

## Original Research Article

# Comparative investigation of two different self-organizing map-based wavelength selection approaches for analysis of binary mixtures with strongly overlapping spectral lines

Lawan Srattaphut<sup>1\*</sup>, Sathit Niratisai<sup>2</sup> and Onoomar Toyama<sup>2</sup>

<sup>1</sup>Department of Health-related Informatics, <sup>2</sup>Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Silpakorn University, Nakhon Pathom, Thailand

\*For correspondence: **Email:** [srattaphut\\_l@su.ac.th](mailto:srattaphut_l@su.ac.th); **Tel:** +6634-253910

Sent for review: 1 February 2017

Revised accepted: 18 June 2017

### Abstract

**Purpose:** To demonstrate the ability and investigate the performance of two different wavelength selection approaches based on self-organizing map (SOM) technique in partial least-squares (PLS) regression for analysis of pharmaceutical binary mixtures with strongly overlapping spectra.

**Methods:** Two different variable selection methods were compared, namely, SOM1-PLS and SOM2-PLS. The main difference between these methods involved the structure of neurons in input layer and the algorithm for variable selection. Adjustable parameters for each technique were optimized for better comparison. The performance of these methods was statistically verified for predictive ability using both synthetic mixtures and a real combination product of sulfamethoxazole (SMX) and trimethoprim (TMP), which exhibited strongly overlapping of spectral lines.

**Results:** The results obtained indicate that SOM2-PLS was more efficient than SOM1-PLS technique with 30 and 6 % improvement in predictive ability for SMX and TMP, respectively. Furthermore, the mean difference between the results obtained from SOM2-PLS method and those from the official method was not statistically significant as p-value was more than 0.01.

**Conclusion:** Although, SOM2-PLS method is more efficient than SOM1-PLS method for the analysis of pharmaceutical binary mixtures with severely overlapping spectra, some problems associated with SOM2-PLS technique include difficult computations of some parameters.

**Keywords:** Co-trimoxazole, Self-organizing map, Wavelength selection, Pharmaceutical analysis, Overlapping spectral lines

Tropical Journal of Pharmaceutical Research is indexed by Science Citation Index (SciSearch), Scopus, International Pharmaceutical Abstract, Chemical Abstracts, Embase, Index Copernicus, EBSCO, African Index Medicus, JournalSeek, Journal Citation Reports/Science Edition, Directory of Open Access Journals (DOAJ), African Journal Online, Bioline International, Open-J-Gate and Pharmacy Abstracts

## INTRODUCTION

In modern analytical methods, using multiple component spectral analysis plays an important role. Various wavelengths are measured with minimal effort measuring per test and yield multivariate information. There are several literatures reporting the improvement of predictive ability when wavelength choosing is accomplished before multivariate calibration [1-3] and describing the different strategies of

wavelength selection [4-6]. In the previous two decades, a standout among fascinating variable selection approaches was self-organizing map (SOM).

Self-organizing map (commonly called Kohonen neural network) was first developed by Kohonen in 1975 [7]. SOM algorithm is based on competitive ('Winner Takes All') and unsupervised learning. The concept of SOM is that it tries to map the similar inputs to similar or

nearby map unit (neuron position or node). Two significant wavelength selection approaches based on SOM technique have been found useful in pharmaceutical field [8-11]. In general, SOM strategy can be applied in both space (sample and variable spaces). The first method (SOM1) has been proposed to specify wavelengths by clustering of variables, while, the second method (SOM2) has been proposed to select wavelengths according to the measure of topological relevance (MTR) on SOM's array which is generated from grouping a set of samples.

The strategies of wavelength selection for SOM1 and SOM2 were introduced by Todeschini *et al* [8,9] and Corona *et al* [10,11], respectively. Because of the quick computation of the analysis results of enormous samples, chemometric methods for spectroscopy are still today's common multivariate method seen in in-process control in pharmaceutical industry. This work was highlighted to support the analyst's decision when developing new analytical methods. The work also demonstrated the ability and investigated the performance of these two SOM methods for wavelength selection using partial least-squares (PLS) as linear regression and co-trimoxazole formulation as an example of pharmaceutical binary mixtures with strongly overlapping spectra. Co-trimoxazole formulation is commonly used for the treatment of pneumocystis pneumonia and for the prophylaxis against this disease in people with HIV/AIDS-defining infections [12]. Co-trimoxazole preparation also provides almost 100 % protection against malaria [13].

## EXPERIMENTAL

### Reagents and chemicals

The analytical-grade reagents and pharmaceutical-grade chemicals (SMX 99.45 % and TMP 99.41 %) were employed for all investigations. The compounds and co-trimoxazole tablets were provided by Sea Pharm Co. Ltd, Thailand. Co-trimoxazole tablets contained 80 mg of trimethoprim (TMP) and 400 mg of sulfamethoxazole (SMX). The stock solutions of SMX (100 mg/L) and TMP (25 mg/L) were accurately prepared in 95% ethanol.

### Apparatus and software

All absorption measurements were recorded on a double-beam UV spectrophotometer (UV160A, Shimadzu) attached to a computer loaded with UVPC software, using quartz cuvettes with a 10-

mm path length. The apparatus parameters which were used to determine the absorption spectra of all standard and test solutions included the following: wavelength range: 200 – 400 nm; spectral sampling interval: 1 nm. The data measured were processed by a workstation equipped with Intel Core i3 processor, 2 GB RAM, and Windows 7 operating system. PLS modelling, SOM1 and SOM2 wavelength selection were achieved by PLS\_Toolbox [14], kohonen\_cpnn\_toolbox 2.0 [15] and SOM\_Toolbox [16] respectively. All toolboxes were run under MATLAB [17].

Chromatographic evaluations were operated on a high-performance liquid chromatography (LC-20AT, Shimadzu), connected to a 3.9-mm x 30-cm column with ODS (C-18) packing material, 254-nm UV detector, single pump and manual injector (20 µl). The mobile phase was prepared by mixing of water, acetonitrile, and triethylamine (1400:400:2), and also adjusted to a pH of 5.9 ± 0.1.

### Calibration and validation sets

Twenty-two standard samples in calibration set were prepared from the combination of 12 samples corresponding to central composite design (CCD) with four center points and 10 samples of pure substances. For model validation, a validation set of nine samples was also built by randomization. The standard concentrations used within the calibration set and validation set were spread over the linearity ranges of 8.00 - 24.23 and 2.00 - 6.22 mg/L for SMX and TMP, respectively. They are shown in Table 1.

### Commercial sample preparation

The commercial co-trimoxazole tablets containing a combination of 80 mg TMP and 400 mg SMX were studied. Twenty tablets were weighed, finely powdered and blended in a mortar. An accurately weighed quantity of finely powdered tablet equivalent to about 50 mg of SMX was transferred to a 50-ml volumetric flask containing approximately 30 ml of 95 % ethanol. Subsequent to being shaken by sonication for 30 min, the volume was completed to 50 ml with the same solvent, and the solid was filtered. Then, 1 ml of clear aliquot was pipetted into a 25-ml volumetric flask containing 5 ml of ammonia-ammonium chloride buffer (pH 10.64) and made up to volume with water. The preparation was analyzed in six replicates.

**Table 1:** Combination of SMX and TMP in calibration set corresponding to central composite design

No.	Code		Concentration (mg/L)	
	SMX	TMP	SMX	TMP
1	+1	+1	21.66	5.41
2	+1	-1	21.66	2.59
3	-1	-1	10.34	2.59
4	-1	+1	10.34	5.41
5	0	0	16.00	4.00
6	+1.414 <sup>a</sup>	0	24.00	4.00
7	-1.414	0	8.00	4.00
8	0	+1.414	16.00	6.00
9	0	-1.414	16.00	2.00
10	0	0	16.00	4.00
11	0	0	16.00	4.00
12	0	0	16.00	4.00
13	-1.414	-	8.00	0
14	-1	-	10.34	0
15	0	-	16.00	0
16	+1	-	21.66	0
17	+1.414	-	24.00	0
18	-	-1.414	0	2.00
19	-	-1	0	2.59
20	-	0	0	4.00
21	-	+1	0	5.41
22	-	+1.414	0	6.00

<sup>a</sup>Star point of 2 factors in CCD ( $\alpha = \pm 1.414$ )

### Wavelength selection by SOM methods

SOM is a kind of artificial neural networks that provides a topology preserving mapping from a space of many dimensions onto a space of few dimensions that is ordinarily two dimensions (2D). In this study, SOM comprised two layers of neurons (nodes or cells): a one-dimensional input layer and a two-dimensional output layer (also called output map). Input layer was assigned to receive and transmit input data while output layer was responsible for competitive learning that held all weight vectors. The connection of both layers would be feed-forward. SOM methods had the following steps.

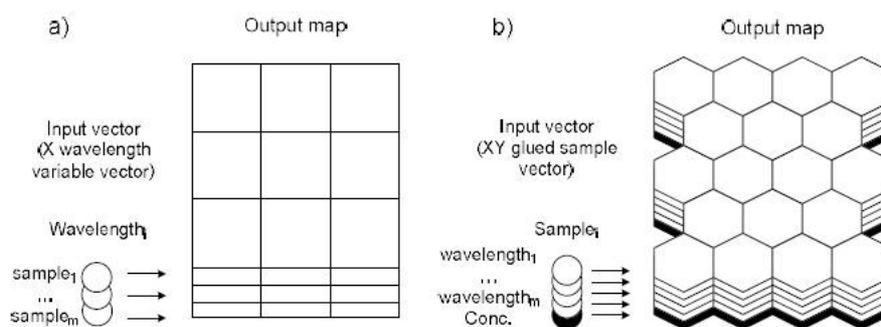
### Step 1: Input data preparation and preprocessing by mean-centering

According to SOM1 method, each wavelength of all samples in the calibration set was assigned to an input vector, leading to an input layer made of 101 input vectors (210 - 310 nm). Therefore, each input vector composed of 22 components. Also, the output map was characterized by being a squared toroidal lattice that consisted of a network of  $N^2$  neurons, where  $N$  is the number of cells for each side of network.

According to SOM2 method, the absorbance values of all wavelength glued together with concentration value were assigned to an input vector (observation sample), leading to an input layer made of 22 input vectors. Hence, each input vector contained 102 components (101 wavelengths + 1 concentration). Furthermore, the output map was produced using hexagonal toroidal grid of neurons initialized in the space spanned by the eigenvectors corresponding to the two largest eigenvalues of the covariance matrix of data [11]. The map size was computed from the ratio between these eigenvalues. The schema of input vector and output map of SOM1 and SOM2 are shown in Figure 1.

### Step 2: SOM algorithm processing by SOM software

There were five major steps which were regularly executed within each SOM software. First, initiation step, the weight vectors were initiated through a principal component analysis applied to the data or small random values. Consequently, all weights and input vectors were normalized to standard length 1. Second, competitive learning step, an input vector was selected by randomization from the input data set and, Euclidian distances between this input vector and all weight vectors (neurons) were calculated. The neuron with the shortest Euclidian distances was designated as the winner.



**Figure 1:** Schema of self-organizing map. a) SOM1 method; b) SOM2 method

Third, cooperation step, the winning node identified the spatial area of a topological neighborhood of excited neurons, consequently providing the basis for cooperation among neighboring neurons. Fourth, adaptation step, the weights of winner neuron and the weights of all neurons (neighborhood) existing in a neighboring area around the winner were updated. Finally, iteration step, the next input vector was fed and the process was repeated. The weight were modified until no further change in the output map or observation of some other termination conditions. At the end of SOM process, output map was built up.

**Step 3: Subset selection of wavelengths among the 101 full spectrum wavelengths**

In accordance with SOM1 method, the relative wavelengths, which fell in the same place in output map, carried the same information. As was described in paper [9], the wavelength closest to each neuron centroid was selected as the representative among all the wavelengths within the same neuron and it was subjected for regression analysis.

In accordance with SOM2 method, the Unified-distance matrix (U-matrix) of each component variables (wavelength and concentration variables) were calculated independently along each direction of the data space [10]. MTR or distance (D) between each wavelength-concentration pair of component U-matrices was estimated using Frobenius norm method as in Eq 1 [10].

$$D(x_j, y) = \|U_{x_j} - U_y\|_F \dots\dots\dots (1)$$

where  $U_{x_j}$  is U-matrices of wavelength component,  $U_y$  the U-matrix of concentration component, and  $\| \cdot \|_F$  the Frobenius norm. Frobenius norm was used to determine the closeness between U-matrices. The closer to zero, the more relevant measure of two components is [10], which means that the wavelength was suitable for reconstructing the concentration. As stated by Corona *et al*, MTR values should be inverted (which is called topological relevance index) in order to clearly represent the relevancies, with larger values indicating stronger relevancy. The median cut-off of topological relevance index was used to perform wavelength selection. A more detail of SOM1 and SOM2 method can be found in [8,9] and [10,11], respectively.

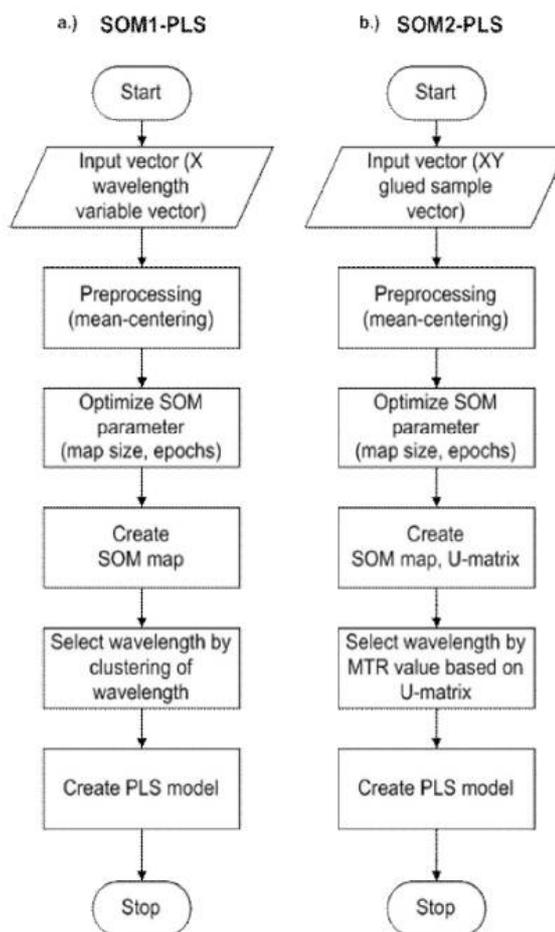
**PLS modeling**

The flowchart of the development process of

SOM1-PLS and SOM2-PLS models were shown in Figure 2. In calibration phase, the absorbance data from each SOM-based wavelength selection strategy were processed by PLS\_toolbox program to develop each PLS calibration model. Before processing, the searching for suitable number of factors was done by leave-one-out cross-validation method described by Haaland and Thomas [18]. In prediction phase, each PLS calibration model with optimum number of factors was applied for estimation of each compound in validation samples and pharmaceutical tablets. Then, the performance of each calibration model was measured by the mean score of percentage recovery and the root mean square error in prediction (RMSEP). The value of RMSEP was calculated as in Eq 2.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \dots\dots\dots (2)$$

where  $n$  is the sample size of the validation set.



**Figure 2:** Flow chart showing development process of two different SOM-PLS models

## RESULTS

### Spectral characteristics

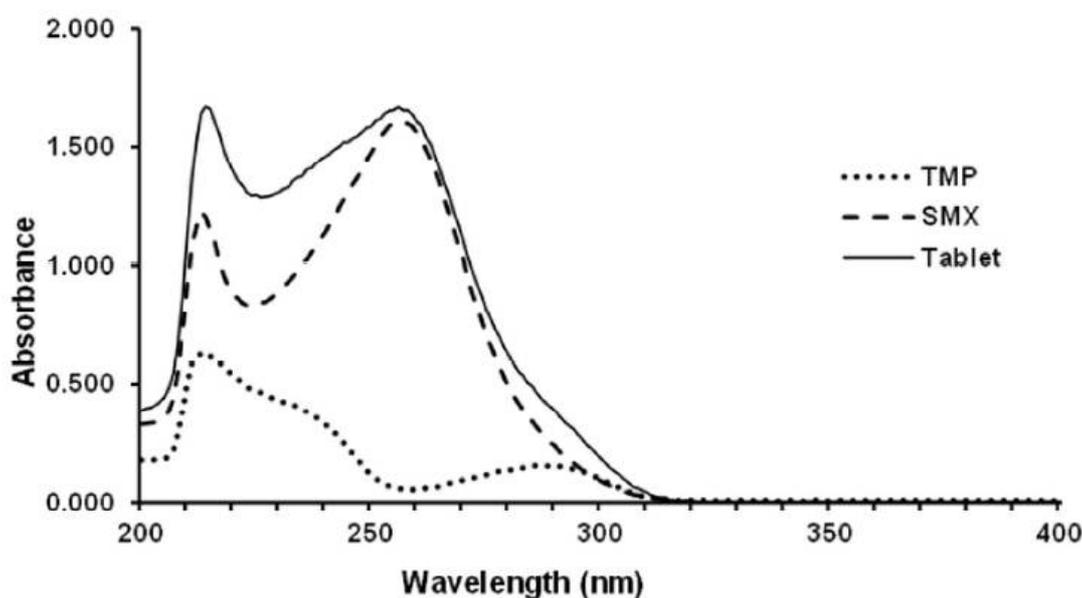
Before data collection, the impact of noises and solvent on the absorption spectra of all active ingredients and their mixtures was studied. Figure 3 indicated the strongly overlapping spectra of solutions of SMX and TMP over the UV region between 200 and 400 nm that might cause problems for regression analysis. Hence, the informative wavelength region should be selected before analysis. Moreover, in the preliminary step, the absorption spectra in the wavelength range between 210 and 310 nm were selected due to a cut-off value of solvent and low absorbance value (under 0.05) of points in the lower and higher end of the range 210 - 310 nm.

### SOM-PLS modeling

For better comparison, SOM was adopted in advance to select the wavelengths for PLS analysis and the adjustable parameters of SOM including the number of training loops (epochs) and the SOM map size were optimized. A test set of five synthetic mixtures was employed for

optimization of these parameters by trial and error.

To optimize and develop SOM1-PLS model, the map sizes were varies with different size: 4 x 4, 8 x 8, 12 x 12 and 16 x 16, and for each map size, the number of training epochs were examined with different epochs: 200, 500, 1000 and 2000. In this manner, 16 networks were produced. The subsets of wavelengths were chosen at each set of parameters and were transmitted into PLS algorithm. SOM1-PLS calibration models for each substance were developed and applied to test set. The mean percentage recovery and RMSEP were calculated for measuring the performance of each calibration model and the results were illustrated in Figure 4. The sets of parameters with minimum RMSEP and the percentage recovery closer to 100 were chosen. The summary specifications for SOM1-PLS models were presented in Table 2. The distribution of 101 wavelengths into 8x8 sized map at the number of training epochs 2000 was shown in Figure 5. The 39 selected wavelengths used for constructing the models of both compounds were presented in Table 3.



**Figure 3:** Absorption spectra of SMX (25.0 mg/L), TMP (5.0 mg/L) and commercial tables

**Table 2:** Specifications and optimized parameters of SOM-PLS models

Parameter	SOM1-PLS		SOM2-PLS	
	SMX	TMP	SMX	TMP
Map size	8x8	8x8	39x3	36x3
Number of training epochs	2000	2000	500	200
Number of wavelengths	39	39	51	53
Number of factors	4	6	3	5

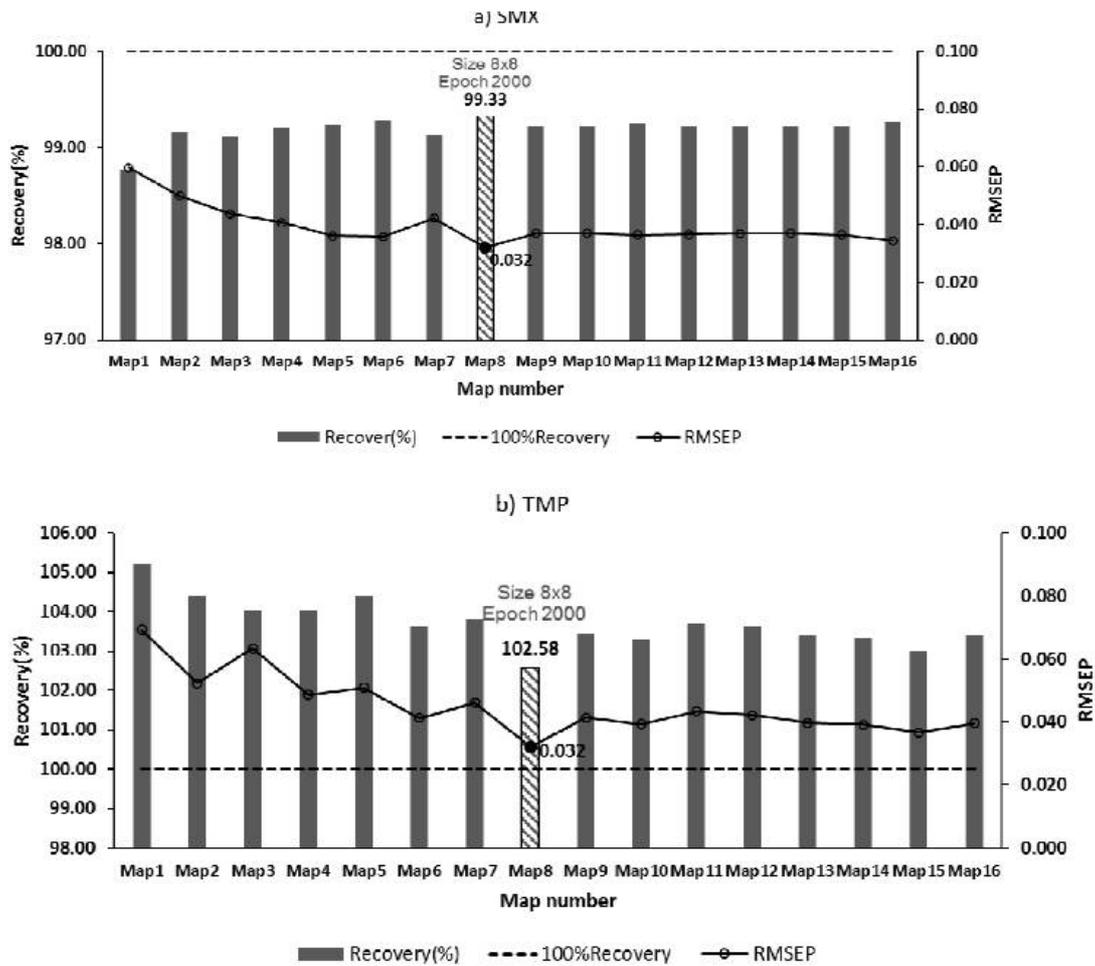


Figure 4: Recovery (%) and RMSEP of SOM1-PLS models. a) SMX; b) TMP

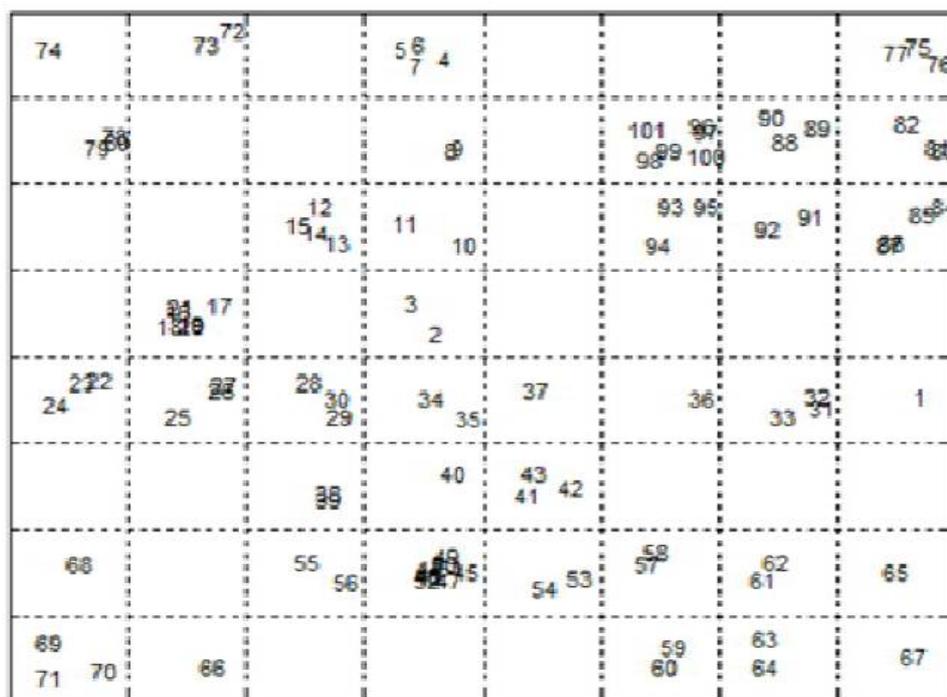


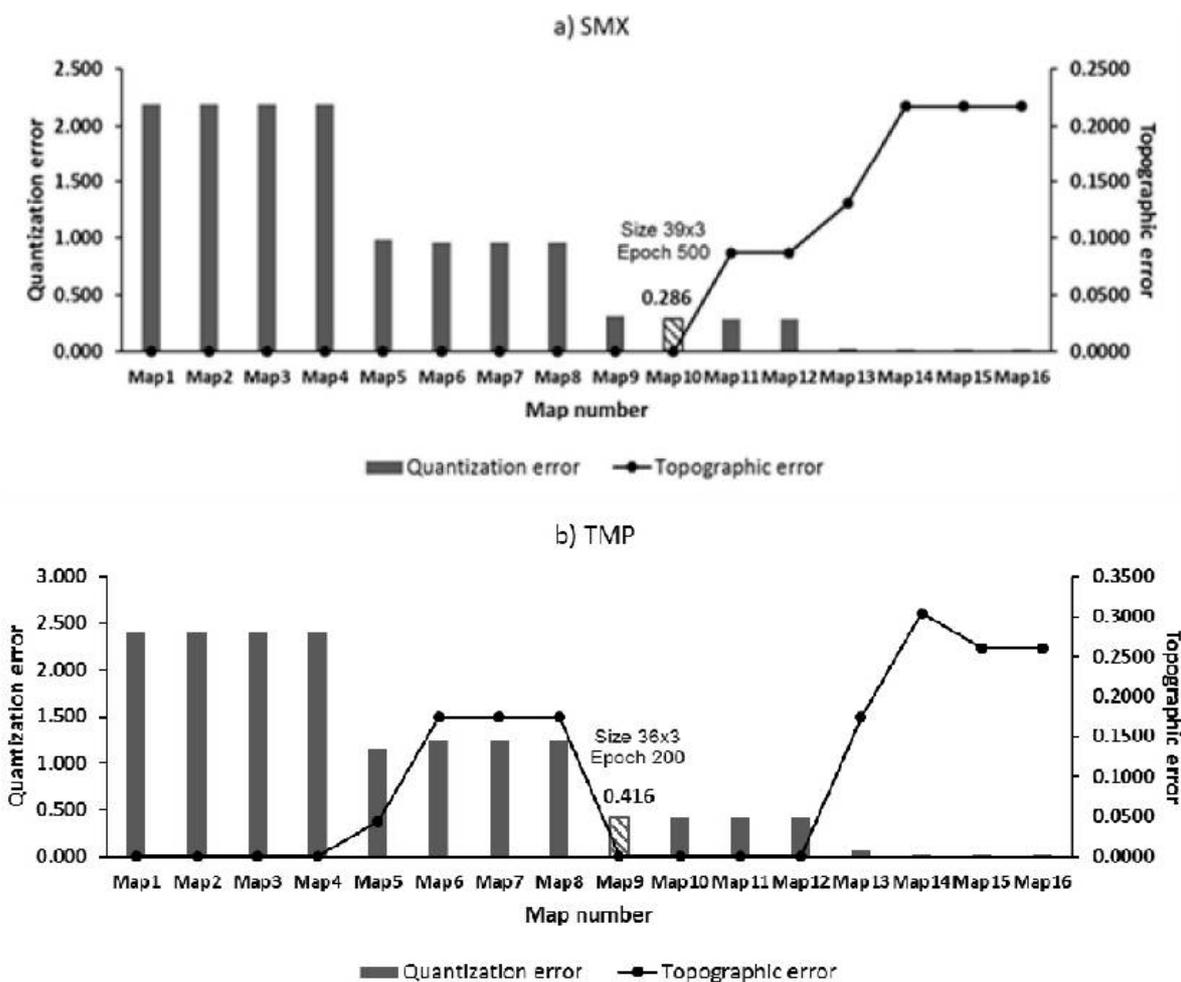
Figure 5: Distribution of 101 wavelengths into 8x8 map at 2000 epochs

**Table 3:** Set of 39 wavelengths for SOM1-PLS models to analyze SMX and TMP

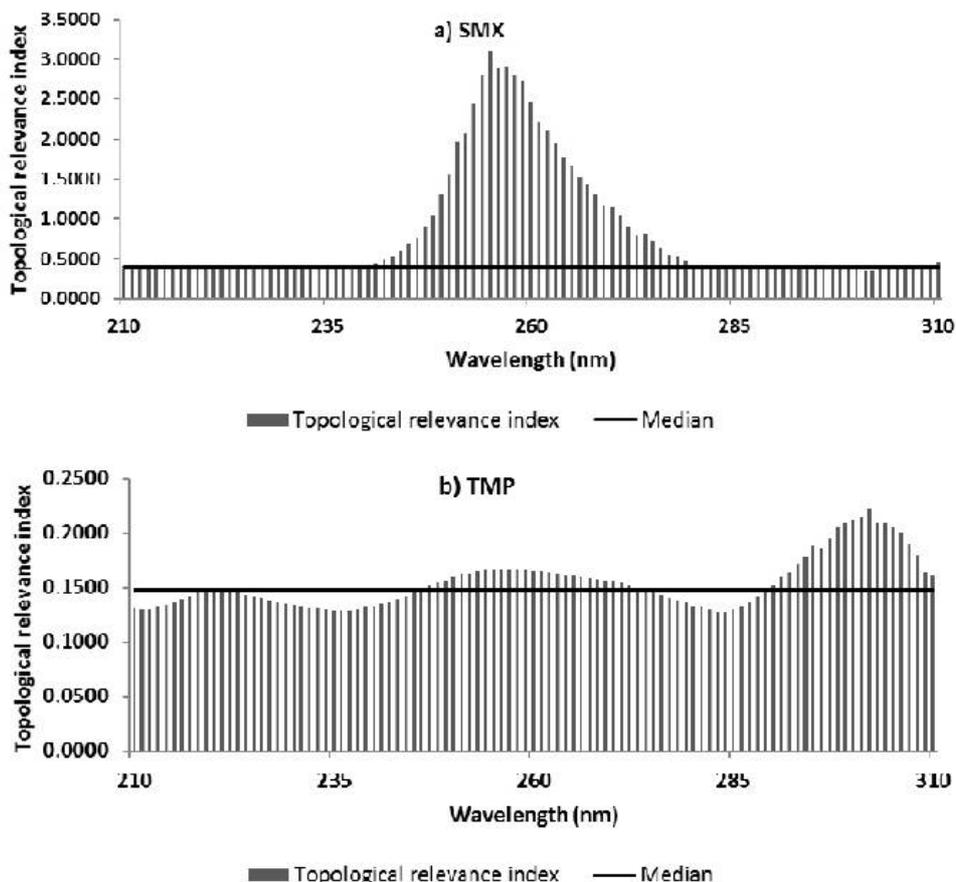
210	212	215	217	220	223	228	232	235	239
242	243	245	246	247	249	252	256	263	264
266	268	271	272	274	275	276	277	278	282
283	284	288	291	294	297	301	302	308	

To optimize and develop SOM2-PLS model, the grid sizes were varied with different size according to the ratio between eigenvalues of training data: 13x1, 26x2, 39x3, 52x4 for SMX and 12x1, 24x2, 36x3, 48x4 for TMP. Also, the number of training epochs were observed with different epochs: 200, 500, 1000 and 2000 for each grid size. A total of 16 networks for each drug were designed in this way. For each network, the quantization error and topographic error were calculated by SOM Toolbox and the results were illustrated in Figure 6. The sets of parameters with low quantization error and minimum topographic error were chosen to construct output maps for calculation of U-matrix of each component variable and topological relevance index. The wavelengths which had the

topological relevance index values (refer to inverted  $D(x_i, y_j)$ ) larger than the median cut-off line indicated strong relevancy and they were selected to create SOM2-PLS models. The values of topological relevance index and median cut-off line for SMX and TMP were shown in Figure 7. Sets of 51 wavelengths from 39x3 sized map at 500 epochs and set of 53 wavelengths from 36x3 sized map at 200 epochs were collected for analysis of SMX and TMP, respectively. These sets were presented in Table 4 and Table 5. Once the number of the wavelengths were selected, they were transmitted into PLS algorithm. SOM2-PLS models were developed. The summary specifications for SOM2-PLS models were listed in Table 2.



**Figure 6:** Quantization error and topographic error of SOM2 models. a) SMX; b) TMP



**Figure 7:** Topological relevance index and median cut-off line obtained from SOM2 models. a) SMX; b) TMP

**Table 4:** Set of 51 wavelengths for SOM2-PLS model to analyze SMX

240	241	242	243	244	245	246	247	248	249
250	251	252	253	254	255	256	257	258	259
260	261	262	263	264	265	266	267	268	269
270	271	272	273	274	275	276	277	278	279
280	289	290	291	292	293	294	307	308	309
310									

**Table 5:** Set of 53 wavelengths for SOM2-PLS model to analyze TMP

220	221	246	247	248	249	250	251	252	253
254	255	256	257	258	259	260	261	262	263
264	265	266	267	268	269	270	271	272	273
274	289	290	291	292	293	294	295	296	297
298	299	300	301	302	303	304	305	306	307
308	309	310							

**Table 6:** Results from applying SOM1-PLS and SOM2-PLS models to the validation samples

Statistical parameter	SOM1-PLS		SOM2-PLS	
	SMX	TMP	SMX	TMP
Mean recovery, %	99.33	102.58	99.55	102.27
RSD <sup>a</sup> , %	0.531	1.420	0.304	1.152
RMSEP	0.097	0.097	0.067	0.090

<sup>a</sup>Relative standard deviation

**Table 7:** Statistical results obtained from the proposed and official methods

Statistical parameter	SMX		Official method	TMP		
	SOM1-PLS method	SOM2-PLS method		SOM1-PLS method	SOM2-PLS method	Official method
Mean % LA ( <i>n</i> = 6)	98.04	99.78	100.91	107.02	101.12	101.93
RSD, %	0.490	0.466	0.560	0.893	0.852	0.380
<i>P</i> -value of variance	0.37	0.34		0.03	0.05	
<i>P</i> -value of mean	<0.01	0.11		<0.01	0.06	

By mean of recovery, 9 test samples were analyzed for accuracy consideration. The percentage recovery values of SMX and TMP, obtained from SOM1-PLS and SOM2-PLS models, were presented in Table 6.

### Application to pharmaceutical dosage form

The calibration models which were constructed from the wavelengths chosen earlier by SOM techniques were applied for determination of the active ingredients in co-trimoxazole tablets in accordance with preceding procedures. From SOM1-PLS models, the quantity of SMX and TMP in samples varied from 97.59 to 98.66 % LA (percent of the labeled amount) and 105.95 to 108.29 %LA, respectively, while for for SOM2-PLS models, the quantity of SMX and TMP varied from 99.20 to 100.60 % LA and 100.48 to 102.42 % LA, respectively. The acceptance limits of both drugs corresponding to official method in the United States Pharmacopeia 39th ed. (USP 39) are between 93.0 and 107.0 % LA [19].

The Student's *t*-test (*P*,0.01) data for the statistical significance of the difference among two means of SOM-PLS methods and the HPLC official method are provided in Table 7.

### DISCUSSION

The finding of this research study showed that SOM was an efficient option for wavelength selection in establishing new analytical approaches for quality assurance enhancement in pharmaceutical industry. From Table 6, the outcomes indicated that SOM1-PLS models and SOM2-PLS models offered good recoveries. Especially, for values of RMSEP, SOM2-PLS models provided 30 % enhancement of predictive ability for SMX and 6 % for TMP, when comparing to SOM1-PLS models. This suggested that SOM2 strategy has actually been confirmed better selection by cause of the supervising criterion for wavelength selection.

Furthermore, from Table 7, there were statistically significant differences between two

mean scores of SOM1-PLS method and official method ( $p < 0.01$ ), while there were no statistically significant differences ( $p > 0.01$ ) between two mean scores of SOM2-PLS method and official method. According to these outcomes, it was possible that there were two reasons why SOM2-PLS method provided better estimates than SOM1-PLS method. SOM2 algorithm chose wavelengths that had much information, involving relation between wavelength variables and concentration of the analytes, whereas SOM1 algorithm selected wavelengths by clustering of collinear wavelengths with no considerable in concentration variable. Moreover, SOM2-PLS model reduced noise because it provided lower optimum number of PLS factors when compared to SOM1-PLS model (Table 2). Hence, the number of factor loadings decreased.

### CONCLUSION

Successful comparison of SOM1 method against SOM2 method as a wavelength selection technique in PLS regression for analysis of pharmaceutical binary mixtures with severely overlapping spectra has been undertaken. Although, SOM2 method is more efficient than SOM1 method, some problems associated with SOM2 technique include the difficult computations of topological relevance index values and the distance values between each pair of input and output values estimated from U-matrices.

### DECLARATIONS

#### Acknowledgement

The authors are grateful to Sea Pharm Company Ltd, Thailand, for providing financial assistance and generous donation of drug samples used in this study.

#### Conflict of Interest

No conflict of interest associated with this work.

### Contribution of Authors

The authors declare that this work was done by the authors named in this article and all liabilities pertaining to claims relating to the content of this article will be borne by them.

### Open Access

This is an Open Access article that uses a funding model which does not charge readers or their institutions for access and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>) and the Budapest Open Access Initiative (<http://www.budapestopenaccessinitiative.org/read>), which permit unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

### REFERENCES

- Hemmateenejad B, Ghavami R, Miri R, Shamsipur M. Net analyte signal-based simultaneous determination of antazoline and naphazoline using wavelength region selection by experimental design-neural networks. *Talanta* 2006; 68(4): 1222-1229.
- De Luca M, Loele G, Spatari C, Ragno G. Optimization of wavelength range and data interval in chemometric analysis of complex pharmaceutical mixtures. *J Pharm Anal* 2016; 6(1): 64-69.
- Sratthaphut L, Phechkrajang C. Genetic algorithms-based approach for wavelength selection in spectrophotometric determination of paracetamol and chlorzoxazone in tablet preparation by partial least-squares. *Indian J Pharm Edu Res* 2012; 46(1): 59-65.
- Sasaki K, Kawata S, Minami S. Optimal wavelength selection for quantitative analysis. *Appl Spectrosc* 1986; 40: 185-190.
- Kalivas JH, Roberts N, Sutter JM. Global optimization by simulated annealing with wavelength selection for ultraviolet-visible spectrophotometry. *Anal Chem* 1989; 61: 2024-2030.
- Xu L, Schechter I. Wavelength selection for simultaneous spectroscopic analysis. *Experimental and theoretical study. Anal Chem* 1996; 68: 2392-2400.
- Kohonen T. *Self-Organization and Associative Memory*. Berlin Heidelberg: Springer Verlag; 1988.312p.
- Todeschini R, Galvagni D. Kohonen artificial neural networks as a tool for wavelength selection in multicomponent spectrofluorimetric PLS modelling: application to phenol, o-cresol, m-cresol and p-cresol mixtures. *Trend Anal Chem* 1999; 18: 93-98.
- Capitan-Vallvey LF, Navas N, del Olmo M, Consonni V, Todeschini R. Resolution of mixtures of three nonsteroidal anti-inflammatory drugs by fluorescence using partial least squares multivariate calibration with previous wavelength selection by Kohonen artificial neural networks. *Talanta* 2000; 52: 1069-1079.
- Corona F, Reinikainen S, Aaljoki K, Perkiö A, Liitiainen E, Barratti R, Simula O, Lendasse A. Wavelength selection using the measure of topological relevance on the self-organizing map. *J Chemometrics* 2008; 22: 610-620.
- Corona F, Liitiainen E, Lendasse A, Saaau L, Melis S, Baratti R. A SOM-based approach to estimating product properties from spectroscopic measurements. *Neurocomputing* 2009; 71: 71-79.
- Wordell CJ, Hauptman SP. Treatment of *Pneumocystis carinii* pneumonia in patients with AIDS. *Clin Pharm* 1988; 7(7): 514-527.
- Thera MA, Sehdev PS, Coulibaly D, Traore K, Garba MN, Cissoko Y, Kone A, Guindo A, Dicko A, Beavogui AH, et al. Impact of trimethoprim-sulfamethoxazole prophylaxis on falciparum malaria infection and disease. *J Infect Dis* 2005; 192: 1823-1829.
- Wise BM, Gallagher NB. *PLS\_Toolbox 2.0. Eigenvector Research, Manson, 1997.*
- Ballabio D, Consonni V, Todeschini R. The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for self-organizing maps and counterpropagation artificial neural networks. *ChemIntell Lab Syst* 2009; 98: 115-122.
- Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. SOM Toolbox for MATLAB 5. Technical Report A57: Helsinki University of Technology; 2000. [cited 2016 Jan 15]. Available from <http://www.cis.hut.fi/projects/somtoolbox/package/papers/techrep.pdf>.
- MATLAB 6.0. MathWorks, Massachusetts, 1996.
- Haaland DM, Thomas EV. *Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information.* *Anal Chem* 1988;60:1193-1202.
- United States Pharmacopeia and National Formulary (USP 39-NF 34), vol 3. Rockville, MD: United States Pharmacopeial Convention, 2016; pp 5956-5957.