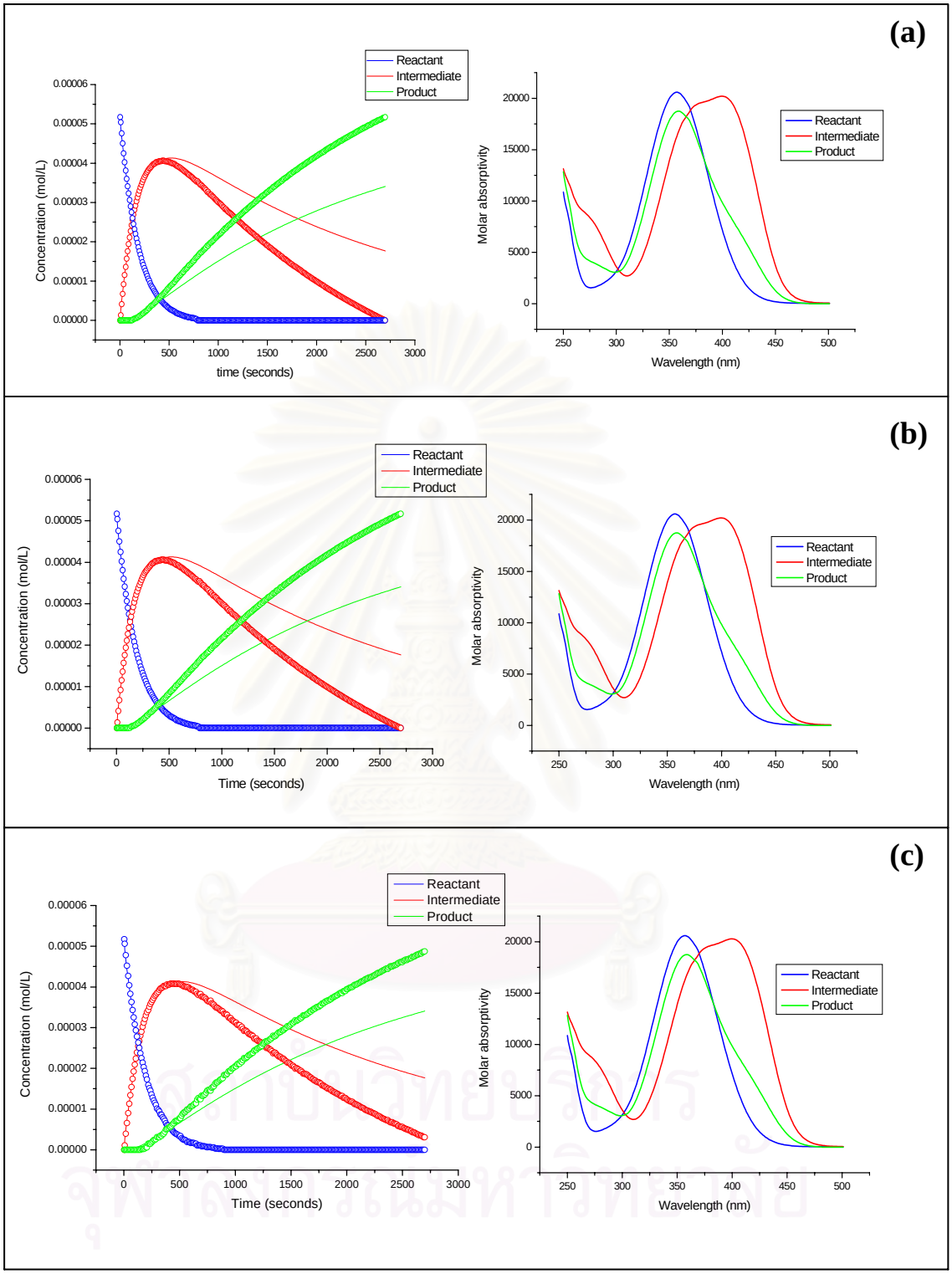


Figure 5.4 Comparison between literature data (solid line) and calculated data (circles) of concentration profiles and the resolved absorptivity profiles of spectra recorded at 250 – 500 nm were obtained by using different method for chemical rank analysis: **(a)** OPA, **(b)** SIMPLISMA and **(c)** EFA.

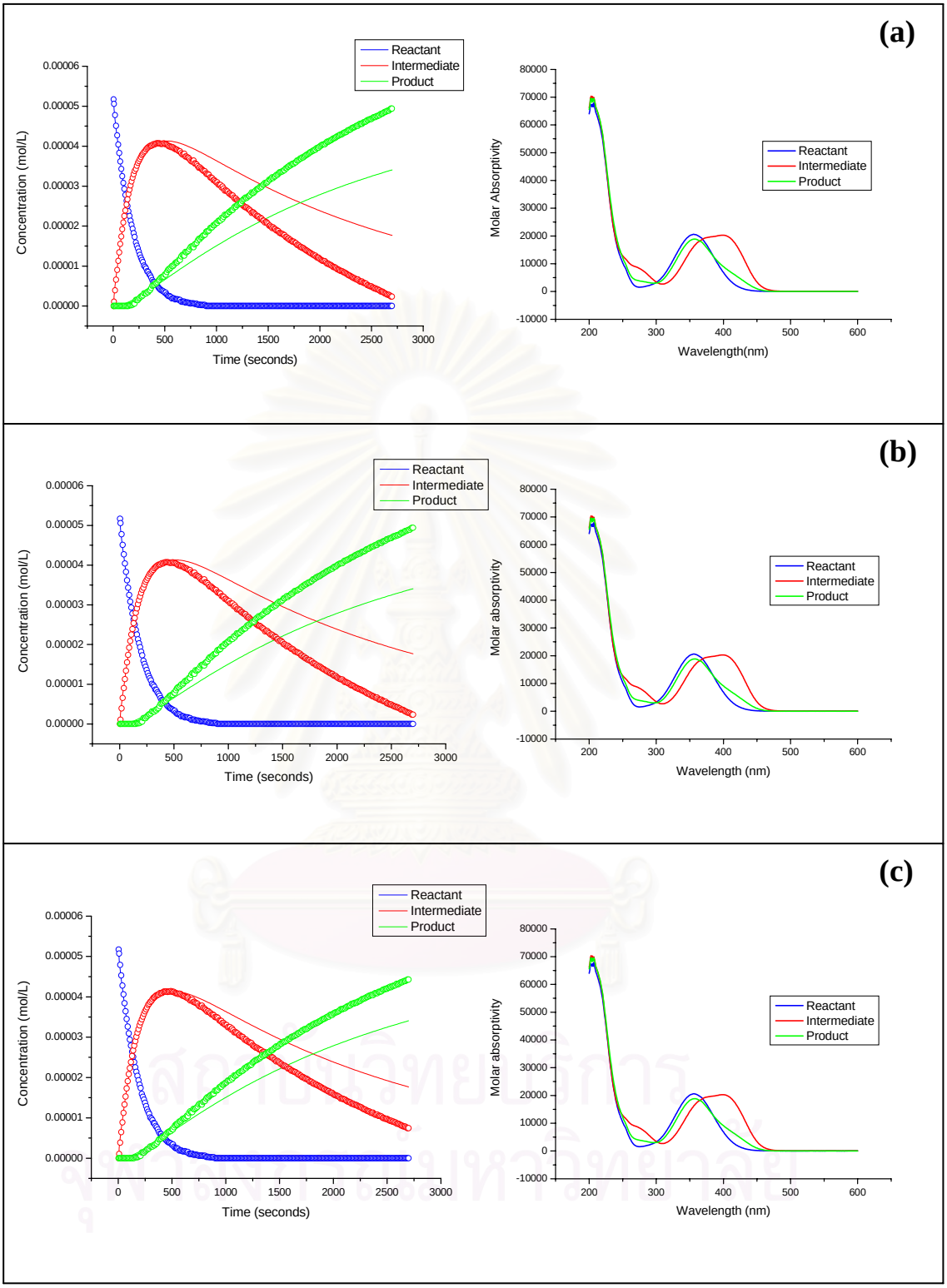


Figure 5.5 Comparison between literature data (solid line) and calculated data (circles) of concentration profiles and resolved absorptivity profiles of spectra recorded at 200 – 600 nm. were obtained by using different method for chemical rank analysis: **(a)** OPA, **(b)** SIMPLISMA and **(c)** EFA.

Form Table 5.12 and Figure 5.4, the resolved pure component spectra and concentration profile of spectra recorded at 250-500 nm. were illustrated. The error estimations in concentration profiles obtained form three methods were almost identical. From the observation, we notice that the resolved concentration profile in the terminal time (time = 2700 s.) consists only product species which are contradict to the previous study that showed both product and intermediate species at the terminal time.

According to Table 5.13 and Figure 5.5, the pure component spectra and concentration profile of spectra recorded 200 – 600 nm. were predicted. They have larger data matrix (401 x 271) than the absorbance data recorded from 250 – 500 nm. (251 x 271). Interestingly, the error estimation in concentration profiles from three methods is different. Only the program based on EFA method can predict the concentration profile in good agreement with the previous study. The results consents with the results in section 5.2 and 5.3 concluded that the EFA method is the most sensitivity method of three methods. The percentage of relative rms error (nrms) from EFA method in concentration profile is 20.37. It can be concluded that the size of data matrix is the important factors for resolving by the SMCR program.

5.6 Spectra of Copper-Glycine and Copper-Alanine complexation

We consider the output results of both systems of copper-glycine and copper-alanine together. The results from both systems possess a similar trend in the properties.

- Determination of the Number of Copper species. For both systems of copper glycine and copper-alanine, three methods, i.e. OPA SIMPLISMA and EFA, for determination of chemical ranks were used. The OPA and SIMPLISMA methods given the consensus of "3" primary components while the EFA method predicted "4" primary components. From the literature [15], the absorption spectra of copper-glycine system is consisted of 4 species; Cu^{2+}, $Cu(GlyH)^{2+}$, $Cu(Gly)^{+}$ and $Cu(Gly)_2$ and the absorption spectra of copper-alanine system is consisted of 4 species; Cu^{2+}, $Cu(AlaH)^{2+}$, $Cu(Ala)^{+}$ and $Cu(Ala)_2$. It should be noted that the fourth component has very small amount with respect to the first three components in the equilibria.

From the results, only EFA method can determine the number of primary component correctly. It can be concluded that EFA method that generates concentration profiles first is the most sensitivity of three methods. The result is in good agreement with the results from validation section.

After determining the number of significant component, the concentration and absorptivity profiles were resolved with MCR algorithm. The resolved concentration profiles were constrained to be nonnegative by nonnegative least square (nnls) algorithm and unimodality. The total concentration of reactant (set to 0.002 mol/L) remains constant along the whole two-step reaction process, the condition of a closed system was also applied as a closure constraint in the optimization. In contrast, spectra in absorptivity profiles were constrained to be only nonnegative. Resolved concentration and absorptivity profiles from spectra data for copper-glycine and copper-alanine complexation system are shown respectively in Figure 5.6 and 5.7.

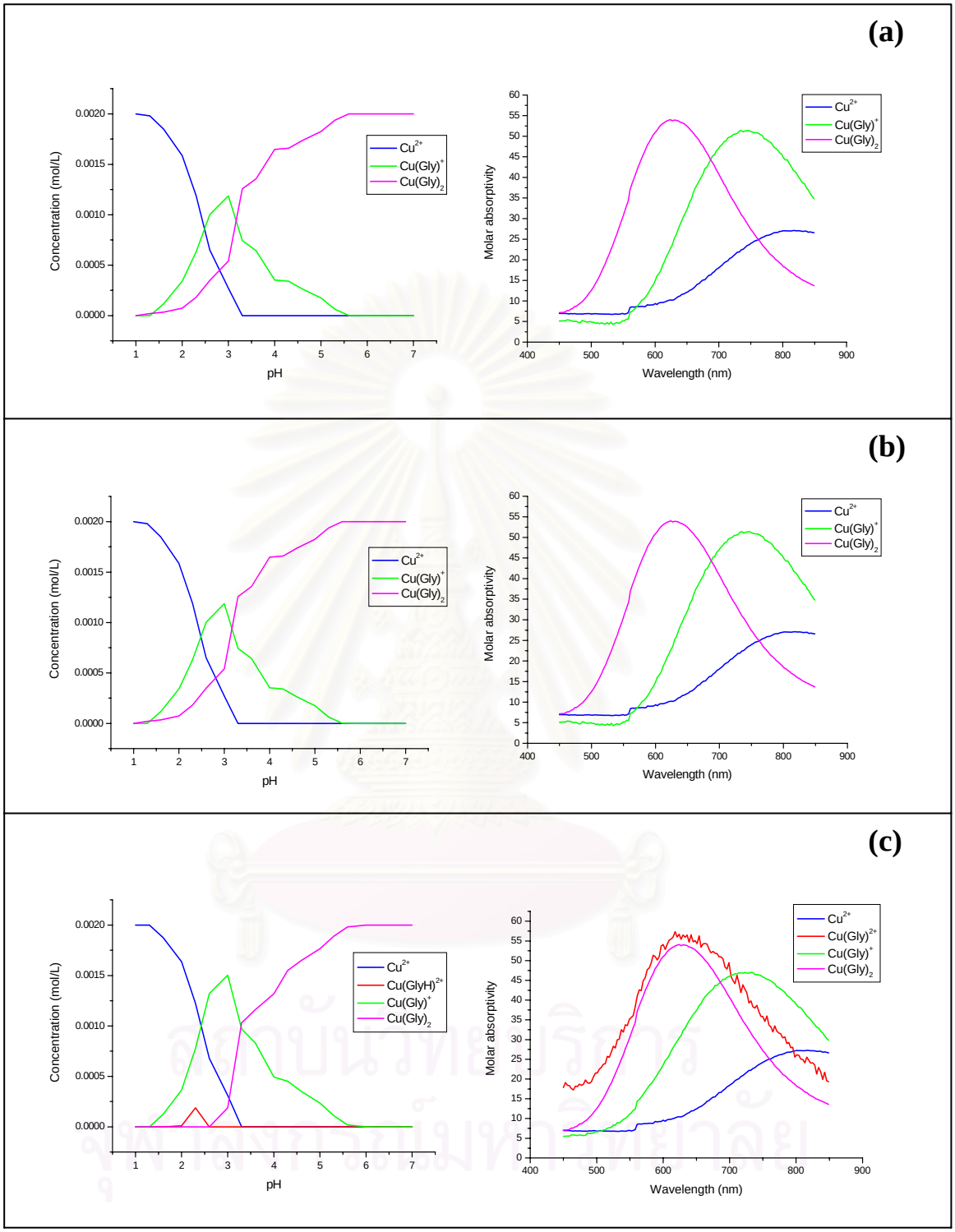


Figure 5.6 The resolved concentration and corresponding absorptivity profiles of copper-glycine complexation equilibria were obtained by using different method for chemical rank analysis: **(a)** OPA, **(b)** SIMPLISMA and **(c)** EFA.

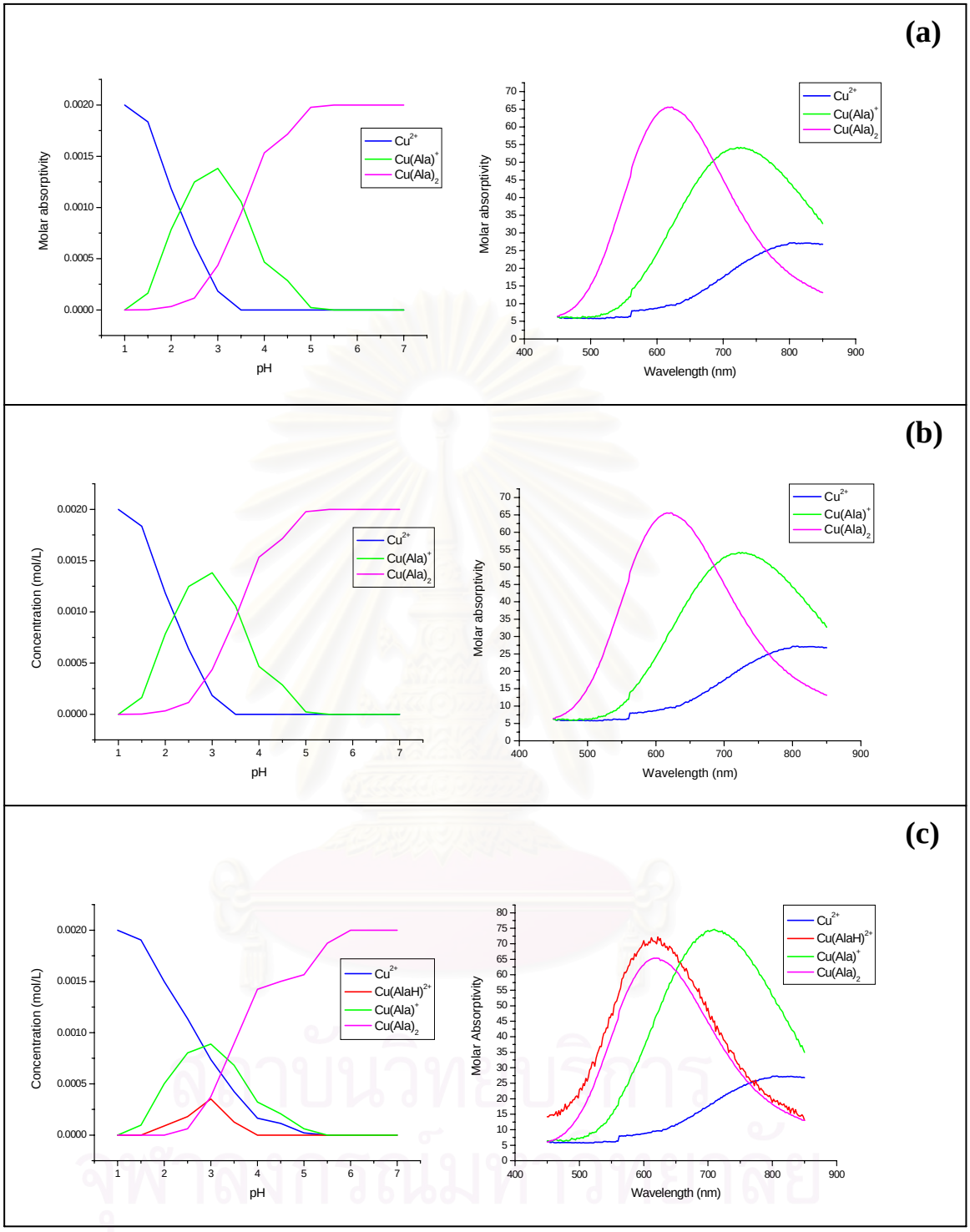


Figure 5.7 The resolved concentration and corresponding absorptivity profiles of copper-alanine complexation equilibria were obtained by using different method for chemical rank analysis: **(a)** OPA, **(b)** SIMPLISMA and **(c)** EFA.

From Figure 5.6 and 5.7, concentrations of $Cu(GlyH)^{2+}$ and $Cu(AlaH)^{2+}$ are smaller than other copper forms and exist in a narrow pH region between pH 2.5 – 3.5. It should be noted that the concentration of $Cu(GlyH)^{2+}$ was overlapped by those of $Cu(Gly)^{+}$, $Cu(Gly)_2$ and Cu^{2+} . Analogous to the copper-alanine, the concentration of $Cu(AlaH)^{2+}$ was overlapped by those of $Cu(Ala)^{+}$, $Cu(Ala)_2$ and Cu^{2+}. The system of copper-glycine and copper-alanine is look like the simulated metal-ligand complexation system in the section 5.3. It was found that resolution for number of components and absorptivity profile was very difficult to obtain accurately. Most of errors were accumulated in the molar absorptivity profile especially the species which contained small amount such as $Cu(GlyH)^{2+}$ and $Cu(AlaH)^{2+}$. Only EFA method can predict the number of components correctly while OPA and SIMPLISMA can not predict the $Cu(GlyH)^{2+}$ or $Cu(AlaH)^{2+}$ species.

In the simulated M-HL complexation system in section 5.3, the spectra data was collected in the wide range of wavelength. This makes three methods for chemical rank analysis determine the number of component accurately. But for the copper-glycine and copper-alanine complexation, the spectra data was collected only from 450 – 850 nm thus some important informations of the system were excluded. This is probably a main reason for the failure of the program in predicting the results with high accuracy. The reason is in good agreement in the two-step reaction system in section 5.5. The more size of data matrix would make the program resolve with high accuracy. In the another system, we used to apply the program to resolve other spectra of copper-glycine complexation produced by Irving and Pettit [22]. The program can resolve the spectra with high accuracy. The error of resolve concentration profile is under 5%.

CHAPTER 6

CONCLUSTIONS

Program SMCR version 1.0 was developed to resolve the UV/VIS spectra and chromatogram of multicomponent mixtures using the chemometric methods especially self-modeling curve resolution techniques such as Orthogonal Projection Approach (OPA), SIMPLe-to-use Interactive Self-Modeling mixture Analysis (SIMPLISMA), Evolving Factor Analysis (EFA) and Multivariate Curve Resolution. The obtained results are the concentration profiles and absorptivity profiles of each component in the mixture. The program works well on the data that obtained by simulation and experimental data.

The program SMCR version 1.0 contains 16 subroutines for calculation and the 10 subroutines for building graphical user interface (GUI) as listed in Table 3.1. The subroutines *OPA.m*, *SIMPLISMA.m* and *EFA.m* were used to determine the number of significant components whiles the subroutines *ALS.m* was used to optimize the concentration and absorptivity profiles from the spectra of multicomponent mixtures.

According to the theoretical simulations, the SMCR program can be used to resolve the spectra of acid-base dissociation, metal-ligand complexation and chromatographic separation systems.

- For acid-base dissociation equilibria, the system consists areas where no overlapping between component spectra. The accuracy of results, number of significant component, corresponding concentration and absorptivity profiles of each component from OPA, SIMPLISMA and EFA methods were not different and all methods are acceptable.

- For metal-ligand complexation, the efficiency of all three methods in resolving the spectra is different. It can not be found the region of non-overlapping for pure component. Three methods can resolve the concentration profile with high accuracy but can not predict the absorptivity profile as accurately. The program based on EFA method gives the best results of the three methods because the EFA algorithm based on the

elucidation of concentration profiles gives higher accuracy than OPA and SIMPLISMA based on the resolution of pure component spectra.

.

- For chromatographic separation system, the OPA and EFA can resolve the concentration profiles and absorptivity profiles with acceptable accuracy. This is not the case for SIMPLISMA method. In this system, their components were separated by chromatographic technique prior to absorbance detection. The SIMPLISMA method which searches the pure component spectra dependently with other components was not suitable for this system.

In the known experimental spectra, the program was employed to resolve the two-step reaction system and Copper-Glycine and Copper-alanine complexation system.

- For two-step reaction system, the spectra recorded between 200 – 600 nm. can be resolved by the program SMCR based on all of three methods with high accuracy. The program based on EFA method can predict the concentration profile with highest accuracy of 3 methods and are good agreement with the previous study. For resolving the spectra recorded between 250-500 nm , the results from three methods did not agree with the previous study. It can be concluded that the size of absorbance data matrix is the important factors for resolving by the program.

- For Copper-Glycine and Copper-alanine complexation system, the output results of the both systems of copper-glycine and copper-alanine were possessed a similar trend in the properties. EFA algorithm generates concentration profiles first and then absorptivity, while OPA and SIMPLISMA both search for pure component spectra and absorptivity profiles first. For this reason, only program based on EFA method can predict the number of significant components correctly while the OPA and SIMPLISMA can not.

The applicability of the program depends on several factors such as the size of data matrix, degree of spectral overlapping between the components, baseline variation, and the formation of unexpected species; signal to noise of spectrum, and so on. The effect of these factors causes the error of calculations.

REFERENCES

1. Jiang, J.H.; Liang, Y.Z.; Ozaki, Y. Principles and Methodologies in Self-modeling Curve Resolution. *Chemometrics and Intelligence Laboratory System.* **2004**, 71, 1-12.
2. Sanchez, F. C.; Bogaert, B.V.; Rutan, S.C.; Massart, D.L. Multivariate peak purity Approaches. *Chemometrics and Intelligence Laboratory System.* **1996**, 34, 139 – 171.
3. Paatero, P. Least square formulation of robust non-negative factor analysis. *Chemometrics and Intelligence Laboratory System.* **1997**, 37, 23 – 35.
4. Sanchez, F.C.; Rutan, S.C.; Gracia, M.D.; Massart, D.L. Resolution of Multicomponent overlapped peaks by the orthogonal projection approach, evolving factor analysis and window factor analysis. *Chemometrics and Intelligence Laboratory System.* **1997**, 36, 153-164.
5. Maeder, M. Evolving Factor Analysis for the resolution of overlapping peaks. *Anal. Chem.* **1987**, 59, 527-530.
6. Malinowski, E.R. Window Factor Analysis: Theoretical derivation and application to flow injection analysis data. *Journal of Chemometrics.* **1992**, 6, 29 -40.
7. Manne, R.; Shen, H.L.; Liang, Y.Z. Subwindow Factor Analysis. *Chemometrics and Intelligence Laboratory System.* **1999**, 45, 171-176.
8. Frenich, A.G.; Galera, M.M.; Martinez, J.L.; Massart, D.L.; Torres-Lapasio, J.R.; Brakeleer, K.D.; Wang, J.H.; Hopke, P.K. Resolution of multicomponent peaks by orthogonal projection approach, positive matrix factorization and alternating least squares. *Analytica Chimica Acta.* **2000**, 411, 145-155.
9. Shen, H.; Liang, Y.; Kvalheim, O.M.; Manne, R. Determination of chemical rank of two-way data from mixtures using subspace comparison. *Chemometrics and Intelligence Laboratory System.* **2000**, 51, 49-59.
10. Frenich, A.G.; Zamora, D.P.; Martinez, J.L.; Galera, M.M. Resolution (and qualtitation) of mixtures with overlapped spectra by orthogonal projection approach and alternative least squares. *Analytica Chimica Acta.* **2001**, 449, 143-155.

11. Gourvenece, S.; Massart, D.L.; Rutledge, D.N. Determination of the number of components during mixture analysis using the Durbin-Watson criterion in the Orthogaonal Projection Approach and in the SIMPLe-to-use Interactive Self-modelling Mixture Analysis approach. *Chemometrics and Intelligence Laboratory System.* **2002**, 61, 51 – 61.
12. Chen, Z.P.; Morris, J.; Martin, E.; Yu, R.Q.; Liang, Y.Z.; Gouy, F. Recursive evolving spectral projection for revealing the concentration windows of overlapping peaks in two-way chromatographic experiments. *Chemometrics and Intelligence Laboratory System.* **2004**, 72, 9-19.
13. Sanchez, F.C.; Toft, J.; Bogaert, B.V.D.; Massart, D.L. Orthogonal Projection Approach applied to peak purity assessment. *Anal. Chem.* **1996**, 68, 79-85.
14. Rubinson, K.A.; Rubinson, J.F. Contemporary instrumental analysis. New York: Prentice Hall international INC., **2000**, 1-214.
15. Bijlsma, S.; Louwerse, D.J.; Smilde, A.K. Estimating rate constants and pure UV- vis spectra of a two-step reaction using trilinear models. *Journal of chemometrics.* **1999**, 13, 311-329.
16. วุฒิชัย พาราสุข และ อรุณชัย ตั้งเจริญบำรุงสุข. รายงานประกอบโปรแกรมวิเคราะห์ตัวประกอบอย่างง่ายสำหรับการประยุกต์ทางสเปกโทรเคมี. กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย, 2542.
17. Hartley, F.R.; Burgess, F.R.; Alcock, R.M. Solution Equilibria. Ellis Horwood, **1980**.
18. Brereton R.G. Chemometrics : Data analysis for the laboratory and chemical plant. Chichester: John Wiley & Sons, **2003**, 1-323.
19. Malinowski, E.R. Factor analysis in chemistry. New York: John Wiley & Sons, **1981**.
20. Wasim, M.; Brereton, R.G. Determination of the number of significant components in liquid chromatography nuclear magnetic resonance spectroscopy. *Chemometrics and Intelligence Laboratory System.* **2004**, 72, 133-151.
21. Jiang, J.H.; Liang, Y.Z.; Ozaki, Y. On simplex-based method for self-modeling curve resolution of two-way data. *Chemometrics and Intelligence Laboratory System.* **2003**, 65, 51-65.
22. Irving, H.; Pettit, L.D. *Journal of Chemical Society*. **1963**, 1546 – 1553.

APPENDICES

APPENDIX I

SMCR version 1.0: Program manual

Program SMCR version 1.0 was developed to resolve the spectroscopic spectra of multicomponent mixtures using the chemometrics methods based on self-modeling techniques such as Orthogonal Projection Appoarch (OPA), SIMPLe to use Interactive Self-modeling Mixture Analysis approach (SIMPLISMA), Evolving Factor Analysis(EFA) and multivariate curve resolution. The obtained results are the profiles of concentration and absorptivity of each component the mixtures. The source codes of the program were implemented in MATLAB version 6.5 (Math Works, Inc.). The program is running on the graphical user interface (GUI). The compact manual provides step by step instruction and demonstration with an example of input and output in an easy understanding.

Running the Program

1) In MATLAB command window, execute the program by typing SMCR_v1 as:

```
>> SMCR_v1
```

After pressing the ENTER key, the program would clear the screen and start to run as the following:

```
---------------------------------------------------------------------------------------
 SMCR v1.01 - Self-modeling multivariate curve resolution program

 Cite this work as:
 Assoc. Prof. Dr. Vudhichai Parasuk; Kanet Wongravee
 Department of Chemistry, Faculty of science, Bangkok THAILAND 10330
---------------------------------------------------------------------------------------
```

The Graphical User Interface (GUI) of main program SMCR will display as shown in the Figure 1. The GUI of main program composed 5 subprograms as:

- INPUT DATA
- COMPONENT ANALYSIS
- MCR-ALS
- PREDICTION
- SAVE RESULTS
- EXIT

In the first step, the only button of “INPUT DATA” is active, whereas the other buttons are inactive.

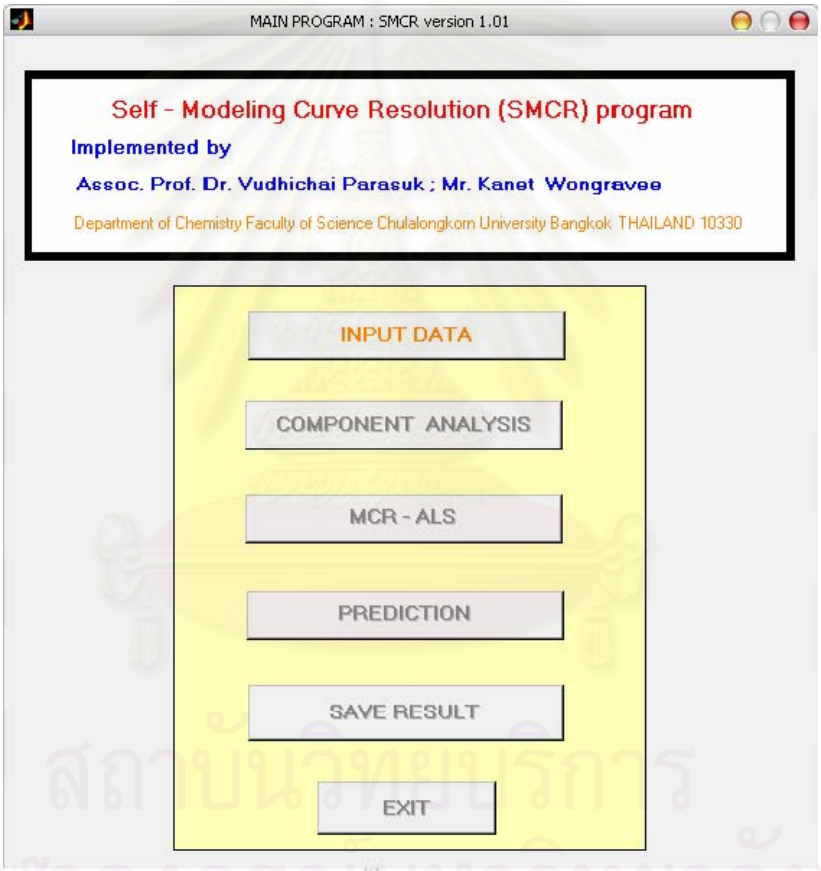


Figure 1. Graphical User Interface of main program: SMCR version 1.01

2) After pressing the button of “INPUT DATA” on the GUI of main program, the GUI of input data will display as shown in the Figure 2:

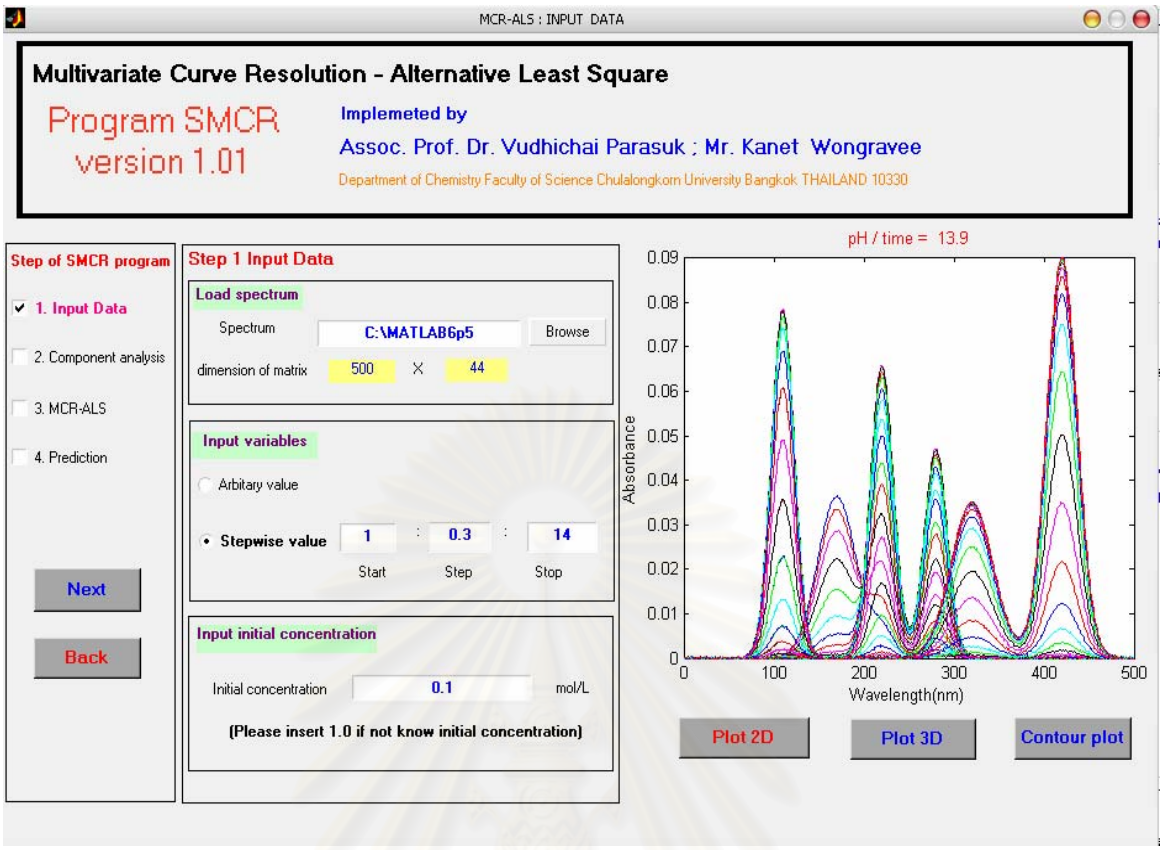


Figure 2 The graphical user interface for INPUT DATA

2.1) Load Spectrum

Begin to input the spectral filename : Use the “Browse button” in the GUI to find your data filename and upload to the program as shown in Figure 3.

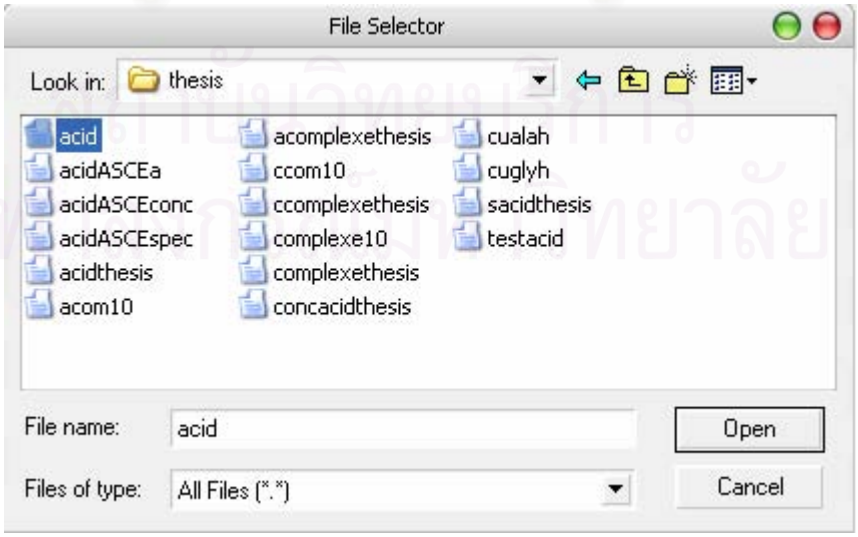


Figure 3 the example for uploading spectra file, namely “acid.txt”, as the input data

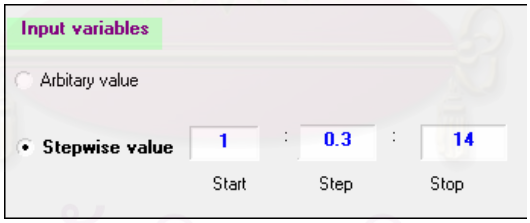
The format of the spectra file would be an ASCII format which created by a text editor such as Notepad or Microsoft editor programs. The structure of the file is correspondent to the spectra matrix which each row associated to the i^{th} wavelength, and each column associated to the j^{th} mixture except for the first column stands for the column of wavelength. The example of spectra file was shown below:

200	0.000372	0.000271	0.000290	0.000170	0.000169
201	0.000179	0.000450	0.000163	0.000027	0.000197
202	0.000562	0.000485	0.000217	0.000200	0.000100
203	0.000621	0.000352	0.000429	0.000053	0.000202
204	0.000895	0.000834	0.000770	0.000357	0.000129
205	0.000919	0.000878	0.000730	0.000378	0.000131
.....					
.....					

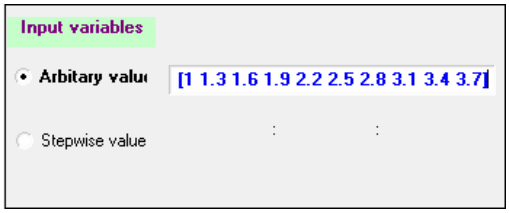
After loading the file, the program would display the dimension of spectra matrix.

2.2) Input variables

Specify the range of variables such as Time or pH. Here there are two possible ways. The first is to specify with an equally step variables as shown below.



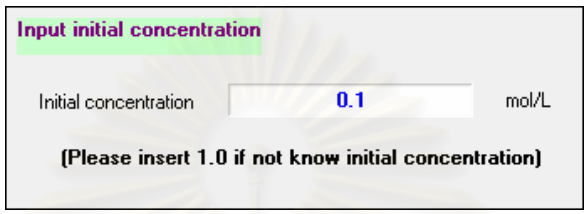
The input ‘1:0.3:14’ means that starting at 1 unit pH and stopping at 14 unit pH with the equal interval of 0.3 unit pH. The second choice is to specify with a series of typical values as shown below.



At this step, the spectra that loaded from the input file would be plotted and displayed as two, three dimensions and contour plot. For example, the spectra were plotted as two dimensions shown in Figure 2.

2.3) Input initial concentration

Specify the initial concentration of the determining component. The initial concentration can be input in the mol/L unit as shown below:



3) In the next step, the program will call the GUI of component analysis. The GUI of component analysis is shown in the Figure 4. There are three methods (OPA, SIMPLISMA and EFA) to be selected in this step. The chemical rank was determined by the plot. Here, there are two possible ways. Firstly, the number of component is to specify with the plot of overall component factor as shown in Figure 4.

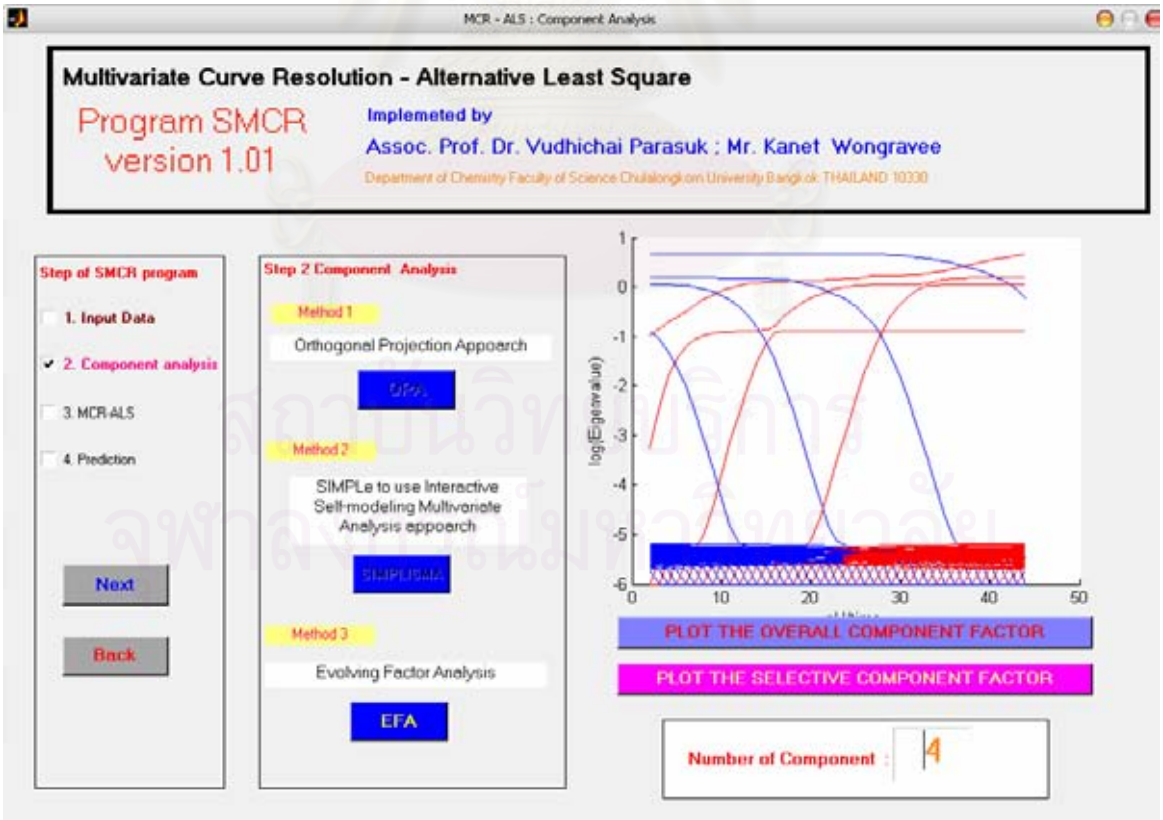


Figure 4. The Graphical user interface of component analysis

Another way, the number of component would be specified with the plot of the selective component factor as shown below:

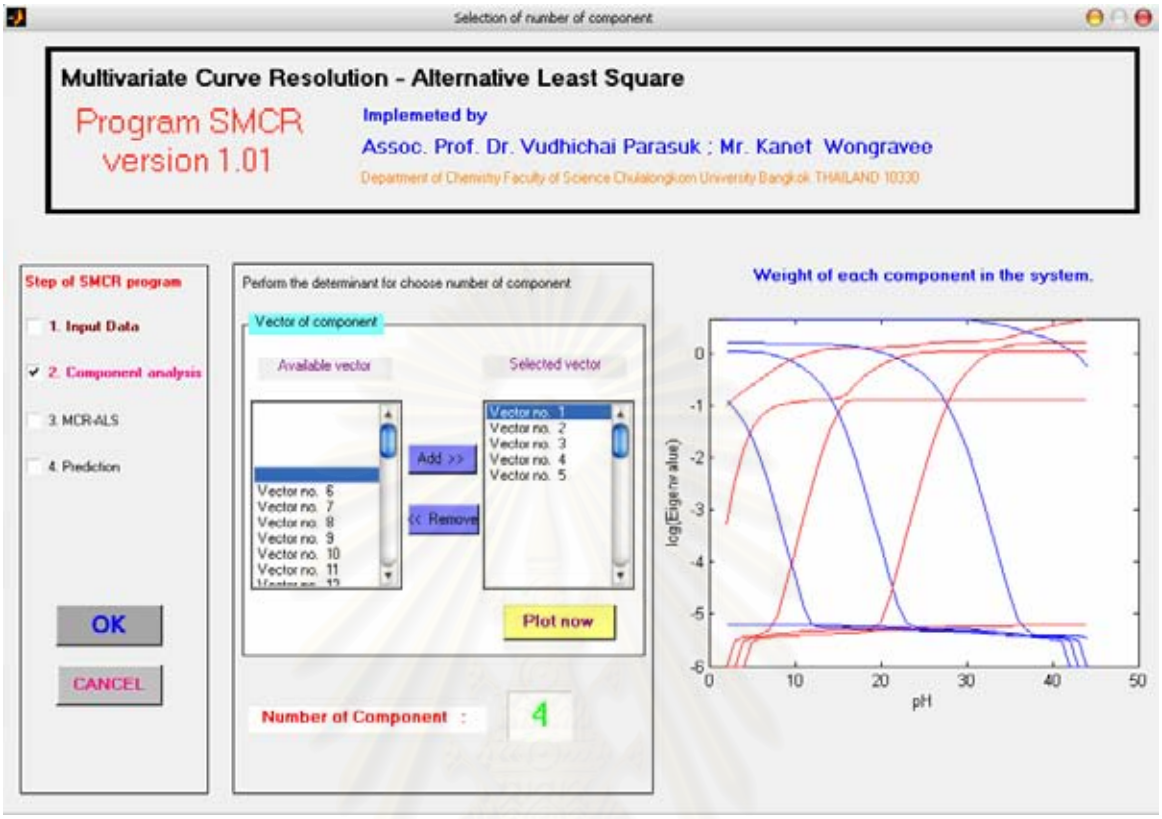


Figure 5. The Graphical user interface of component analysis using selective component factor

4) In the next step, assign the convergence criteria for the process of MCR as shown in the Figure 5. However, the user also has choice to use either the default values of the program: - Maximum cycle = 150; - Tolerance = 0.01 or specify their own values.

Then, the user has choice to apply constraint between optimization for avoid rotational ambiguity. There are three constraints that the user can use.

- Non-negativity
- Unimodality
- Closure

After setting, the user can plot estimated concentration and absorptivity profiles from the chemical rank analysis before optimization shown in Figure 6.

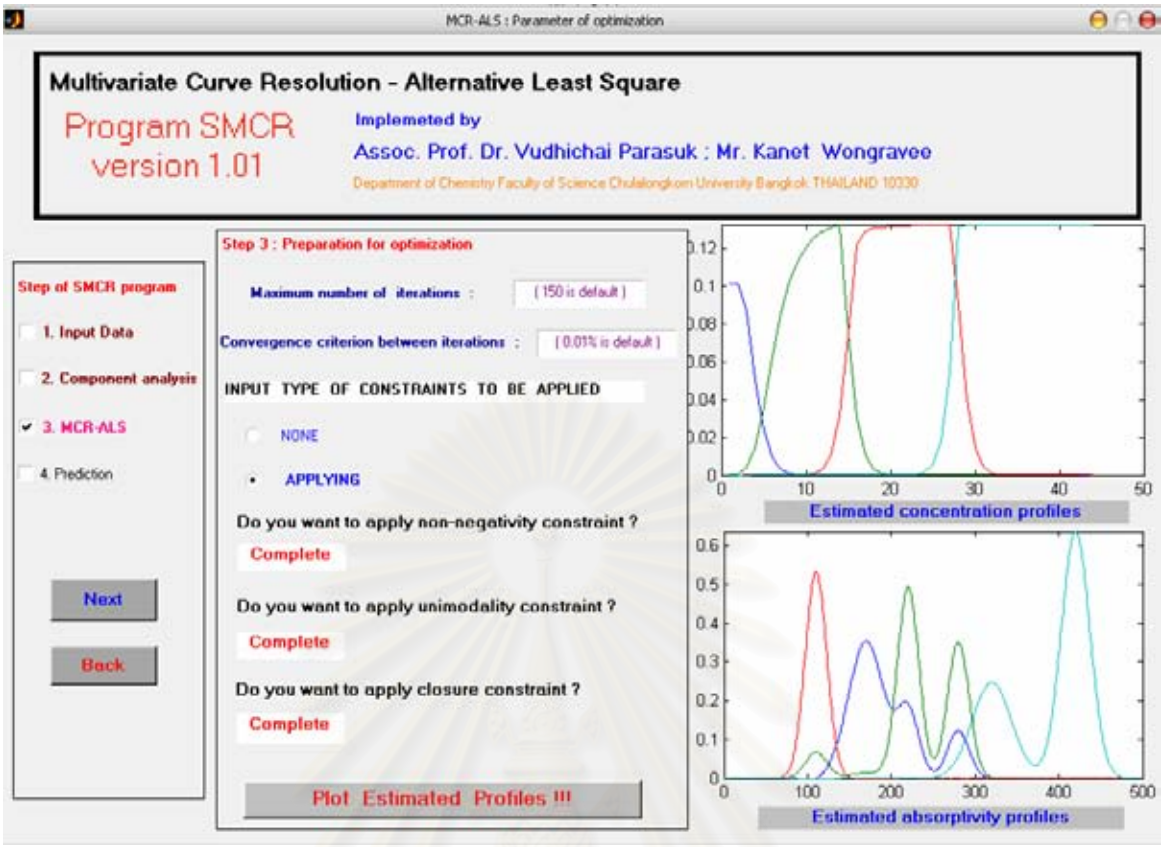
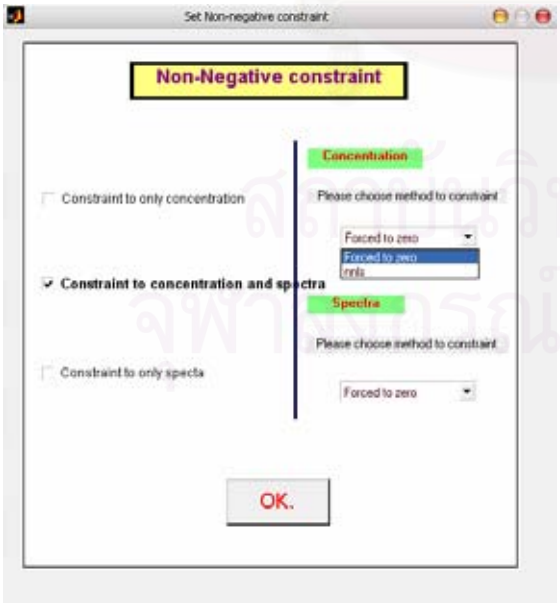


Figure 6. The Graphical user interface of setting parameter for MCR optimization.

4.1) Non-negativity constraint



The user has choice to apply the constraint that suitable for the system. The constraint forces the elements in a profile to be positive.

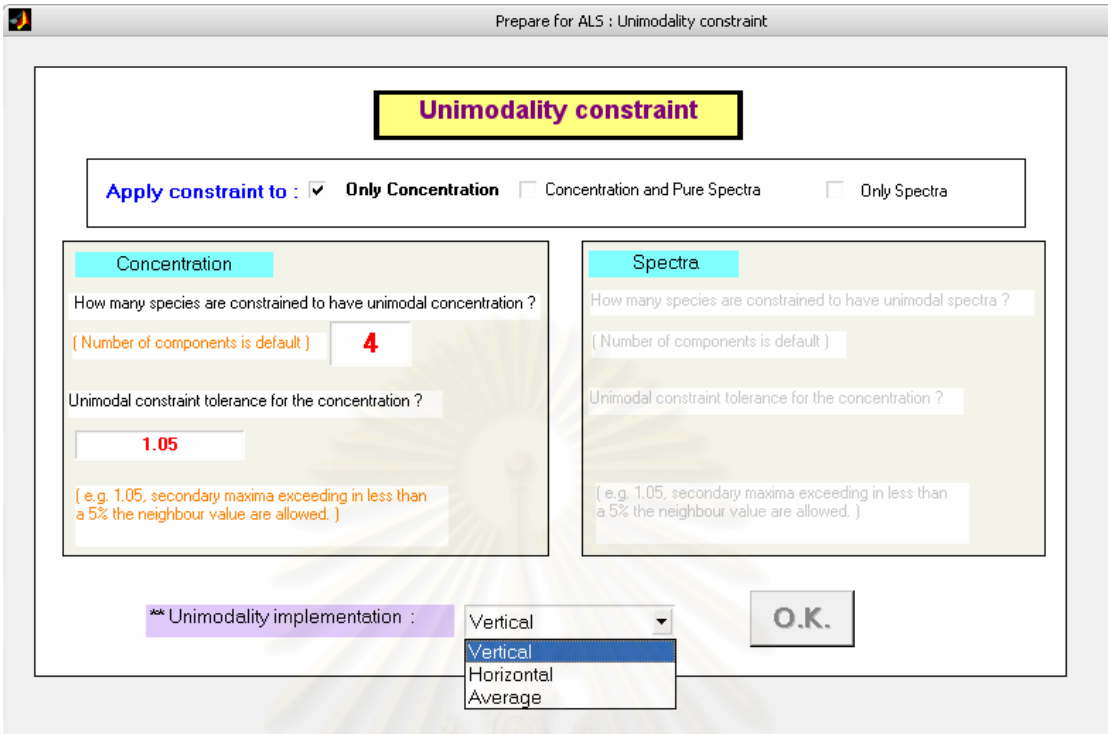
Applied Constraint

- For only concentration
- For Concentration and absorptivity (*recommended*)
- For only spectra

Algorithm for Constraint

- Fored to zero (set negative value to zero)
- nnls (change negative value to positive value used nonnegative least square algorithm)

4.2) Unimodality constraint



The user could change the setting that suitable for system. The constraint allows for the presence of only one maximum per profiles.

Applied Constraint

- For only concentration (*recommended*)
- For Concentration and absorptivity
- For only spectra

Number of species that user would to constraint

- The number of profiles to be constrained (The Number of component is default)

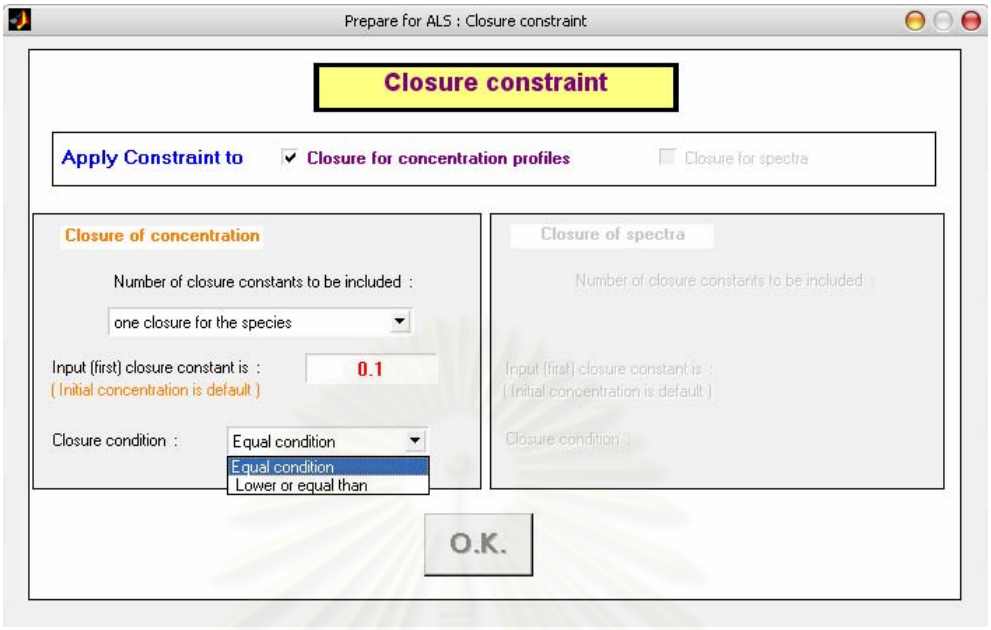
Unimodal constraint tolerance

- Defines how strictly the constraint should be applied. If the answer is 1.0, no departures of the unimodal condition are allowed; if the value is higher than 1.0, slight departures from unimodality are allowed. (e.g., 1.05, secondary maxima exceeding in less than a 5% the neighbour value are allowed).

Unimodality implementation

- Algorithms used in the application of unimodality : Vertical, Horizontal and Average (recommended).

4.3) Closure constraint



The user could change the setting that suitable for system. The constraint is the fulfillment of a mass balance condition. The different profiles (compounds) involved in the closed system are simultaneously constrained.

Closure

- For only concentration profiles (recommended)

- For only spectra profiles

Number of closure constraints

- Number of closed systems in the same data matrix. One closure for the species (Have one maximum values of sum of all species) is default.

Input closure constraint

- Value of the total concentration in the closed system. The initial condition is default. If not known, you may enter 1.

Closure condition

- Defines the tolerance of the constraint. *"Equal condition"* forces the sum of the concentratins in the closed system to equal exactly the total concentration at each stage of the process. "Lower or equal than" allows for some departures of this condition, i.e., slight variations of the total concentration in the system may be allowed.

5) After setting the criteria convergence, the program determines the concentration and absorptivity profiles of each component in the mixtures. The user should wait until the process is complete. During optimization, the program would plot the concentration and absorptivity profiles in the each iteration. The GUI of optimized calculation is shown in Figure 7.

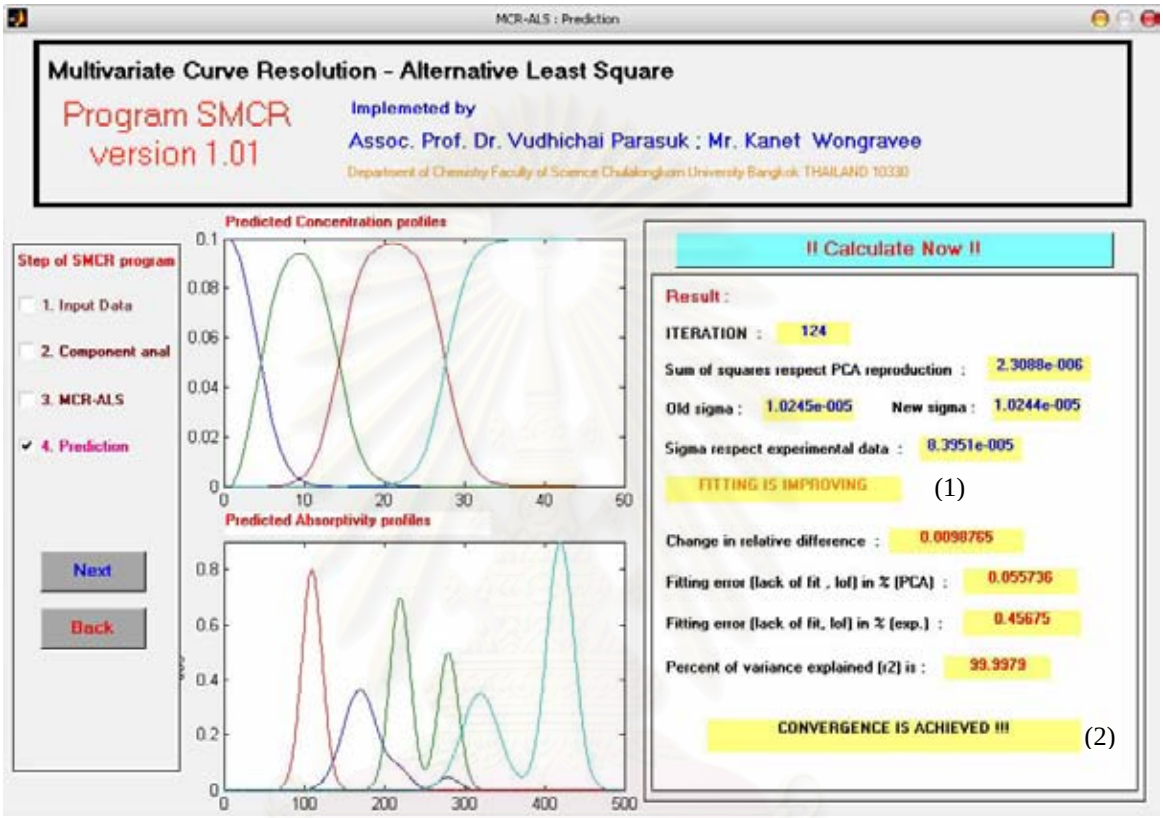


Figure 7. The Graphical User Interface of optimized calculation.

Per each iteration *n*, the following results are shown:

Iteration: iteration of optimized calculation.

Sum of squares respect PCA reproduction: Sum of squares of residuals between the MCR-ALS reproduction and the PCA model

Old sigma and New sigma: Std. deviation of the residuals (MCR-ALS vs. PCA) for iteration *n-1* (old sigma) and *n* (new sigma).

Sigma respect experiment data: Std. deviation of the residuals (MCR-ALS vs. experimental data) for iteration *n*.

(1): Diagnostic on the evolution of the fit based on comparison of iteration *n* and last converging iteration.

- Fitting is improving: Std. deviation of the residuals for iteration new sigma is less than old sigma.
- Fitting is not improving: Std. deviation of the residuals for iteration new sigma is higher than old sigma.

Change in relative difference: of sigma change between iteration *n* and last converging iteration

Fitting error (lack of fit) in % (PCA): compares the matrix obtained from the resolution results with the matrix obtained by PCA reproduction using the same number of components as the raw data set.

Fitting error (lack of fit) in % (exp.): compares the matrix obtained from the resolution results with the raw data set.

Percent of variance explained (r2) : % variance

(2) : Criteria for optimization

- Convergence is achieved: Change in sigma is less than criteria setting.
- Number of iteration is exceeded the allowed: The optimization would need more iteration number for optimization.
- Fit not improves for 20 times consecutively (divergence): New sigma is higher than old sigma more than 20 iteration numbers.

6) After optimized calculation, the selection an option for saving the results as shown in the Figure 8.

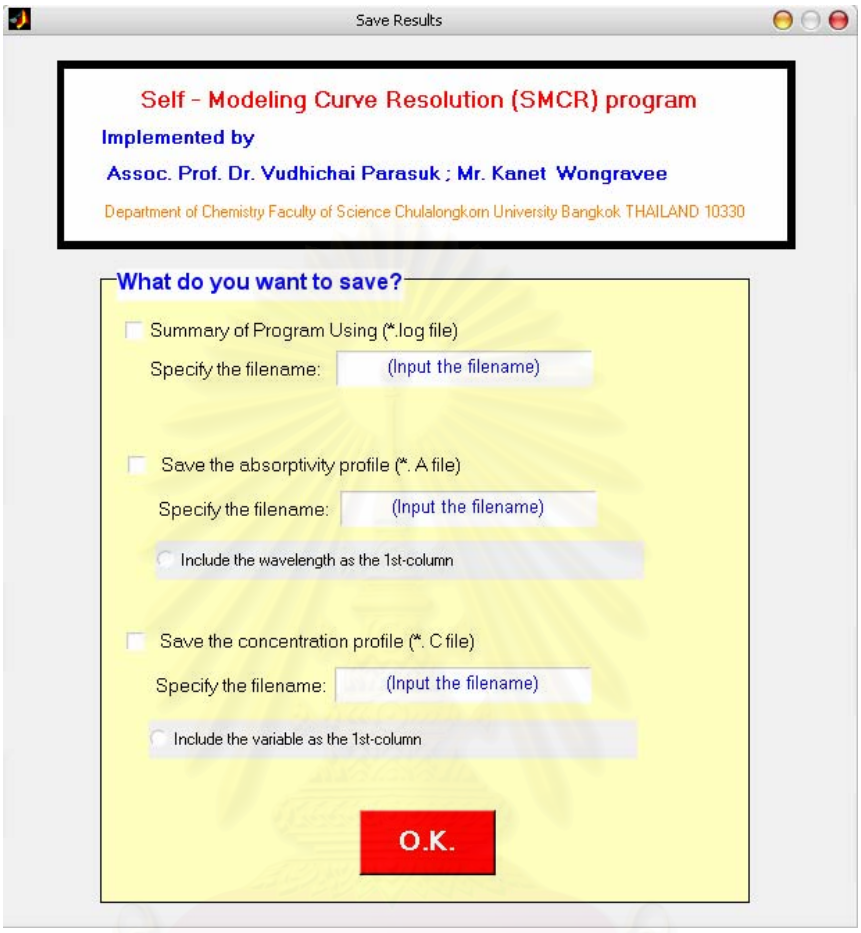


Figure 8. The Graphical User Interface for saving options.

The results of summary of Program, concentration and absorptivity profiles would be an ASCEII format that can use Notepad or Microsoft Excel to open it.

APPENDIX II

The example of source code

This is the main subroutine of SMCR program written with MATLAB language.

```
%Main program for determination of significant components
%
%Method
%   Orthogonal projection Approach(OPA)
%   SIMPLe to use interaction self-modeling mutivarite analysis approach
%   Evolving Factor Analysis(EFA)
%
%Estimated pure spectra and concentration
%   OPA esitmate pure spectra
%   SIMPLISMA estimate pure spectra
%   EFA estimate concentration profiles
%
%Implemented by Mr Kanet Wongravee
%Modified 18/10/2004
%
%
disp('***********************************************')
disp('Self-modeling curve resolution program (SMCR)  ')
disp('Method OPA,SIMPLISMA,EFA                       ')
disp('Implemented by Mr. Kanet Wongravee             ')
disp('CCUC unit cell Chulalongkorn University        ')
disp('Bangkok Thailand 10310                         ')
disp('Modified 18/10/2547                            ')
disp('***********************************************')
disp('** PLEASE INSERT FILE NAME OF SPECTRA DATA  :  **')
%Input
%Pre data set
global Conc Sref Spec
clr = ['b' 'r' 'g' 'm' 'k' 'c'];
Spec = input('Filename that you want to calculate     :');
Xph  = input('Input pH/Time in you experiment [. . .]  :');
[nrow ncol] = size(Spec);

%
%Choice of estimated significant components
disp('Method for determination of significant components')
disp('OPA (1) ')
disp('SIMPLISMA (2) ')
disp('EFA (3) ')
disp('Enter the Choice ,e.g.[1,2,3] ')
Choice = input('Enter the Choice that you should  :')
```

```
%************************************
%Orthogonal projection appoarch(OPA)
%************************************
ml = find(Choice==1);
if ~isempty(ml)
   [Determinant,MaxDis,Xref,pH]=OPAnumber(Spec);
   j=0;
   Terminate = 'y';
   while Terminate == 'y'
   i = j+1;
   figure(i+4);
   plot(pH,Determinant(i,:),clr(1,i));
   title(['Factor number ' int2str(i)])
   Terminate = input('Other num. of factors to be considered:? (y/n)','s');
   j = i;
   end

   n=input('The Number of component from OPA method  :  ');

   Sref = Xref(:,1:n); %Initail estimate of pure spectra from OPA method
   x0 = Sref';

end

%************************************
%SIMPLISMA Method
%************************************
ml = find(Choice==2);
if ~isempty(ml)
   [Determinant,MaxDis,Xreff,pH,offset]=SIMnumber(Spec);
   j=0;
   Terminate = 'y';
   while Terminate == 'y'
   i = j+1;
   figure(i+4);
   plot(pH,Determinant(i,:));
   title(['Factor number ' int2str(i)])
   Terminate = input('Other num. of factors to be considered:? (y/n)','s');
   j = i;
   end
   number = input('The Number of component from SIMPLISMA method  : ');
   [Xref,index] = pure(Spec,number,offset);
   Sref = Xref(:,1:number); %Initial estimated of pure spectra form SIMPLISMA
method
   x0 = Sref';
end
```

```
%*********************************
% Evolving Factor Analysis (EFA)
%*********************************
ml = find(Choice ==3);
if ~isempty(ml)
   [e,eforward,ebackward] = efaop(Spec);
   % e = Conc (etimated conce from efa method)
   x0 = e;
end
d = Spec';
isp = input('correspondence among the species in the experiments  : ' );
nexp = input('number of data matrices analyzed simultaneously  :  ');
nit = input('maximum number of iterations (50 is the default ) :  ');
tolsigma = input('convergence criterion in the difference of sd of residuals between
iterations (0.1% is the default) :  ');

%************************************
% Alternating least square (ALS) optimization
%************************************
[copt,sopt,sdopt,ropt,time]=alsmo(d,x0,nexp,nit,tolsigma,isp);
```

VITAE

Name : Kanet Wongravee

Date of Birth : January 30, 1981.

Birth Place : Nakhon Nayok, Thailand.

Marial Status : Single

Current Address :

269/19 Soi Suwannasorn 9,
Suwannasorn Rd. Meung.
Nakhon Nayok, 26000, Thailand.

Education :

High school	Bodindecha (Sing Singhasanee), Thailand.
Bachelor of Science	Chulalongkorn University, Thailand, 2nd Class Honours (Chemistry).

Awards and Scholarships :

Financial support from the Development and Promotion of Science and Technology talents project of Thailand (DPST) during my graduate studies at Chulalongkorn University and scholarship of Siam Cement Thai for visiting Mizui and Kubota group, Japan.