Classification of saline water for irrigated agriculture using near infrared spectroscopy coupled with pattern recognition techniques

Ravipat Lapcharoensuk^{1,*}, Jirawat Phuphanutada¹, and Patthranit Wongpromrat²

¹Post Harvest Innovation Research and Development Laboratory, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

²Department of Chemical Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

Abstract. This research aimed to create near infrared (NIR) spectroscopy models for the classification of saline water with a pattern recognition technique. A total of 112 water samples were collected from the Tha Chin river basin in Thailand. Water samples with salinity less than 0.2 g/l were identified as suitable for agriculture, while water samples with salinity higher than 0.2 g/l were found to be unsuitable. The NIR spectra of water samples were recorded using a Fourier transform (FT) NIR spectrometer in the wavenumber of 12,500–4,000 cm⁻¹. The salinity of each water sample was analysed by electrical conductivity meter. Identification models were established with 5 supervised pattern recognition techniques including k-nearest neighbour (k-NN), support vector machine (SVM), artificial neural network (ANN), soft independent modelling of class analogies (SIMCA), and partial least squares-discriminant analysis (PLS-DA). The performance of the NIR model was carried out with a split-test method. About 80% of spectra (90 spectra) were randomly selected to develop the classification models. After model development, the NIR spectroscopy models were used to classify the categories of the remaining samples (22 samples). The ANN model showed the highest performance for classifying saline water with precision, recall, F-measure and accuracy of 84.6%, 100.0%, 91.7% and 90.9%, respectively. Other techniques presented satisfactory classification results with accuracy greater than 68.2%. This point indicated that NIR spectroscopy coupled with the pattern recognition technique could be applied to classify saline water for agricultural use according to salinity level in natural resources.

1 Introduction

Tha Chin river basin is located in the central region of Thailand and is separated from the Chao Phraya River. It is an important river for agricultural plantations, especially the lower part of the river, where many kinds of fruits such as mangoes and coconuts as well as valuable

* Corresponding author: ravipat.la@kmitl.ac.th

orchids are grown. From 2014 to 2018, the income gained by exported orchids was around 1.9–2.3 billion baht (about 54–66 million USD) per year [1]. The key to this success was the quality of the soil and water making high-quality, valuable products. However, the location is also near the Gulf of Thailand. The seawater level in the Gulf of Thailand sometimes increases due to the ocean tides, leading to the inverse flow of seawater into the lower Tha Chin, resulting in the increased salinity of the water.

High salinity in the water is a big problem for agricultural plantations because it affects the growth of plants, i.e. decreased photosynthesis, changes in the metabolic process which inhibit mineral absorption and dehydration of the plants due to the high adsorption of sodium ions (Na^+) and chloride ions (Cl^-). Additionally, the salinity affects the breath of the plants resulting in inhibited growth. Staff from the US Salinity Laboratory (USSL1954) categorise water quality based on the salinity of water and specified that high-quality water for planting agricultural products contains salinity less than 0.2 g/L [2]. Therefore, it is necessary to monitor the salinity of water periodically in order to ensure high-quality products. When the salinity of the water exceeds the control level, treated water from Chao Phraya Dam is drained to force the seawater to flow out to the Gulf of Thailand. The Royal Irrigation Department is responsible for monitoring the water quality and draining the water if and when necessary. The department uses a water-quality meter (multi-parameter) to roughly measure the salinity of water in real-time, so a more precise method is still desirable.

There are several methods for measuring the salinity in water. The general principle is to determine the properties of water, i.e. the electrical conductivity by conductometer, the specific gravity by hydrometer, the angles of incidence and reflection of light by refractometer and the total dissolved solids (TDS). These methods can be done easily, but their results can be disturbed by other chemicals, especially other ionic salts, leading to erroneous experiment results. Hence, a more modern and feasible method is needed. Near Infrared Spectroscopy (NIR), especially the portable-NIR, is known to be effective for detecting the H-bond consisting in the covalent structures. Although NaCl is an inorganic chemical solution, it can be detected by NIR since Cl^- is surrounded by the molecule of water. This force induces the change in dipole moment in the bonds and NIR absorbance is then presented [3]. Many researchers have reported on NaCl in many mediums, such as in canned sardine [4], in hot-smoked salmon [5], and in the soil [6], which could be detected using NIR spectroscopy. However, the application of NIR spectroscopy for detecting salinity in natural water is very difficult due to the variety and complexity of natural water composition. The chemical composition of natural water is usually compound with main ions, dissolved gases, biogeneous substances, organic substances, microelements and pollutants [7]. These chemical compositions may affect to the NIR absorbance spectra. Therefore, a study on the classification of saline water using NIR spectroscopy is an interesting research. The outcome of this research obtains different knowledge from previous researches which chemical composition of the sample is not complexity.

In this work, Fourier Transform Near Infrared Spectroscopy (FT-NIR) models were developed with a pattern recognition technique for classifying excellent water quality containing salinity less than 0.2 g/L from higher salinity values. The confusion matrix was calculated to evaluate the feasibility of the classification model. After achieving the most feasible model, the NIR spectroscopy models could be applied to more efficiently manage water, leading to reduced risks for agricultural product damage.

2 Methodology

2.1 Water sample collection

One hundred and twelve water samples were collected from the Tha Chin river basin in Thailand. The samples were randomly gathered from the estuary along 82 km of the river, which is an area shared by the provinces of Nakhon Pathom and Samut Sakhon. Sample collection was done every month in 2018. Each sample was collected using a stainless-steel bucket at a depth level of 150 cm from the water surface. The water samples were placed in BOD bottles at a volume of 300 ml, after which they were delivered to the laboratory for scanning NIR spectra.

2.2 NIR spectra scanning

Before NIR spectra scanning, each sample bottle was shaken for 10 minutes. Each water sample was taken from the BOD bottle with a plastic dropper. After that, the sample was dropped into a quartz cup and a ceramic lid was put over the quartz cup for reflecting NIR radiation during scanning. The NIR spectra of water samples were measured by a Fourier transform (FT) NIR spectrometer (MPA, Bruker Ltd., Germany) in a wavenumber range of $12,500\text{--}4,000\text{ cm}^{-1}$ ($800\text{--}2,500\text{ nm}$) with a resolution of 8 cm^{-1} . The spectrum of each sample was measured in diffuse reflectance mode and reported on $\log 1/R$. The R value is a reflection of NIR radiation from the water sample. Each spectrum was a product of 32 internal scanning. The NIR spectra scanning procedure was carried out at room temperature of $25 \pm 2^\circ\text{C}$. Fig. 1 shows the NIR spectra scanning of the water samples.

Fig. 1. NIR spectra scanning of water samples.

2.3 Salinity analysis

After NIR spectra scanning, the water samples were delivered to the laboratory for determination of salinity. The samples were kept at room temperature ($25 \pm 2^\circ\text{C}$) for 30 minutes. Salinity value was analysed using an electrical conductivity meter (YSI Ltd., USA). Each sample was measured 3 times and then averaged. In this research, water samples with salinity less than 0.2 g/l were identified as suitable for agriculture (SWA), while the samples with salinity higher than 0.2 g/l were unsuitable for agriculture (UWA).

2.4 Qualitative classification modelling

Qualitative classification models were analysed with five supervised pattern-recognition techniques including k-nearest neighbour (k-NN), support vector machine (SVM), artificial neural network (ANN), soft independent modelling of class analogies (SIMCA) and partial least squares-discriminant analysis (PLS-DA). The SIMCA and PLS-DA models were performed with Unscrambler v10.1 software package (CAMO AS, Trondheim, Norway). Remainder methods (i.e. k-NN, SVM and ANN) were analysed by the open-source framework of RapidMiner Studio (version 9.1, Education Edition). The performance of the NIR model was done with a split-test method. About 80% of spectra (90 spectra) were randomly selected to develop the classification models. After model development, the NIR spectroscopy models were used to classify the categories of the remaining samples (22 samples). The potential of the classification models was evaluated in terms of precision, recall, f-measure and accuracy, which are defined as [8]:

$$\text{precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{f-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

where TP, FN, TN and FP are the number of true positives, false negatives, true negatives and false positives, respectively. TP explains the classification result where the model correctly identifies for SWA samples. TF is a correct identification result for UWA samples. On the other hand, FP is an incorrect outcome where the model predicts UWA samples as SWA. Finally, FN is an incorrect identification result where the model predicts SWA samples as UWA.

3 Results and discussion

3.1 Salinity of water samples

Table 1 shows the statistical results of salinity for the water samples. SWA and UWA comprised 56 samples. The salinity of the SWA samples ranged from 0.10 to 0.20 g/l, while the UWA samples showed salinity between 0.30 and 8.50 g/l. The means of SWA and UWA were 2.67 ± 0.24 and 4.20 ± 1.06 g/l, respectively. The frequency distribution for salinity of water samples is shown in Fig. 2. About sixty-six percentage of the samples were between 0.01–1.78 g/l. The salinity of the remaining samples ranged between 1.78–8.50 g/l.

Table 1. Statistical data for salinity of water samples.

Water category	Number of samples	Salinity (g/l)				
		Min	Max	Mean	Range	SD
SWA	56	0.10	0.20	0.19	0.10	0.03
UWA	56	0.30	8.50	3.58	8.20	2.59

Fig. 2. Frequency distribution for salinity of water samples.

3.2 NIR spectra

The raw and second derivative spectra of SWA and UWA samples are shown in Fig. 3 and 4, respectively. The evident absorbance peaks were obtained between a wavenumber of 6900 and 5155 cm⁻¹ (1450 and 1940 nm) for raw and second derivative spectra plots. Both peaks are the O-H stretching first overtone and combination of H₂O stretching [9]. The absorbance value at both water bands of UWA was higher than SWA. However, NIR radiation cannot specify the absorbance band of salt. The detection of salinity in water is possible due to the effect of sodium chloride on the shape and position of the water band [5, 10–13].

3.3 Classification model

Table 2 presents the classification results for water by pattern recognition techniques. All models showed true positive between 8-10 samples, while the range of true negative was 5–10 samples. False positives for all models were 1–6 samples, while false negatives were found in 0–3 samples. No false negatives were found by the ANN model. The classification results obtained from all models presented precision between 62.5 and 100%. All supervised models showed recall values of 45.5–100%. F-values for all classification models were between 74.1 and 91.7%. Accuracy of SIMCA, PLS-DA, ANN, k-NN and SVM was 68.2, 86.4, 90.9, 81.8 and 81.8%, respectively. The ANN model showed the highest performance of classification with accuracy of 90.9%. This finding shows that NIR spectra combined with

supervised pattern recognition models (i.e. SIMCA, PLS-DA, SVM, k-NN and ANN) is an alternative technique to detect salinity in water for agricultural irrigation.

Regression coefficient plot of the PLS-DA model is shown in Fig. 5. Regression coefficients are the numerical coefficient that specifies the relationship between variation in the independent variable (NIR spectra) and variation in the dependent variable (reference data). The peaks at 6900 and 5155 cm^{-1} (1450 and 1940 nm) were O-H stretching first overtone and combination of H_2O stretching. This point confirmed the effect of sodium chloride on NIR absorption in the water band. Therefore, the PLS-DA model could classify the category of saline water according to salinity level.

Fig. 3. Raw spectra of water samples.

Fig. 4. Second derivative spectra of water samples.

Table 2. Classification results of water by pattern recognition technique.

Model	True positive	True negative	False positive	False negative	Precision	Recall	F-Measure	Accuracy
SIMCA	10	5	6	1	62.5	90.9	74.1	68.2
PLS-DA	10	9	2	1	83.3	90.9	87.0	86.4
ANN	11	9	2	0	84.6	100.0	91.7	90.9
k-NN	8	10	1	3	88.9	72.7	80.0	81.8
SVM	9	9	2	2	81.8	81.8	81.8	81.8

Fig. 5. Regression coefficient plot of the PLS-DA model.

4 Conclusion

Near infrared spectroscopy coupled with pattern recognition techniques was used to classify the salinity of water for agricultural irrigation. In this research, the PLS-DA, ANN, k-NN and SVM models achieved satisfactory classification performance with accuracy greater than 81.8%. The ANN model showed the highest performance for classifying saline water with precision, recall, F-measure and accuracy of 84.6%, 100.0%, 91.7% and 90.9%, respectively. This point indicates that NIR spectroscopy coupled with a pattern recognition technique could be used effectively to classify the salinity in water according natural resource salinity levels for agriculture.

The authors gratefully acknowledge the instrument and software support provided by the NIRS Research Center for Agricultural Product and Food, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.

References

1. Thai Customs, *Thailand Trading Report* (Bangkok: Thai Customs Department, 2019)
2. US Salinity Laboratory Staff (USSL), *Diagnosis and Improvement of Saline and Alkali Soils (Agriculture handbook 60)* (Washington, DC: USDA, 1954)

3. R. Matters, M. A. Sugui, F. Chen, X. Shi, J. Liu, P. A. Gilbert, *Bioprocess Inter.* **4**, 44–50 (2009)
4. S. Srikornkarn, P. Sirisomboon, *Agric. Agric. Sci. Procedia* **2**, 381–385 (2014)
5. M. Lin, A. G. Cavinato, Y. Huang, B. A. Rasco, *Food Res. Int.* **36**, 761–766 (2003)
6. J. Peng, W. Ji, Z. Ma, S. Li, S. Chen, L. Zhou, Z. Shi, *Biosyst. Eng.* **152**, 94–103 (2016)
7. A. Nikanorov, L. Brazhnikova, *Types Prop. Water* **2**, 42–80 (2009)
8. R. Chawuthai, R. Sakdanuphab, *Proc. Int. Conf. on 4th Engineering, Applied Sciences, and Technology*, 1–4 (2018)
9. B. G. Osborne, T. Fearn, *Theory of near infrared spectrophotometry* (In: *Near Infrared Spectroscopy in Food Analysis*, London, Longman Science & Technical, 1993)
10. Y. Huang, A. G. Cavinato, D. M. Mayes, G. E. Bledsoe, B. A. Rasco, *J. Food Sci.* **67**, 7, 2543–2547 (2002)
11. Y. Huang, T. M. Rogers, M. A. Wenz, A. G. Cavinato, D. M. Mayes, G. E. Bledsoe, B. A. Rasco, *J. Agric. Food Chem.* **49**, 9, 4161–4167 (2001)
12. J. Lin, C. W. Brown, *Appl. Spectrosc.* **46**, 1809–1815 (1992)
13. T. H. Begley, E. Lanza, K. H. Norris, W. R. Hruschka, *J. Agric. Food Chem.* **32**, 984–987 (1984)