



## CLASSIFICATION OF DRINKING WATER IN THAILAND BY NEAR INFRARED SPECTROSCOPY

\*Bandit THONGSROY<sup>1</sup> Phalanon ONSAWAI<sup>1</sup> and Panmanas SIRISOMBOON<sup>1</sup>

<sup>1</sup>Curriculum of agricultural engineering, Department of Mechanical Engineering, Faculty of Engineering,  
King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, THAILAND

Corresponding author: Phalanon ONSAWAI. E-mail: Phalanon.onsawai@hotmail.com

**Keywords:** Drinking Water; Near-infrared spectroscopy; Classification.

### ABSTRACT

Near-infrared spectroscopy (NIRS) appears as a prominent technique for non-destructive food quality assessment. This research work was focused on evaluating the use of NIRS in classification of drinking water for the benefit of consumers. The water samples were six groups which were 1) Natural Mineral Water 2) Bottled Mountain Mineral Water 3) Bottled Underground Mineral Water 4) Filtered Water (including bottled and household filtered water) 5) Tap Water 6) Distilled Water. The sample was subjected to FT-NIR spectrometer with the pathlength of 2 mm in transmittance mode. The optical data of full spectral range (12500-3600 cm<sup>-1</sup>) were analyzed by statistic methods of principle component analysis (PCA), soft independent modeling by class analogy (SIMCA), partial least square discriminant analysis (PLS-DA) and cluster analysis (CA). The PLS-DA model confirmed its best performance among the methods used with the average accuracy of 89.22%. It was concluded that the NIRS measurement technique is an attractive and efficient tool for classifying the types of drinking water in Thailand.

### INTRODUCTION

Drinking water is a finite and precious resource essential for sustaining life and health, and for ensuring the preservation of ecosystems [1]. Over 1 billion people lack access to safe drinking water, and an estimated 80% of child deaths from digestive-tract diseases such as diarrhoea (approximately 2 million per year) are caused by consumption of contaminated drinking water.

Chromatography is widely used as reference method to verify the authenticity of drinking water. However, chromatographic methods present many inconveniences like the time-consuming required to prepare a large variety of standard solutions that produces dangerous residues, and the use of instrumentation of high cost and maintenance and moreover, the chromatography is an invasive technique and is relatively slow, what implicates in a low sample throughput [2].

Ideally, an analytical method used to verify the quality and authenticity of drinking water should perform an analysis without sample pre-treatment. The combination of chemometric methods with NIR spectroscopy is a good way to eliminate the problems. In this work, a new strategy for classification of drinking water using NIR spectroscopy is proposed and the methods of chemometric classification, i.e. PCA, PLS-DA, SIMCA and cluster analysis are compared.

### MATERIAL AND METHOD

#### Water samples

A total of 17 drinking water samples from different lots and 6 groups (group A, Natural Mineral Water of 3 samples which were collected directly from the natural mineral hot spring (Ruksa Varin well) in the south of Thailand; group B, Bottled Mountain Mineral Water of 3 commercial brands, group C, Bottled Underground Mineral Water of 2 brands, group D, Filtered Water (including bottled and household filtered water) of 7 brands, group E, Tap Water, and group F, Distilled Water, were collected in Thailand.

#### Near infrared transmittance measurements

The transmittance spectra in the near infrared were obtained from untreated samples by using a Multi-Purpose Analyzer (MPA) FT-NIR spectrometer (Bruker, Bremen, Germany). For instrument control and data acquisition the OPUS program version 7.0.129 from Bruker was employed.

Each sample was transferred into a glass vial of 22 mm, as a measurement cell, covered with a transmittance plate made from stainless steel which provided an optical path length of 2 mm and scanned between 12500-3600cm<sup>-1</sup> with a nominal resolution of 16 cm<sup>-1</sup>, accumulating 32 scans per spectrum using a background of the gold. The scanning was done in a room temperature of 25±1 °C.

#### Data analysis

The principal component analysis (PCA) method was used to detect outlier spectra for each drinking water group. It was also used to assess the effectiveness of a non-supervised statistical method to separate the different drinking water groups according to their average spectral data as represented by the principal components (PCs).

Partial Least Square Discriminant Analysis (PLS-DA) and Soft Independent Modeling of Class Analogy (SIMCA) were then used to develop supervised classification models for drinking water using the calibration data set. PLS-DA is a classification method based on modeling the differences between several classes with PLS. If there are more than two classes to separate, the PLS model uses the corresponding response variables, which code for class membership as follows: 1 for members of one class, 0 for members of the other ones. SIMCA is based on making a PCA model for each class in the training set. Unknown samples are then compared to each class model and are assigned to a particular class according to their similarity to the calibration set samples.

The performances of the PLS-DA or SIMCA models were tested using the validation set based on the percent of

correctly classified into their respective drinking water classes (i.e. correct classification rate). The SIMCA model classification cut-off was determined based on the 95% confidence level.

According to Isaksson and Aastveit [3], cluster analysis (CA) assigned similar spectra, and therefore similar samples, to the same cluster based on a single data table [4]. Usually the dendrogram is illustrated where the Euclidean distances between the centroids of the groups are used to evaluate the similarity of the groups.

All the analyses were conducted using The Unscrambler 9.8 software (CAMO ASA, Oslo, Norway).

## RESULTS AND DISCUSSION

### NIR spectra

Fig. 1 presents the NIR spectra of the 17 drinking water in the range of 12500-3600  $\text{cm}^{-1}$ . As can be clearly seen, the absorption bands of water are at 10300, 8400, 6900, and 5160  $\text{cm}^{-1}$ .

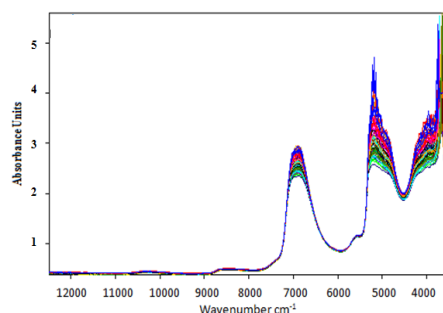


Fig. 1 Spectra of seventeen samples of drinking water analyzed in this work.

### Principal component analysis (PCA)

According to PCA, there was no outlier spectrum of tested samples. In Fig. 2, the scores graph PC1 (86%) $\times$ PC2 (9%) for the drinking water samples is studied, clearly demonstrating two clusters of drinking water samples including firstly, natural mineral water (NMW) and tapped water (TW) and secondly, bottled mountain mineral water (BMMW), filtered water (FW), bottled underground mineral water (BUMW) and distilled water (DW). The TW and NMW were in the same cluster because they were not filtered as much as FW. The DW was in the cluster of the latter since there is no mineral. However, the BUMW might be processed by severe filtering too. PC1 separates the two clusters.

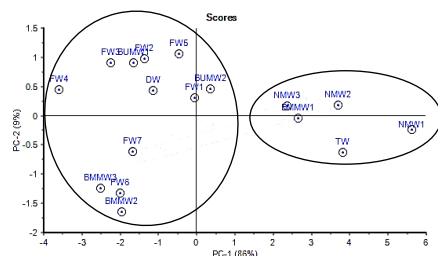


Fig. 2 PC2 $\times$ PC1 score plot for the overall set of 17 drinking water samples (NMW, BMMW, BUMW, FW, TW, and DW). The explained variance by each principal component is put in parentheses.

### SIMCA and PLS-DA

Table 1. Classification success rate (%) by SIMCA and PLS-DA for drinking water

	NMW	BMMW	BUMW	FW	DW	TW
SIMCA	41.18	5.88	11.76	41.18	94.12	82.35
PLS-DA	94.12	88.24	88.24	76.47	94.12	94.12

Table 1 shows that the PLS-DA provided the better performance than the SIMCA in classification of drinking water. The overall classification success rate of SIMCA and PLS-DA were 46.08% and 89.22%, respectively.

### Cluster Analysis (CA)

Fig.3 shows the dendrogram for the study of drinking water samples based on cluster analysis technique. It can be seen that two main classes can be identified. One contains NMW, BMMW (one brand) and TW, and the other contains BUMW, FW and DW. The NMW and TW are more similar than BUMW which more similar to FW. This confirmed the result of PCA.

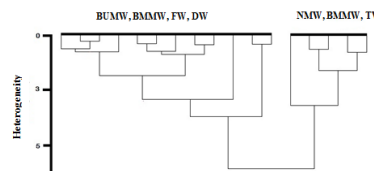


Fig. 3 Dendrogram of cluster analysis (CA) for drinking water samples of NMW, BMMW, BUMW, FW, TW and DW.

## CONCLUSION

This work was proposed a new strategy for classification of drinking water in Thailand using NIR spectrometry and chemometric methods (PCA, PLS-DA, SIMCA and cluster analysis). The PLS-DA model confirmed its best performance among the methods used with the average accuracy of 89.22%. Therefore, NIR spectroscopy combined with PLS-DA may provide an attractive and efficient tool for discrimination of drinking water in Thailand.

## REFERENCES

- Balbus, J.M., Lang, M.E., 2001. Is the water safe for my baby. *Pediatric Clinics of North America* 48, 1129–1152.
- Pontes, M.J.C., Santos, S.R.B., Araujo, M.C.U., Almeida, L.F., Lima, R.A.C., Gaião, E.N., Souto, U.T.C.P., 2006. Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Research International* 39, 182–189.
- Isaksson, T., Aastveit, A.H., 2002. Classification methods. Chalmers J.M., Griffiths (Eds.), *P.R Handbook of Vibrational Spectroscopy* 3 Wiley, Chichester. 2107–2122.
- Andreas Landman, A., 2005. Data reduction, and cluster and discriminant analysis of aluminosilicate infrared spectra—fly ash reacted at 860 °C with sodium carbonate as a model system. *Vibrational Spectroscopy* 37, 209–216