CHAPTER 4
SIMULTANEOUS ASSAY OF EACH SUBSTANCE IN TERNARY MIXTURE OF FOOD COLORANTS VIA SPECTROMETRY

4.1 Introduction

Many analytical techniques have been used for the simultaneous determination of various food colorants in mixture of them [1, 2]. Chromatographic methods are popular methods to determine concentrations of mixed colorants in food analysis [3, 4] such as Thin Layer Chromatography (TLC) [5], Capillary Electrophoresis (CE) [6-8], and High-Performance Liquid Chromatography (HPLC) [9-13]. Some researches transformed signal to increase selectivity of determination of colorants in mixture by using derivative of UV-VIS spectra of colorant mixtures [6, 14-17]. The derivative process can be used well with binary complex mixture but it is hard to deal with the combination of ternary mixture or more complex mixtures.

Although chemical separation methods can be give good results for determinations of multiple colorants in mixture but those techniques need complicated system and take much time to perform experimental. Ultraviolet visible spectrophotometry is used in many fields because of the operation can be done with short time, easy to operate and low cost. Analysis of the analytes in mixture by UV-VIS spectroscopy give a complicated overlapped spectra so it cannot directly determine the analytes from mixture. The mixtures need separation processes before each analyte was detected by UV-VIS spectrometers.

Various multivariate calibration methods were present to determine colors in mixture without separations methods such as Classical Least-Squares (CLS), Inverse

Least-Squares (ILS), Principal Component Regression (PCR) and Partial Least-Squares (PLS) regression techniques [18-21].

PCR and PLS method were bolded out for aim of determination of multicomponent mixture because of their good points. Those 2 methods can extract information from overall data and leaved out useless noise data.

In this study, tartrazine, pouceau 4 R and indigo carmine were selected to be used as analytes component in mixture because of their overlapped spectra. Demonstration of increasing of ability of PCA and PLS model by using limited data. The consideration about training set and signal is important to study for construction of multicalibration model.

Mostly the multivariate calibration models performed on designed optimum training sets to increase quality prediction. Nevertheless, in situation of lack of time and chemical, it cannot do experimental to setup perfect training sets. In this part, management of limited data to setup multivariate calibration models is aimed to study. Important factors obtained algorithm of calculation, number of training set and characteristics of signals were varied to find the best models to predict multiple components chemical concentrations in the mixtures in the limited data. The data of tartarzine, ponceau 4 R and indigo carmine in their mixtures which obtained concentration data and spectra of the mixtures data was used as data model to studied for the aim.

4.2 Experimental

4.2.1 Apparatus

A Lambda 25 spectrophotometer equipped with 1.0 cm cells and connected to an genuineIntel Intel(r) family 6 model 5 processor 160.0 MB RAM computer provided with UV-Winlab version-2.85.04 copyright 2000 PerkinElmer, Inc software in scaning mode. PCR and PLS1 methods were calculated in the "R Project for Statistical Computing" software with the "pls software package" and PLS2, the algorithm was wrote in "Matlab Version 7.1.0.246 (R14) Service Pack 3", all of software are operated on an AMD Athlon™ 64X2 Dual Core Processor 3800+ 2.20 GHz., 2.00 GB of RAM Physical Address Extension, which were used for the statistical treatment of the data and for the application of PCR PLS1 and PLS2 methods.

4.2.2 Reagents

All solvents and reagents were of analytical reagent-grade unless indicated otherwise. Indigo carmine, I was supplied by BDH; Tartrazine, T and Ponceau 4R, P were provided by Adinop Co., Ltd. (Thailand). T, P, and I stock aqueous solutions with a concentration of 200 mg/l were prepared. Acetic acid–sodium acetate (0.1 M and pH 5.5) was used as buffer solution.

4.2.3 Experimental procedure

In 50 ml calibrated flasks, aliquots of the stock solutions were added to obtain concentrations between 4 and 20 mg/l of T, between 4 and 24 mg/l of P, and between 4 and 20 mg/l of I; 10 ml of acetate buffer solution (pH 5.5) and deionized water (milli Q quality) to the mark were also introduced. In Figure 4.1, spectra for solutions of 10 ppm T (1), 10 ppm P (2), 10 ppm I (3), and their mixtures of 10

ppm of T, 10 ppm of P, and 10 ppm of I (4), prepared in 10 ml of acetate buffer solution (pH 5.5) and deionized water (milli Q quality) recorded against a blank of deionized water.

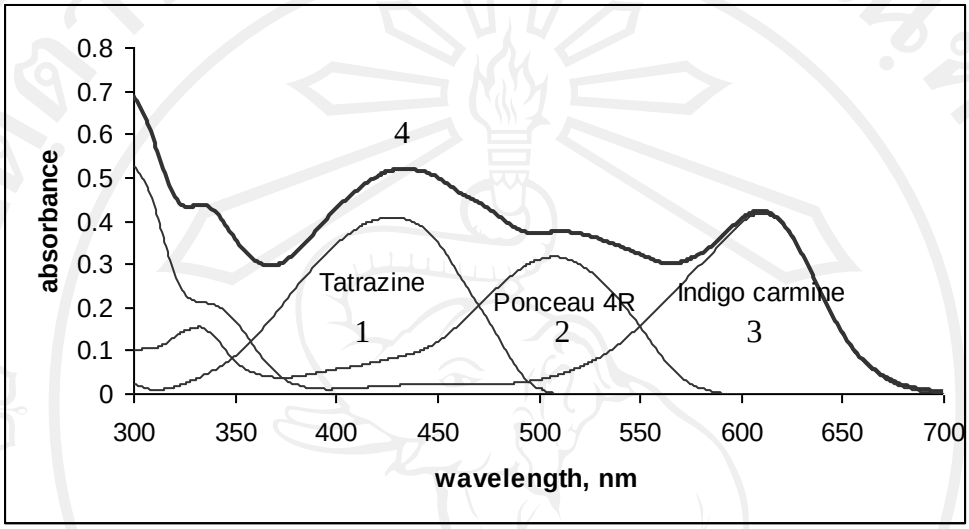

Figure 4.1 Absorption spectra for solutions of 10 ppm T (1), 10 ppm P (2), 10 ppm I (3), and their mixtures of 10 ppm of T, 10 ppm of P, and 10 ppm of I (4), prepared in 10 ml of acetate buffer solution (pH 5.5) and deionized water (milli Q quality) recorded against a blank of deionized water.

The absorption spectra were recorded between 300 and 750 nm with a scan rate of 960 nm/min against a blank of deionized water. The optimized calibration matrix, obtained by the use of recorded absorption spectra of standards (in the spectral ranges: 300–700 nm, 370-700 nm and 400-700 nm) was applied to analyze the spectra and to calculate the concentrations of T, P, and I in synthetic mixtures. The calibration models were further investigated by the cross-validation (obtaining statistical parameters that show the efficiency for a calibration fit model), the internal validation (prediction of dyes concentration in its own designed training set of calibration), and the external validation (prediction of dyes concentration in the test set). Three types of method; PCR, PLS1 and PLS2, training set characteristic and signal characteristic were compared to optimize quality of prediction of each component in mixtures.

4.2.4 Statistical techniques

4.2.4.1 Multivariate calibration techniques [1]

Multivariate calibration techniques are techniques which calculate relationship between multivariate data (X-data) or in this work; spectra of mixtures and the analytical concentration data (Y-data). In this work principal component regression, partial least square 1 (PLS-1) and partial lease square 2 (PLS-2) were applied for the simultaneous determination of colorants in mixture solutions. The data of mixtures spectra and analytes concentrations were extracted by multivariate calibration methods to be the condensed data of the samples called score data of principal latent component of the condensed variables called loading data.

PCR modeling (A) uses regression to convert attracted X-data (also called principal component or score (T) and another matrix is loading (P)) from principal component analysis (PCA) to Y-data. From the regression, the rest of noise data, E, were left out so PCR can deal with noise and collinear problem in raw data by select only important principal components.

$$X = TP' + E$$

$$Y = T.A$$

PLS-1 constructs factor from variation of X-data and the analytical concentration of Y-data. PLS-1 model (Q) is regressed from X -data score (T) with weight with inner relation (H) between X -data and the analytical concentration Y-data. PLS-1 can decrease influent of the analytical concentration Y-data variation by including it to inner relation.

PLS-2 use concept of PLS-1. Then, score T from PLS-2 is calculated from variation of matrices in solution so inner relation between X -data and

all components in Y-data are considered. PLS-2 model can deal with influent of the analytical concentration Y-data and the matrices Y -data variation to signal.

X = T.P' + E

T = X.H (H = inner relation)

Y = TQ' + F

The 3 multivariate calibration methods are extract data in different way. Principal component of PCR is extracted from distribution of signal. Latent component of PLS1 is extracted from distribution of signal and concentration data of analyte and weighted with concentration data of analyte. Latent component of PLS2 is extracted from distribution of signal and concentration data of all interested components in mixture and weighted with concentration data of all interested components in mixture.

4.2.4.2 Validation statistics terms [2, 3]

Cross validation-leave one out is applied for validation of multivariate model. The calculate statistic terms contain:

The prediction error sum of squares (PRESS) explain about how well model can predict the analytical concentration when various number (g) of principal components was used is given by equation 4.1

$$PRESS(g) = \sum (obs_g - pred)^2 \quad \text{.....4.1}$$

The root mean squares difference (RMSD) is the standard deviation of prediction to explain quality of model, given by equation 4.2

$$RMSD = \sqrt{\frac{PRESS}{N}} \quad \text{…….4.2}$$

The square of the correlation coefficient (R^2), is indicate that how well all of data in training set can fit into calibration model, is given by equation 4.3

$$R^2 = 1 - \frac{\sum_i (ops - pred)^2}{\sum_i (ops - \overline{ops})^2} \quad \text{.....4.3}$$

In validation step, the predictive ability of each model ***by validation set*** is evaluated using the root mean square error of prediction (RMSEP). This is defined as equation 4.4

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (ops_i - pred_i)^2}{n}} \quad \text{.....4.4}$$

where ops_i is the actual value of object i and $pred_i$ is predicted value with the model under evaluation of object i, and n is the number of objects in validation set for which pred is obtained by prediction.

Reliable of prediction was considered by concentration and concentration ratio ranges of colorants in training set, when the prediction was out of range of concentration and ratio of concentration of colorants, the prediction was rejected.

4.2.5 Data management

The data obtained concentration data and signal data of 76 mixtures in different ratio of T, P and I compound. In this work, the data was divided into 3 sets for calibration set, validation set and unknown set. The criteria of data dividing is "data covering"; by using concentration range and ratio of colorant concentration. Compositions of each colorant in training sets were shown in Table 4.1 and Table 4.2.

Model validation was done by using 8 mixtures validation samples and leave-one-out cross validation. Compositions of each colorant in validation sets were shown in Table 4.3. The ranges of colorants concentrations and ratios of validation set samples are within training set. The rest of data 35 mixtures data was obtained both within and out of boundary of training set (to demonstrate prediction in real situation).

4.2.6 Optimization of model quality

Many factors are effected quality of prediction by multicalibration model. The main factors which most important to consider for reducing analysis time and also gaining well accuracy are calibration method, characteristic of training set and signal.

4.2.6.1 Multivariate calibration method effect

Multicalibration model is correlation of the extraction of information from analysis data (signal) and concentration data. Extracted data is general known in term principal or latent component. PCR, PLS1 and PLS2 are linear regression which has different data extraction process which already explained in 4.2.4. Models of those three methods are different. If the mixture contained of n components, the models from PCR and PLS1 method of n components are n models, in another hand, 1 model from PLS2 can be used to predict n components in the same time. The complication to deal with prediction of PLS2 is less than PCR or PLS1. In this study the 3 methods were applied by using 7 training sets that contained different number of training mixtures (to confirm the comparison). Cross validation and external validation were used to select the best method, by using RMSD and RMSEP.

4.2.6.2 Training set characteristic effect

Minimization of number of mixtures in training set was studied to decrease analysis time. Studying of training set number effect was performed by decomposing from calibration design and unsystematic design. The boundary of prediction was also studied into 2 ways; 4.2.6.2.1) Colorants concentration ranges was controlled and 4.2.6.2.2) colorants concentration ranges and colorants concentration ratios were controlled.

4.2.6.2.1) study of training set number effect when colorants concentration ranges was controlled

In this part, decreasing of number of mixture in training set was studied from calibration design and unsystematic design. From Table 4.1, 35 mixtures data was selected to study training set number effect.

Decreasing number of training mixtures from calibration design of 5 training sets were selected from 35 mixtures data. Training set 01 was performed by calibration design. Training set 02-05 was decreased to 20, 15, 10, 5 mixtures, respectively.

Training set 01-05 were compared when concentration range is 6-16 ppm, 4-20 ppm, and 6-16 ppm, for T, P and I, respectively.

Training set 06 was performed with increase boundary of prediction by adding 10 mixtures into training set 01 and become to unsystematic training (35 mixtures) and training set 07 (which had same boundary as training set 06) was decrease number of mixtures in training set 2/3 time of training set 01(10 mixtures). Training set 06-07 were compared when concentration range of

T, P and I is 4-18 ppm, 4-24 ppm, and 5-18 ppm, for T, P, I, respectively. Colorants concentrations in the mixtures were predicted within boundary that considered from only concentration range. RMSEP is term to use for explain quality of prediction.

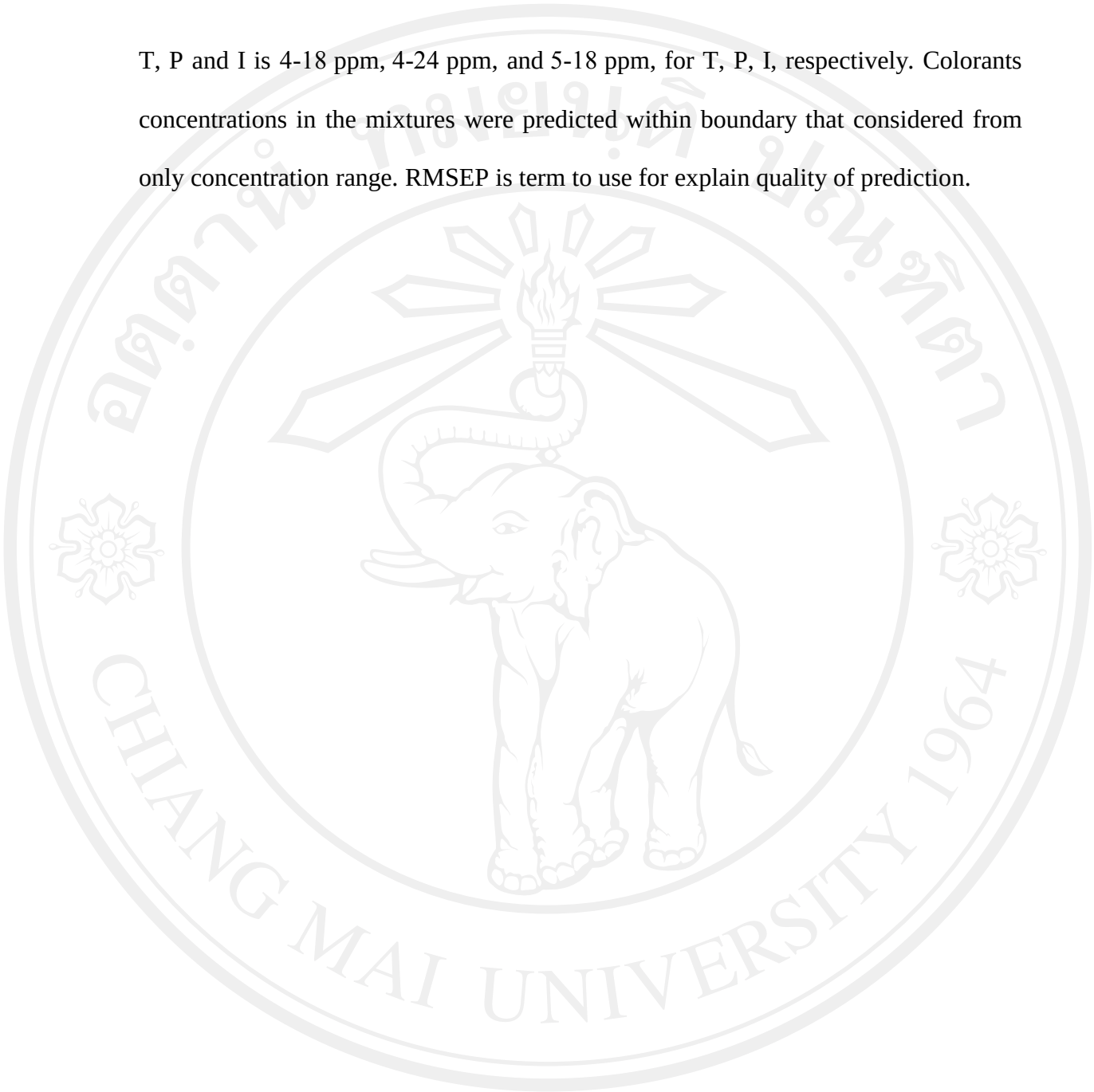

Table 4.1 Composition of T, P and I for training set selection (*colorants concentration ranges was controlled*)

	colorant concentration, ppm			training set #						
#	T	P	I	set01	set02	set03	set04	set05	set06	set07
21	10	12	10	/	/	/			/	
22	6	4	16	/	/	/	/	/	/	
23	16	8	16	/					/	
24	16	12	8	/	/	/	/		/	
25	8	16	16	/	/	/			/	
26	10	20	16	/	/	/	/	/	/	
27	6	16	6	/	/				/	
28	12	4	10	/	/				/	
29	12	16	8	/					/	
30	12	8	6	/	/	/	/		/	
31	8	4	8	/	/				/	
32	6	8	10	/					/	
33	8	12	6	/	/	/	/		/	
34	10	4	6	/	/	/			/	
35	6	20	8	/	/	/	/	/	/	
36	8	20	10	/	/				/	
37	10	8	8	/	/				/	
38	8	8	12	/					/	
39	12	20	12	/	/	/			/	
40	16	16	10	/	/	/			/	
41	12	12	16	/					/	
42	16	20	6	/	/	/	/	/	/	
43	16	4	12	/	/	/	/	/	/	
44	6	12	12	/	/	/	/		/	
45	10	16	12	/	/	/	/		/	
56	10	10	10						/	/
57	10	4	6						/	/
58	5	5	5						/	/
59	5	10	15						/	/
60	15	5	10						/	/
61	4	8	10						/	/
62	10	5	10						/	/
63	18	18	18						/	/
64	10	15	5						/	/
65	12	24	12						/	/
[T] range				6-16	6-16	6-16	6-16	6-16	4-18	4-18
[P] range				4-20	4-20	4-20	4-20	4-20	4-24	4-24
[I] range				6-16	6-16	6-16	6-16	6-16	5-18	5-18
[T/[P] range				0.3-4.0	0.3-4.0	0.3-4.0	0.3-4.0	0.3-4.0	0.3-4.0	0.5-3.0
[T/[I] range				0.4-2.7	0.4-2.7	0.4-2.7	0.4-2.7	0.4-2.7	0.4-2.7	0.3-2.0
[P/[I] range				0.3-3.3	0.3-3.3	0.3-3.3	0.3-3.3	0.3-3.3	0.3-3.3	0.5-3.0
[T]/[P]/[I] range				0.03-0.42	0.03-0.42	0.03-0.42	0.03-0.33	0.03-0.33	0.03-0.42	0.03-0.42
number of training set				25	20	15	10	5	**35**	**10**

Table 4.2 Composition of T, P and I for training set selection *(colorants concentration ranges and colorants concentration ratios were controlled)*

#	Colorant conc, ppm			Training set #				
	T	P	I	0A	0B	0C	0D	0E
s04	10	16	20	/	/	/	/	
s05	20	10	5	/	/	/	/	/
s16	6	10	20	/	/	/	/	/
s21	10	12	10	/				
s22	6	4	16	/	/	/		
s23	16	8	16	/	/	/	/	
s26	10	20	16	/	/			
s27	6	16	6	/	/			
s28	12	4	10	/	/	/		
s29	12	16	8	/	/	/		
s31	8	4	8	/	/			
s33	8	12	6	/				
s35	6	20	8	/	/	/	/	/
s36	8	20	10	/	/	/		
s37	10	8	8	/				
s38	8	8	12	/	/	/	/	
s39	12	20	12	/	/	/		
s40	16	16	10	/				
s42	16	20	6	/	/			
s43	16	4	12	/	/	/	/	
s44	6	12	12	/	/			
s45	10	16	12	/				
s52	10	4	4	/	/	/	/	/
s55	16	12	4	/	/	/	/	
s61	4	8	10	/	/	/	/	/
[T] range				4-20	4-20	4-20	4-20	4-20
[P] range				4-20	4-20	4-20	4-20	4-20
[I] range				4-20	4-20	4-20	4-20	4-20
[T/[P] range				0.3-4.0	0.3-4.0	0.3-4.0	0.3-4.0	0.3-2.5
[T/[I] range				0.3-4.0	0.3-4.0	0.3-4.0	0.3-4.0	0.3-4.0
[P/[I] range				0.3-3.3	0.3-3.3	0.3-3.0	0.3-3.0	0.5-2.5
[T]/[P]/[I] range				0.30-0.63	0.30-0.63	0.30-0.63	0.30-0.63	0.30-0.63

4.2.6.2.2 Study on training set number effect when colorants concentration ranges and colorants concentration ratios were controlled

In this part, training set series was studied by concentration of T, P and I range was fixed at 4-20 ppm. Number of mixtures in the training set (0A-0E) was varied from 5-25 mixtures. The compositions of each colorant in the training set samples were shown in Table 4.2.Concentration range and ratio of concentration range of each training set was controlled and used as criteria for boundary of prediction and RMSEP was used as term to descript prediction quality.

Training set which had fewer amount with same quality was used for study on signal range effect.

Table 4.3 composition of T, P and I of validation set

#	expected concentration		
	T	P	I
v1	16	16	10
v2	12	4	4
v3	20	6	6
v4	4	4	4
v5	20	20	20
v6	8	16	16
v7	6	8	10
v8	12	12	16

4.2.6.3 Signal characteristics effect

Range and interval of signal were studied to decrease recorded time. Selected method from 4.2.4.1 which calculated by the selected training set from 4.2.4.2 was used to studied signal range effect. 3 ranges of spectrum; 300-700 nm,

370-700nm, and 400-700nm were compared. Interval of signal (1, 5, 10, 50, 100 nm^{-1}) was studied within the selected signal range.

4.3 Results and discussion

4.3.1 Multivariate calibration method effect

Training set series (explained in topic 4.2.5.2.1) was used for studying of effect from calibration methods; PCR, PLS1 and PLS2, when use different training sets. Optimum number of component was selected from graft between numbers of component and RMSD (scree plot). When the number of components was increasing, the RMSD value was decreased. The components which significantly decrease RMSD value were selected to use to calculate model. Scree plots of PCR, PLS1 and PLS2 of training set 01 are shown in Figure 4.2, respectively. For other models, the same criteriawas used. The scree pots of the models were omitted.

Mostly numbers of principal components of the PCR, PLS1 and PLS2 models were 3, 3, 3 of T, P and I, respectively. Multicalibration models of training set 06 had different numbers of components; [3,3,5][3,2,5][3,3,3] of T, P and I of PCR, PLS1 and PLS2,respectively. RMSEP from external validation was used to explain quality of models.

From Figure 4.6, RMSEP values of prediction of T and P by PCR of each training set are equal to PLS1 and PLS2. The prediction of T and P didn't affect from type of calculation method. RMSEP of prediction of I from PLS1 and PLS2 were more than PCR when training set 06 was used. When number of principal component of each method was considered, it was found that models from training set 01-05, 07 have similar number of principal component. In another hand, when using training set

06, number of principal component of PCR was more than PLS1 and PLS2.PCR model from training set 06 has error less than PLS1 and PLS2 method.

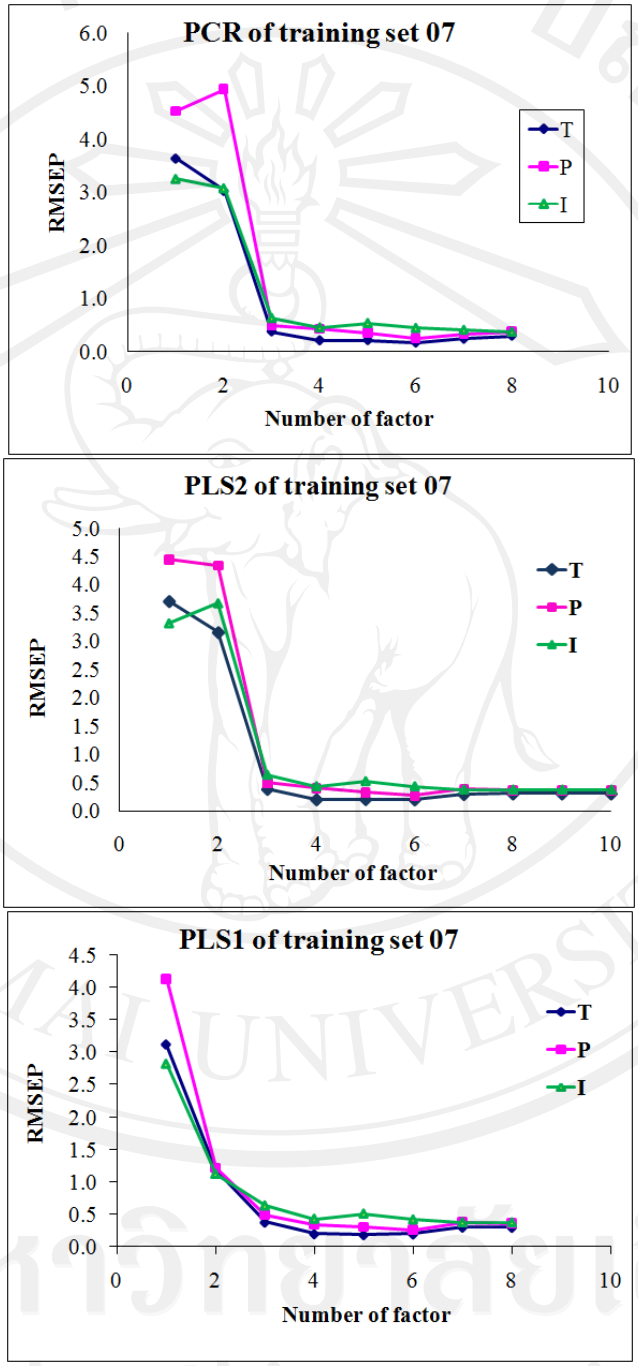

Figure 4.2 The number of factors and RMSD of PCR model (above), PLS1 model (middle) and PLS2 model (bottom) for the resolution of the ternary mixture: T, P, and I of training set 07

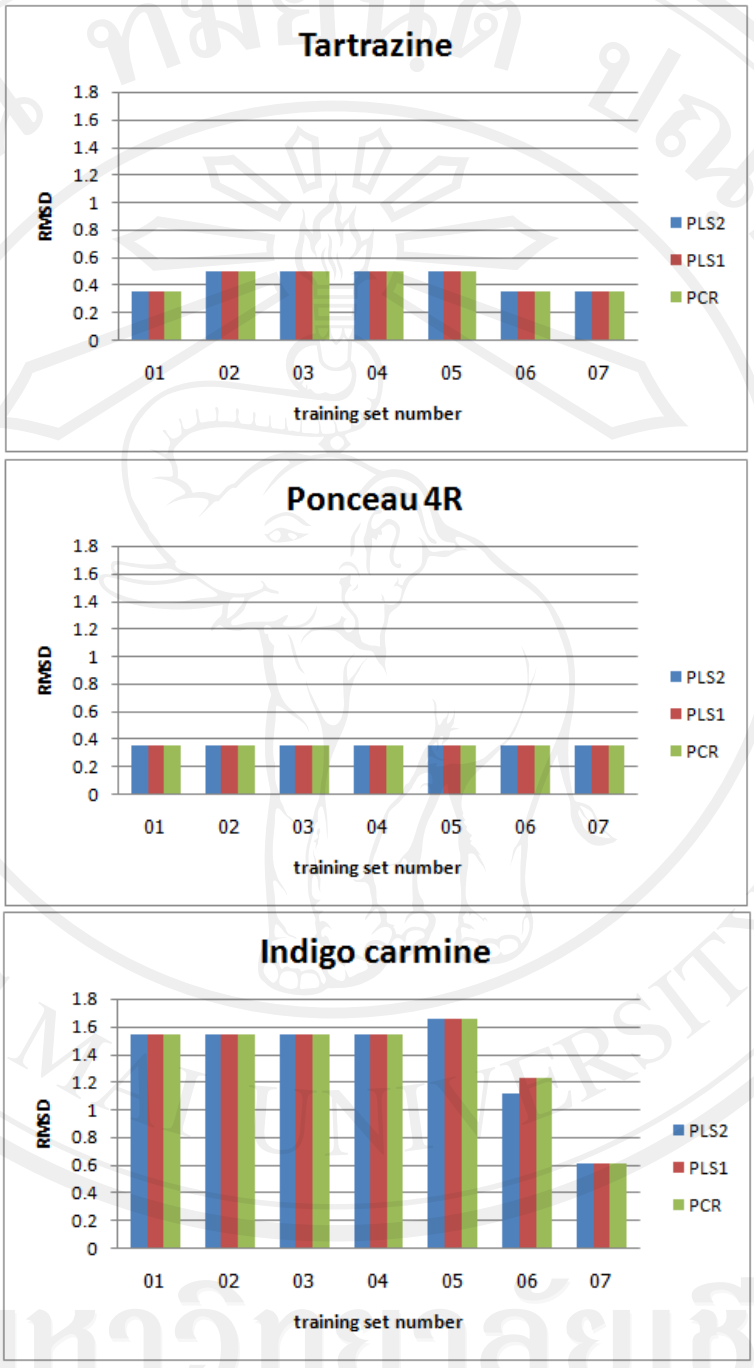

Figure 4.3 RMSEP of T,P and I of PCR, PLS1 and PLS2 model when using different training set (training set 01-07)

For all result, it was found that number of principal component has effect to quality of prediction more than calibration methodology. From internal and

external validation of PCR PLS1 and PLS2 of each training set, the prediction can be done within ±1, ±1 and ±2 ppm of actual concentration of T, P, and I, respectively. When PCR, PLS1 and PLS2 was use as tool for model calculation of training set 01-05 and 07, predicted concentration of T, P and I in unknown set were similar. Predicted concentrations of unknown mixtures by using training set 06 with PCR, PLS1 and PLS2 method mostly were similar except some outlier predicted concentrations. Predicted concentration of P of unknown number 5 and 16 when using PCR and PLS2 were similar but different from PLS1. Predicted concentration of I of unknown number 4, 10, 17, 20, 46, 53, 67, 73, 75, and 77 when using PCR and PLS1 were similar but different from PLS2. When number of component of each method by using training set 016 was considered (number of component of I, P and I of PCR, PLS1 and PLS2 were [3,3,5],[3,2,5], and [3,3,3], respectively), it was found that it correlated with the prediction results. The prediction of unknown by using training set 06 with 3 methods was the one evident for conclusion that number of component is effect to quality of prediction more than methodology. When complication of result management of each method was compared, it was found that PLS2 appropriate to use as multicalibration tool in this study.

4.3.2 Training set characteristic effect

4.3.2.1 Study of training set number effect when colorants concentration ranges was controlled

Quality of multicalibration model when using each training set; training set 01-07; was explained by R^2 and RMSEP value.

From Table 4.2, R^2 of prediction of T and P concentration when using training set 01-07 were more than 0.99 but R^2 of prediction of I are

different between training set 06 and the rest. R^2 of prediction of I concentration when using training set 06 was 0.967 while the rest has R^2 value more than 0.99. The R_2 value of PLS2 model by using training set 06 worse than the rest because of more scatter from bigger training set size.

From Figure 4.7, RMSEP of T, P and I by PLS2 model when using training set 01-07 were shown. RMSEP of P of all training sets was similar. RMSEP of T by PLS2 when using training set 02-05 was decreased from RMSEP of T of training set 01,while when RMSEP of T of training set 06-07 were not differently different from training set 01. RMSEP of I when using training set 01-05 Were compared, it was found that RMSEP of I of PLS2 from training set 01-04 were not significantly different, in another hand, RMSEP of I of PLS2 from training set 05 was increased. The RMSEP values of I of PLS2 from training set 06 and 07 were compared, it was found that when number of mixtures in training set was decrease from 35 mixtures to 10 mixtures, the RMSEP of I of PLS2 was decreased. From R2 and RMSEP of PLS2 of prediction of T, P and I by using training set 01-07,it was proved of possibility of training set number decreasing. The varying of training set which colorants concentration ranges was controlled had not effected to P concentration prediction but affected to P and I concentration prediction.

Table 4.4 Quality of PCR PLS1 and PLS2 when using training set 01-07

training set #	01			02			03			04			05			06			07		
method	PCR	PLS1	PLS2	PCR	PLS1	PLS2	PCR	PLS1	PLS2	PCR	PLS1	PLS2	PCR	PLS1	PLS2	PCR	PLS1	PLS2	PCR	PLS1	PLS2
number of component	3,3,3	3,3,3	3	3,3,3	3,3,3	3	3,3,3	3,3,3	3	3,3,3	3,3,3	3	3,3,3	3,3,3	3	3,3,5	3,2,5	3	3,3,3	3,3,3	3
R2_T	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.999	0.985	0.985	1	0.998	0.998	0.998	0.992	0.992	0.998
R2_P	1	1	1	1	1	1	1	1	1	0.999	0.999	1	0.999	0.999	1	0.999	0.998	0.999	0.994	0.994	0.998
R2_I	0.996	0.996	0.998	0.997	0.998	0.999	0.998	0.998	0.999	0.997	0.997	0.999	0.99	0.99	1	0.986	0.988	0.967	0.975	0.975	0.995
RMSD_T	0.095	0.095	0.095	0.107	0.107	0.107	0.136	0.136	0.136	0.188	0.188	0.188	0.621	0.621	0.621	0.178	0.178	0.178	0.381	0.38	0.381
RMSD_P	0.086	0.086	0.086	0.095	0.095	0.095	0.118	0.118	0.118	0.189	0.189	0.189	0.268	0.268	0.268	0.188	0.255	0.188	0.495	0.495	0.495
RMSD_I	0.227	0.227	0.227	0.166	0.166	0.166	0.178	0.178	0.178	0.214	0.214	0.214	0.464	0.464	0.464	0.433	0.403	0.764	0.632	0.632	0.632
error_T	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1
error_P	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1	±1
error_I	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2	±2

When number of training set was decreased to 5, errors from external prediction (RMSEP) of T and I were increased. Furthermore when PLS2 from training set 06-07 were considered, decreasing of training set number can improved quality of prediction. From external validation, it was found that, difference between predicted concentration and expected concentration of T, P and I were ±1, ±1, and ±2, respectively. When PLS2 model of each training set was performed, the unknown mixture that predicted within boundary was accepted, otherwise was rejected.

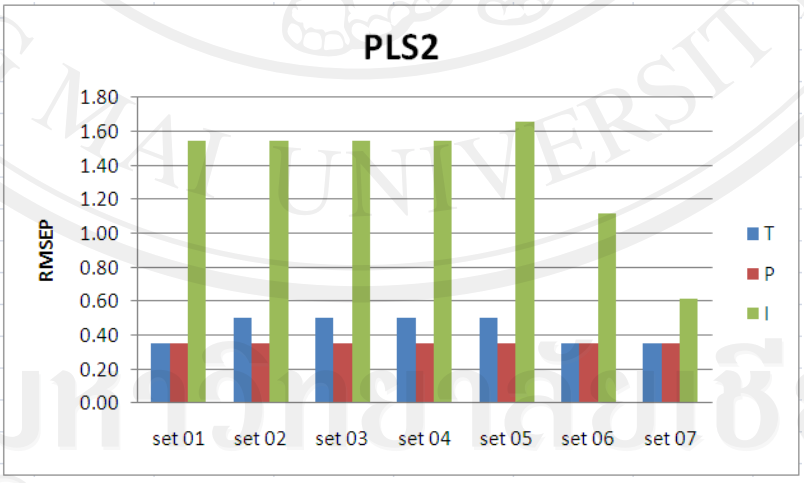

Figure 4.4 RMSEP of T, P and I of PLS2 model when using training set 01-07

From Table 4.3, predicted concentration of T, P and I of unknown mixtures from PLS2 model from each training set when concentration range of training set was used as criteria for to identify confidence of prediction from PLS2 mostly correlate with actual concentration of colorants in mixture except unknown number 10, 19 and 70. Components concentration in sample 10 and 19 were out of boundary of each training set but prediction of PLS2 from training set 06 was in boundary. Actual concentration of I of unknown number 70 was in boundary of all training set but prediction of I when using training set 01-05 were out of boundary. The study of number of training set effect was increase criteria and increase boundary of prediction to eliminate weak point of reliable of prediction of concentration near boundary.

Table 4.5 Unknown prediction of PCR, PLS1 and PLS2 when using training set 01-07(☒ is ejected prediction result, ☑ is accepted prediction result)

	expected concentration			PLS2-01			confidence	PLS2-02			confidence	PLS2-03			confidence	PLS2-04			confidence	PLS2-05			confidence	PLS2-06			confidence	PLS1-06			confidence	PCR-06			confidence	PLS2-07			confidence
#	T	P	I	T	P	I		T	P	I		T	P	I		T	P	I		T	P	I		T	P	I		T	P	I		T	P	I		T	P	I	
1	1	2	1	1	2	1	☒	1	2	1	☒	1	2	1	☒	1	2	1	☒	1	2	2	☒	1	2	1	☒	1	2	1	☒	1	2	1	☒	1	2	1	☒
2	2	4	6	2	4	8	☒	2	4	8	☒	2	4	8	☒	2	4	8	☒	2	4	8	☒	2	4	7	☒	2	4	7	☒	2	4	7	☒	2	4	6	☒
4	10	16	20	10	16	24	☒	10	16	24	☒	10	16	24	☒	9	16	24	☒	10	16	24	☒	10	16	23	☒	10	16	24	☒	10	16	24	☒	10	16	21	☒
5	20	10	5	19	10	6	☒	19	10	6	☒	19	10	6	☒	19	10	6	☒	19	9	6	☒	19	10	6	☒	19	9	6	☒	19	10	6	☒	19	10	5	☒
8	6	6	6	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	8	☑	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	6	☑
10	12	4	4	11	4	5	☒	11	4	5	☒	11	4	5	☒	11	4	5	☒	11	4	5	☒	11	4	5	☑	11	4	4	☒	11	4	4	☒	11	4	4	☒
12	16	20	8	16	20	10	☑	16	20	10	☑	16	20	10	☑	16	20	10	☑	16	20	10	☑	16	20	10	☑	16	20	10	☑	16	20	10	☑	16	20	9	☑
14	16	18	10	15	18	12	☑	15	18	12	☑	15	18	12	☑	15	18	12	☑	15	18	12	☑	16	18	12	☑	16	18	12	☑	16	18	12	☑	16	18	11	☑
15	20	6	6	19	6	7	☒	19	6	7	☒	19		8	☒	19	6	8	☒	20	6	8	☒	20	6	7	☒	20	6	7	☒	20	6	7	☒	20	6	7	☒
16	6	10	20	6	10	25	☒	6	10	25	☒	6	10	25	☒	6	10	25	☒	6	11	25	☒	6	10	23	☒	6	11	23	☒	6	10	23	☒	6	10	22	☒
17	4	2	2	4	2	3	☒	4	2	3	☒	4	2	3	☒	4	3	3	☒	4	2	4	☒	4	2	3	☒	4	2	2	☒	4	2	2	☒	4	2	2	☒
18	18	6	3	18	6	4	☒	18	6	4	☒	18	6	4	☒	18	6	4	☒	18	6	4	☒	18	6	4	☒	18	6	4	☒	18	6	4	☒	18	6	4	☒
19	4	4	4	4	4	5	☒	4	4	5	☒	4	4	5	☒	4	4	5	☒	4	4	6	☒	4	4	5	☑	4	4	5	☑	4	4	5	☑	4	4	4	☒
20	20	20	20	19	19	23	☒	19	19	23	☒	19	19	23	☒	18	19	23	☒	19	19	23	☒	19	19	22	☒	19	19	23	☒	19	19	23	☒	19	19	21	☒
46	2	2	2	2	2	2	☒	2	2	2	☒	2	2	2	☒	2	2	2	☒	2	2	2	☒	2	2	2	☒	2	2	1	☒	2	2	1	☒	2	2	1	☒
48	2	4	6	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	5	☒
49	8	4	4	7	4	4	☒	7	4	4	☒	7	4	4	☒	7	4	4	☒	7	4	4	☒	7	4	4	☒	7	4	4	☒	7	4	4	☒	7	4	3	☒
52	10	4	4	9	4	4	☒	9	4	4	☒	9	4	4	☒	9	4	4	☒	9	4	4	☒	9	4	4	☒	9	4	4	☒	9	4	4	☒	9	4	3	☒
53	2	4	10	2	4	10	☒	2	4	10	☒	2	4	10	☒	2	4	10	☒	2	4	10	☒	2	4	9	☒	2	4	10	☒	2	4	10	☒	2	4	8	☒
54	6	16	2	5	16	2	☒	5	16	2	☒	5	16	2	☒	5	16	2	☒	5	16	2	☒	5	16	2	☒	5	16	2	☒	5	16	2	☒	6	16	2	☒
55	16	12	4	14	12	4	☒	14	12	4	☒	14	12	4	☒	14	12	4	☒	14	12	4	☒	15	12	4	☒	15	12	4	☒	15	12	4	☒	15	12	4	☒
66	1	1	1	1	1	2	☒	1	1	1	☒	1	1	1	☒	1	1	1	☒	1	1	2	☒	1	1	1	☒	1	1	1	☒	1	1	1	☒	1	1	1	☒
67	30	30	30	29	30	31	☒	29	30	31	☒	29	30	31	☒	29	29	31	☒	29	30	31	☒	29	30	30	☒	29	30	27	☒	29	30	27	☒	30	30	29	☒
70	15	15	15	15	15	17	☒	15	15	17	☒	15	15	17	☒	14	15	17	☒	15	15	17	☒	15	15	16	☑	15	15	15	☑	15	15	15	☑	15	15	15	☑
71	15	20	15	15	20	16	☑	15	20	16	☑	15	20	16	☑	15	20	16	☑	15	20	16	☑	15	20	16	☑	15	20	14	☑	15	20	14	☑	15	20	15	☑
72	12	24	12	12	24	13	☒	12	24	13	☒	12	24	13	☒	12	24	13	☒	12	24	13	☒	12	24	13	☑	12	24	12	☑	12	24	12	☑	12	24	12	☑
73	16	26	6	16	26	6	☒	16	26	6	☒	16	26	6	☒	16	26	6	☒	16	26	6	☒	16	26	7	☒	16	26	6	☒	16	26	6	☒	16	26	7	☒
74	8	4	2	8	4	2	☒	8	4	2	☒	8	4	2	☒	8	4	2	☒	8	4	3	☒	8	4	2	☒	8	4	2	☒	8	4	2	☒	8	4	2	☒
75	14	7	10	14	7	11	☑	14	7	11	☑	14	7	11	☑	14	7	11	☑	14	7	11	☑	14	7	11	☑	14	7	10	☑	14	7	10	☑	14	7	10	☑
76	10	10	10	9	10	12	☑	9	10	12	☑	9	10	12	☑	9	10	12	☑	9	10	12	☑	9	10	12	☑	9	10	12	☑	9	10	12	☑	9	10	11	☑
77	0	0	10	0	0	12	☒	0	0	12	☒	0	0	12	☒	0	0	13	☒	0	0	13	☒	0	0	11	☒	0	0	12	☒	0	0	12	☒	0	0	10	☒
78	0	10	0	0	10	0	☒	0	10	0	☒	0	10	0	☒	0	10	0	☒	0	9	0	☒	0	10	0	☒	0	10	0	☒	0	10	0	☒	0	10	0	☒
79	10	0	0	10	0	0	☒	10	0	0	☒	10	0	0	☒	10	0	0	☒	10	0	0	☒	10	0	0	☒	10	0	0	☒	10	0	0	☒	10	0	0	☒

Range of concentration and concentration ratio of each concentration in mixture of training set were used as criteria for prediction, the result of study was shown in 4.3.2.2.

4.3.2.2 Study on training set number effect when colorants concentration ranges and colorants concentration ratios were controlled

Training set series (0A-0E) was shown in table 4.4; number of mixtures in training set was varied from 5-25 mixtures. Training set series that studied in 3.2.2 was selected by controlled ratio of components in mixture.

Concentration range of each colorant in training set was fixed at 4-20 ppm (Table 4.5). Ratio of each colorant in mixture of each training set was controlled to be equal. When ratio of [T]/[P]/[I] of range all training sets were fixed at 0.03-0.63,ratios of [T]/[P] and [T]/[I] of training set 0A-0D were same at 0.3-4.0 and ratios of [T]/[P] and [T]/[I] of training set 0E were 0.3-2.5 and 0.3-4.0, respectively.

Table 4.6 Quality of PLS2 when using training set 0A-0E

	PLS2-0A	PLS2-0B	PLS2-0C	**PLS2-0D**	PLS2-0E
R2_T	0.992	0.992	0.992	**0.995**	1
R2_P	0.999	0.999	0.999	**0.998**	1
R2_I	0.964	0.964	0.964	**0.978**	1
RMSD_T	0.45	0.498	0.549	**0.831**	0.506
RMSD_P	0.226	0.255	0.308	**0.465**	2.521
RMSD_I	1.093	1.217	1.361	**1.623**	0.157
number of component	3	3	3	**3**	3
error_T	±1	±1	±1	**±1**	±1
error_P	±1	±1	±1	**±1**	±1
error_I	±2	±2	±2	**±2**	±3

While [T]/[P]/[I], [T]/[P], [T]/[I] of each training set were mostly equal but [P]/[I] of each training set was different; [P]/[I] of training set 0A-0B was 0.3-3.3, training set 0C-0D was 0.3-3.0 and training set 0E was 0.5-2.5. RMSD of T and I were increased when number of training set was decreased from 25 to 10 except training set 0E had RMSD of T and I were not significantly different from training set 0A and 0B. RMSD of P was increasing when number of training set

number was increased from 25 to 5, PLS2 from training set 0E had greater RMSD of P more than the other. From external validation, it was found that, difference between predicted concentration and expected concentration of T, and P were ±1 ppm for all training set but difference between predicted concentration and expected concentration of I was ±2 ppm for PLS2 from training set 0A-0D and ±3 ppm from training set 0E.

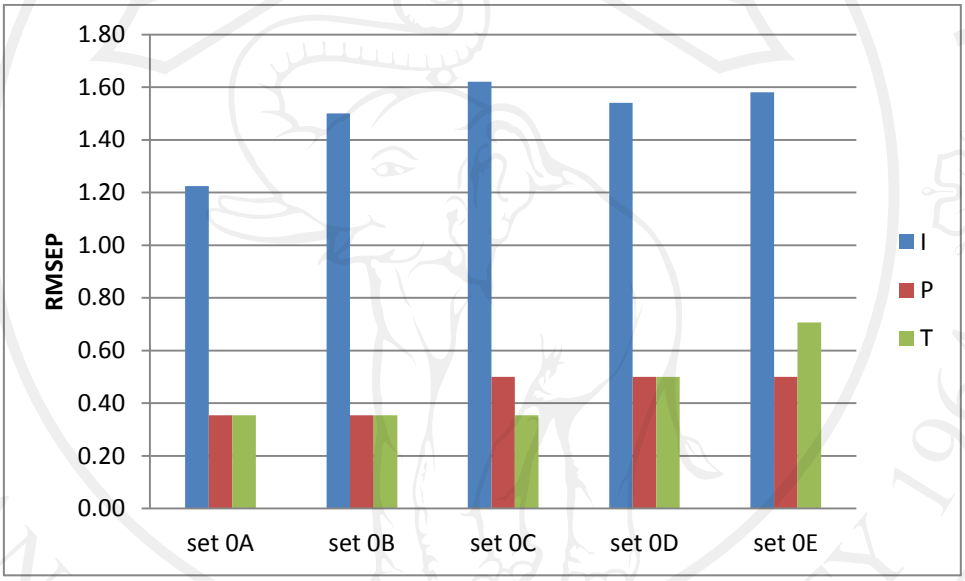

Figure 4.5 RMSEP of T, P and I of PLS2 model when using training set 0A-0E

From Figure 4.8, RMSEP from external validation of each colorant slightly increased within ±0.20 when training set number was decreased. Concentration of each colorant range and ratio of concentration of each colorant were used as criteria for prediction.

From Table 4.6, prediction of colorants in unknown mixture was presented. The colorants in unknown mixtures can be predicted within concentration error. Number of mixtures which were accepted was affected from working range of PLS2 model so accepted prediction of PLS2 from training set 0E prediction was less than other PLS2 models because of narrow acceptation range. Furthermore, mixture which contained of colorant concentration nearby working

range, sample no 18 and 74, was risk to predict uncorrelated with actual concentration of each colorant in the mixture. Actual concentration of colorants in mixture no 18 were within boundary of all training set. Concentration of I of sample no. 18 was nearby working range so the prediction when using PLS2 model from 0A-0D was accepted but rejected when using PLS2 from 0E because [T]/[I] from the prediction was more than working range. The prediction of sample no 74 was accepted when using PLS2 from training set 0C uncorrelated with expected concentration which I concentration was out of range. It was found that quality of prediction of PLS2 from training set 0D as good as PLS2 from training set 0A and was selected to use for study of signal characteristics effect.

4.3.3 Signal characteristics effect

Spectrum range and interval of spectrum of training set 0D were studied to improve quality of prediction. Three spectrum ranges were compared; i. Full spectrum range which contain variation of each colorant in UV and VIS range; 300-700 nm, ii. Main variation of each colorant; 370-700 nm and iii. UV range of each colorant variation; 400-700 nm. The quality of each PLS2 model when using different spectrum range was same when working range, validation and unknown prediction were considered in table 4.7-4.8.

Table 4.7 Unknown prediction of PLS2 when using training set 0A-0E (☒ is ejected prediction result, ☑ is accepted prediction result)

unknown	expected conc			PLS2-0A			confidence	PLS2-0B			confidence	PLS2-0C			confidence	PLS2-0D			confidence	PLS2-0E			confidence
	T	P	I	T	P	I		T	P	I		T	P	I		T	P	I		T	P	I	
1	1	2	1	1	2	3	☒	1	2	3	☒	1	2	3	☒	1	2	2	☒	2	3	2	☒
2	2	4	6	2	4	8	☒	2	4	8	☒	2	4	8	☒	2	4	7	☒	3	4	7	☒
3	10	10	10	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	10	☑
8	6	6	6	6	6	7	☑	6	6	8	☑	6	6	8	☑	6	6	7	☑	6	6	7	☑
9	8	10	12	7	9	12	☑	7	9	12	☑	7	9	12	☑	7	9	12	☑	7	9	11	☑
11	8	10	8	8	10	9	☑	8	10	9	☑	8	10	9	☑	8	10	9	☑	8	10	8	☑
12	16	20	8	16	20	10	☑	16	20	9	☑	16	20	9	☑	17	20	9	☑	16	20	9	☑
13	6	10	8	6	10	9	☑	6	10	9	☑	6	10	9	☑	6	10	9	☑	6	10	9	☑
14	16	18	10	16	18	11	☑	16	18	11	☑	16	18	11	☑	16	18	11	☑	16	18	10	☑
17	4	2	2	5	3	4	☒	5	3	4	☒	5	3	4	☒	5	3	4	☒	5	3	4	☒
18	18	6	3	18	6	5*	☑	19	6	5*	☑	19	6	5*	☑	19	7	5*	☑	19	7	4	☒
24	16	12	8	17	12	8	☑	17	12	8	☑	17	12	8	☑	17	12	8	☑	17	12	7	☑
30	12	8	6	12	8	7	☑	13	8	7	☑	13	8	7	☑	13	8	6	☑	13	8	6	☑
34	10	4	6	10	4	6	☑	11	4	6	☑	11	4	7	☑	11	4	6	☑	11	4	5	☑
46	2	2	2	2	2	3	☒	2	2	3	☒	2	2	3	☒	2	2	3	☒	2	2	3	☒
47	6	6	6	6	6	6	☑	6	6	7	☑	6	6	7	☑	6	6	6	☑	6	6	6	☑
48	2	4	6	2	4	6	☒	2	4	6	☒	2	4	7	☒	2	4	6	☒	2	4	6	☒
49	8	4	4	8	4	5	☑	8	4	5	☑	8	4	5	☑	8	4	5	☑	8	4	4	☑
53	2	4	10	2	4	10	☒	2	4	10	☒	2	4	10	☒	2	4	9	☒	2	4	9	☒
54	6	16	2	6	16	3	☒	6	16	3	☒	6	16	3	☒	6	16	2	☒	6	16	3	☒
56	10	10	10	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	10	☑	10	10	10	☑
57	10	4	6	10	4	7	☑	10	4	7	☑	10	4	7	☑	10	4	7	☑	11	4	6	☒
58	5	5	5	5	5	6	☑	5	5	6	☑	5	5	7	☑	5	5	6	☑	6	5	6	☑
59	5	10	15	5	10	16	☑	5	10	16	☑	5	10	16	☑	5	10	15	☑	5	10	15	☑
60	15	5	10	16	5	11	☑	16	5	11	☑	16	5	11	☑	16	5	11	☑	16	5	10	☒
62	10	5	10	10	5	11	☑	10	5	11	☑	10	5	11	☑	10	5	11	☑	11	5	10	☑
63	18	18	18	18	18	17	☑	18	18	17	☑	18	18	17	☑	18	18	17	☑	18	18	16	☑
64	10	15	5	10	15	6	☑	10	15	6	☑	10	15	6	☑	11	15	5	☑	10	15	6	☑
65	12	24	12	12	24	12	☒	12	24	12	☒	12	24	12	☒	12	24	11	☒	12	24	12	☒
66	1	1	1	1	1	3	☒	1	1	3	☒	1	1	3	☒	1	1	2	☒	2	2	2	☒
67	30	30	30	29	30	27	☒	29	30	27	☒	29	30	26	☒	30	30	26	☒	29	29	24	☒
68	8	8	8	8	8	9	☑	8	8	9	☑	8	8	9	☑	8	8	8	☑	8	8	8	☑
69	8	10	12	8	10	13	☑	8	10	13	☑	8	10	12	☑	8	10	12	☑	8	10	11	☑
70	15	15	15	15	15	15	☑	15	15	15	☑	15	15	15	☑	15	15	15	☑	15	15	14	☑
71	15	20	15	15	20	15	☑	15	20	15	☑	15	20	14	☑	16	20	14	☑	15	20	13	☑
72	12	24	12	12	24	12	☒	12	24	12	☒	12	24	12	☒	12	24	11	☒	12	24	12	☒
73	16	26	6	16	26	7	☒	16	26	7	☒	16	26	7	☒	17	26	6	☒	16	26	7	☒
74	8	4	2	8	4	3	☒	8	4	3	☒	8	4	4*	☑	9	4	3	☒	9	4	3	☒
75	14	7	10	14	7	11	☑	14	7	11	☑	14	7	11	☑	14	7	10	☑	15	7	9	☑
76	10	10	10	9	10	12	☑	9	10	12	☑	10	10	11	☑	10	10	11	☑	10	10	10	☑
77	0	0	10	0	0	11	☒	0	0	12	☒	0	0	12	☒	0	0	11	☒	0	0	11	☒
78	0	10	0	0	10	1	☒	0	10	1	☒	0	10	1	☒	0	10	1	☒	0	10	2	☒
79	10	0	0	10	0	1	☒	10	0	1	☒	10	0	2	☒	11	0	2	☒	11	0	1	☒

Table 4.8 Quality of PLS2 of training set 0D from varying of spectrum range and interval

	PLS2-0D_1	PLS2-0D_2	PLS2-0D_3	PLS2-0D_4	PLS2-0D_5	PLS2-0D_6	PLS2-0D_7	PLS2-0D_8
Wavelength range	300-700	370-700	400-700	300-700	300-700	300-700	300-700	300-700
interval	1	1	1	5	10	20	50	100
number of training set	10	10	10	10	10	10	10	10
R2_T	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R2_P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R2_I	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
RMSD_T	0.83	0.83	0.79	0.79	0.79	0.80	0.81	0.84
RMSD_P	0.47	0.47	0.48	0.48	0.47	0.47	0.47	0.47
RMSD_I	1.62	1.70	1.71	1.71	1.71	1.70	1.70	1.68
number of component	3	3	3	3	3	3	3	3
error_T	±1	±1	±1	±1	±1	±1	±1	±1
error_P	±1	±1	±1	±1	±1	±1	±1	±1
error_I	±2	±2	±2	±2	±2	±2	±2	±2

Spectrum range 400-700 nm was selected to decrease analysis time. Interval of spectrum 400-700 nm was varied; 1 5 10 20 50 100 nm-1.From the results in Table 4.7-4.8, validation and unknown prediction of all PLS2 models from training set 0D (when interval was varied) were not significant differently so spectrum range 400-700 nm with 100 nm^{-1} was selected to use as signal of PLS2 calibration. It was shown that analysis time can be reduced by decreasing of spectrum range to 400-700 nm and increasing interval to 100 nm^{-1} and quality of prediction was not significant differently.

Table 4.9 Unknown prediction of PLS2 of training set 0D from varying of spectrum range and interval (☒ is ejected prediction result, ☑ is accepted prediction result

unknown	expected conc.			PLS2-0D_1			confidence	PLS2-0D_2			confidence	PLS2-0D_3			confidence	PLS2-0D_4			confidence	PLS2-0D_5			confidence	PLS2-0D_6			confidence	PLS2-0D_7			confidence	PLS2-0D_8			confidence
	T	P	I	T	P	I		T	P	I		T	P	I		T	P	I		T	P	I		T	P	I		T	P	I		T	P	I	
1	1	2	1	1	2	2	☒	1	2	2	☒	1	2	2	☒	1	2	2	☒	1	2	2	☒	1	2	2	☒	1	2	2	☒	1	2	2	☒
2	2	4	6	2	4	7	☒	2	4	7	☒	2	4	7	☒	2	4	7	☒	2	4	7	☒	2	4	7	☒	2	4	7	☒	2	4	7	☒
3	10	10	10	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑
8	6	6	6	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	7	☑	6	6	7	☑
9	8	10	12	7	9	12	☑	7	9	12	☑	7	9	12	☑	7	9	12	☑	7	9	12	☑	7	9	12	☑	7	9	12	☑	7	9	12	☑
11	8	10	8	8	10	9	☑	8	10	9	☑	8	10	9	☑	8	10	9	☑	8	10	9	☑	8	10	9	☑	8	10	9	☑	8	10	9	☑
12	16	20	8	17	20	9	☑	17	20	9	☑	17	20	9	☑	17	20	9	☑	17	20	9	☑	17	20	9	☑	17	20	9	☑	17	20	9	☑
13	6	10	8	6	10	9	☑	6	10	9	☑	6	10	9	☑	6	10	9	☑	6	10	9	☑	6	10	9	☑	6	10	9	☑	6	10	9	☑
14	16	18	10	16	18	11	☑	16	18	11	☑	16	18	11	☑	16	18	11	☑	16	18	11	☑	16	18	11	☑	16	18	11	☑	16	18	11	☑
17	4	2	2	5	3	4	☒	5	3	4	☒	5	3	4	☒	5	3	4	☒	5	3	4	☒	5	3	4	☒	5	3	4	☒	5	3	4	☒
18	18	6	3	19	7	5	☑	19	6	5	☑	19	6	5	☑	19	6	5	☑	19	6	5	☑	19	6	5	☑	19	6	5	☑	19	6	5	☑
24	16	12	8	17	12	8	☑	17	12	8	☑	17	12	8	☑	17	12	8	☑	17	12	8	☑	17	12	8	☑	17	12	8	☑	17	12	8	☑
30	12	8	6	13	8	6	☑	13	8	6	☑	13	8	6	☑	13	8	6	☑	13	8	6	☑	13	8	6	☑	13	8	6	☑	13	8	6	☑
34	10	4	6	11	4	6	☑	11	4	6	☑	11	4	6	☑	11	4	6	☑	11	4	6	☑	11	4	6	☑	11	4	6	☑	11	4	6	☑
46	2	2	2	2	2	3	☒	2	2	3	☒	2	2	3	☒	2	2	3	☒	2	2	3	☒	2	2	3	☒	2	2	3	☒	2	2	3	☒
47	6	6	6	6	6	6	☑	6	6	6	☑	6	6	6	☑	6	6	6	☑	6	6	6	☑	6	6	6	☑	6	6	6	☑	6	6	6	☑
48	2	4	6	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒	2	4	6	☒
49	8	4	4	8	4	5	☑	8	4	5	☑	8	4	5	☑	8	4	5	☑	8	4	5	☑	8	4	5	☑	8	4	5	☑	8	4	5	☑
53	2	4	10	2	4	9	☒	2	4	9	☒	2	4	9	☒	2	4	9	☒	2	4	9	☒	2	4	9	☒	2	4	9	☒	2	4	9	☒
54	6	16	2	6	16	2	☒	6	16	3	☒	6	16	3	☒	6	16	3	☒	6	16	3	☒	6	16	3	☒	6	16	3	☒	6	16	3	☒
56	10	10	10	10	10	10	☑	10	10	10	☑	10	10	10	☑	10	10	10	☑	10	10	10	☑	10	10	10	☑	10	10	10	☑	10	10	10	☑
57	10	4	6	10	4	7	☑	10	4	7	☑	10	4	7	☑	10	4	7	☑	10	4	7	☑	10	4	7	☑	10	4	7	☑	10	4	7	☑
58	5	5	5	5	5	6	☑	5	5	6	☑	5	5	6	☑	5	5	6	☑	5	5	6	☑	5	5	6	☑	5	5	6	☑	5	5	6	☑
59	5	10	15	5	10	15	☑	5	10	15	☑	5	10	15	☑	5	10	15	☑	5	10	15	☑	5	10	15	☑	5	10	15	☑	5	10	15	☑
60	15	5	10	16	5	11	☑	16	5	11	☑	16	5	11	☑	16	5	11	☑	16	5	11	☑	16	5	11	☑	16	5	11	☑	16	5	11	☑
62	10	5	10	10	5	11	☑	10	5	11	☑	10	5	11	☑	10	5	11	☑	10	5	11	☑	10	5	11	☑	10	5	11	☑	10	5	11	☑
63	18	18	18	18	18	17	☑	18	18	17	☑	18	18	17	☑	18	18	17	☑	18	18	17	☑	18	18	17	☑	18	18	17	☑	18	18	17	☑
64	10	15	5	11	15	5	☑	11	15	5	☑	11	15	5	☑	11	15	5	☑	11	15	5	☑	11	15	5	☑	11	15	5	☑	10	15	5	☑
65	12	24	12	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒
66	1	1	1	1	1	2	☒	1	1	2	☒	1	1	2	☒	1	1	2	☒	1	1	2	☒	1	1	2	☒	1	1	2	☒	1	1	2	☒
67	30	30	30	30	30	26	☒	30	30	26	☒	30	30	26	☒	30	30	26	☒	30	30	26	☒	30	30	26	☒	30	30	26	☒	30	30	26	☒
68	8	8	8	8	8	8	☑	8	8	9	☑	8	8	9	☑	8	8	9	☑	8	8	9	☑	8	8	9	☑	8	8	9	☑	8	8	9	☑
69	8	10	12	8	10	12	☑	8	10	12	☑	8	10	12	☑	8	10	12	☑	8	10	12	☑	8	10	12	☑	8	10	12	☑	8	10	12	☑
70	15	15	15	15	15	15	☑	15	15	15	☑	15	15	15	☑	15	15	15	☑	15	15	15	☑	15	15	15	☑	15	15	15	☑	15	15	15	☑
71	15	20	15	16	20	14	☑	16	20	14	☑	16	20	14	☑	16	20	14	☑	16	20	14	☑	16	20	14	☑	16	20	14	☑	16	20	14	☑
72	12	24	12	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒	12	24	11	☒
73	16	26	6	17	26	6	☒	17	26	6	☒	17	26	6	☒	17	26	6	☒	17	26	6	☒	17	26	6	☒	17	26	6	☒	17	26	6	☒
74	8	4	2	9	4	3	☒	9	4	3	☒	9	4	3	☒	9	4	3	☒	9	4	3	☒	9	4	3	☒	9	4	3	☒	8	4	3	☒
75	14	7	10	14	7	10	☑	14	7	10	☑	14	7	10	☑	14	7	10	☑	14	7	10	☑	14	7	10	☑	14	7	10	☑	14	7	10	☑
76	10	10	10	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑	10	10	11	☑
77	0	0	10	0	0	11	☒	0	0	11	☒	0	0	11	☒	0	0	11	☒	0	0	11	☒	0	0	11	☒	0	0	11	☒	-1	0	11	☒
78	0	10	0	0	10	1	☒	0	10	1	☒	0	10	1	☒	0	10	1	☒	0	10	1	☒	0	10	1	☒	0	10	1	☒	0	10	1	☒
79	10	0	0	11	0	2	☒	11	0	2	☒	11	0	2	☒	11	0	2	☒	11	0	2	☒	11	0	2	☒	11	0	2	☒	10	0	2	☒

4.4 Conclusions

Assay of tartrazine, ponceau 4R and indigo carmine (T, P and I) in mixture by no need of chemical separation can be done by using PLS2 method. Effects of calculation method, structure and number of training set, and signal characteristic were studied led to decrease analysis time. PCR, PLS10 and PLS2 by using training set 01-07 were compared. PLS2 was selected because it was easier than other methods to manage prediction result. Number of training set from 5-25 mixtures was varied by

comparing between controlled of colorants concentration ranges and with or without colorants concentration ratios. It was found that, quality of prediction by the training set number can be decrease to 10 mixtures when .colorants concentration ranges and with colorants concentration ratios was controlled (training set 0D) was as well as 25 mixtures. Furthermore, 100 nm^{-1} interval of spectrum was selected to decrease analysis time. Colorants in mixture can be determined by PLS2 from training set 0D which 400-700 nm, 100nm^{-1} interval. Colorants concentration ranges (4-20 ppm of T, P and I) and colorants concentration ratios (0.03-0.63, 0.3-4.0,0.3-4.0 and 0.3-3.0 for [T]/[P]/[I], [T]/[P] ,[T]/[I] ,[P]/[I], respectively) were used as criteria for the determination of T, P and I in mixture with error ranges of prediction were ±1, ±1, and ±2 for T, P and I, respectively.

4.5 References

1 G. K. Karaoglan, G. Gumrukcu, M. U. Ozgur, A. Bozdogan, B. Asci, *Anal. Lett.* **2007**, *40*, 1893-1903.

2 J. J. B. Nevado, J. R. g. Flores, M. J. V. Llerena, N. R. g. Fariñas, *Talanta* **1999**, *48*, 895-903

3 M. Puttemans, L. Dryon, D. Massart, *J. Assoc. Anal.Chem.* **1982**, *65*, 737.

4 M. Puttemans, L. Dryon, D. Massart, *J. Assoc. Anal.Chemical Papers* **1983**, *66*, 1039.

5 J. J. B. Nevado, J. R. Flores, M. J. V. Llerena, *Talanta* **1997**, *44*, 467-474.

6 J. J. B. Nevado, J. R. Flores, G. C. Peñalvo, *Analytica Chimica Acta* **1997**, *340*, 257-265.

7 J. J. B. Nevado, C. G. Cabanillas, A. C. Salcedo, *Analusis* **1994**, *22*, 5.

8 D. Haaland, E. Thomas, *Anal. Chem* **1988**, *60*, 1193.

9 D. Haaland, E. Thomas, *Anal. Chem.* **1990**, *62*, 1091.

10 H. Wold, *Research Papers in Statistics*. Editor, Wiley, New York, **1966**.

11 H. Jores-Kong, H. Wold, *Systems Under Indirect Observation vol. 2*. Editor, Amsterdam, North-Holland, **1982**.

12 H. Wold, H. Martens, *Multivariate CalibrationProblems in Chemistry Solved by PLS*. Editor, A. Ruhe, B. Kagstrom (Eds.), Heidelberg, **1983**.

13 M. Ma, X. Luo, B. Chen, S. Su, S. Yao, *Journal of Chromatography A* **2006**, *1103*, 170-176.

14 C. Bosch Ojeda, F. Sanchez Rojas, *Analytica Chimica Acta* **2004**, *518*, 1-24.

15 N. E. Llamas, M. Garrido, M. S. D. Nezio, B. S. F. Band, *Analytica Chimica Acta* **2009**, *655*, 38-42.

16 J. J. B. Nevado, C. G. Cabanillas, A. M. C. Salcedo, *Talanta* **1995**, *42*, 2043-2051.

17 L. F. Capitán-Vallvey, N. N. Iglesias, I. de Orbe Payá, R. A. Castaneda, *Microchimica Acta* **1997**, *126*, 153-157.

18 E. Dinç, E. Baydan, M. Kanbur, F. Onur, *Talanta* **2002**, *58*, 579-594.

19 D. M. Haaland, E. V. Thomas, *Analytical Chemistry* **1988**, *60*, 1193-1202.

20 E. V. Thomas, D. M. Haaland, *Analytical Chemistry* **1990**, *62*, 1091-1099.

21 Y. Ni, G. Zhang, S. Kokot, *Food Chemistry* **2005**, *89*, 465-473.