

Using Spectroscope to Fast Screen Hinoki Hydrolat Concentration

Chang Lung Yen^{1*}, Meng-Shiu Lee², Jian-Hung Chen¹, Hung-Yu Chien¹ and Jen-Son Cheng¹

¹College of Management, National Chi Nan University, Nantou County 545, Taiwan

²Assistant Professor, University Rd, China

ISSN: 2576-8816



***Corresponding author:** Chang Lung Yen, College of Management, National Chi Nan University, Nantou County 545, Taiwan

Submission: 📅 November 22, 2024

Published: 📅 December 17, 2024

Volume 11 - Issue 3

How to cite this article: Chang Lung Yen*, Meng-Shiu Lee, Jian-Hung Chen, Hung-Yu Chien and Jen-Son Cheng. Using Spectroscope to Fast Screen Hinoki Hydrolat Concentration. Res Med Eng Sci. 11(3). RMES.000761. 2024. DOI: [10.31031/RMES.2024.11.000761](https://doi.org/10.31031/RMES.2024.11.000761)

Copyright@ Chang Lung Yen. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract

Hinoki hydrosol is the remaining by-product of the distillation process of hinoki essential oil. In traditional applications, although hinoki hydrosol has many health benefits, due to the different concentrations and uncontrollable changes, most of the hinoki hydrosol is discarded and not used, in order to achieve the purpose of circular economy utilization. In order to distinguish the concentrations of hinoki hydrosol and pure water samples, this study designed a set of experimental methods to scan the spectra of various hinoki hydrosol and pure water samples, pre-process the spectra, and then apply two types of machine learning Calculation methods (K-means and PLSR).

Hinoki hydrosol is the remaining by-product of the distillation process of hinoki essential oil. In traditional applications, although hinoki hydrosol has many health benefits, due to the different concentrations and uncontrollable changes, most of the hinoki hydrosol is discarded and not used, in order to achieve the purpose of circular economy utilization. In order to distinguish the concentrations of hinoki hydrosol and pure water samples, this study designed a set of experimental methods to scan the spectra of various hinoki hydrosol and pure water samples, pre-process the spectra, and then apply two types of machine learning Calculation methods (K-means and PLSR).

In the future, follow-up research will plan to combine GC-MS experimental instruments and obtain more accurate concentration data to train the PLSR model studied, so that the improved PLSR model can further improve the accuracy of quantitative prediction of hinoki hydrosol. And conduct research on the subsequent application of hinoki hydrosol products.

Keywords: Hinoki hydrolat; K-means; GC-MS; Spectra; K-means; PLSR

Introduction

Red cypress and chamaecyparis: There are only a few areas in the world that grow cypress, and Taiwan is located at the southernmost geography of the growth of cypress as well as the only area with subtropical climate that has cypress. Taiwan cypress consists of two kinds of cypress: Meniki (*Chamaecyparis formosensis*) and Hinoki (*Chamaecyparis obtuse*), and they are the coniferous species commonly seen at intermediate high altitude in Taiwan. They are of excellent quality and very high economic value, and they are species used intensively for forestation in mountains of high attitude [1]. In the past of more than half a century, numerous researchers put lots of efforts to study Meniki and Hinoki, exploring their habitat, physiological ecology, chemical elements, physical characteristics, and bioactivity [2].

Qualitative research was carried out for the distillation treatment on the outer bark of Hinoki to explore the possibility of using the residual forest biomass to produce valuable chemicals [3]. Most Hinoki leaves contain essential oil, and it includes the element of terpenoid

glycoside. Raw terpenoid glycoside is extracted using conventional distillation extraction on Hinoki leaves [4]. The ethanolic extracts on the bark of Hinoki is able to greatly inhibit fusarium and pseudomonas, and the inhibition is mainly caused by the volatile oil and non-volatile matter in the bark while neutral substances and acidic substances are both with high activity towards pathogen [5]. The report [6] highlighted that the cedar leaf oil of Hinoki can be used to stimulate olfactory sensation to significantly reduce the oxygenation concentration of hemoglobin in the right prefrontal cortex as well as increase parasympathetic activities. Academic studies revealed that touching Hinoki with palms is able to calm the activity on prefrontal cortex as well as increase the parasympathetic activities and further cause physical relaxation [7]. According to the academic research, several chemical elements found on Hinoki contain terpenoids; some of them are recognized to be good for antitumor, anti-malaria, and antimicrobial activity [8]. In terms of the application of Hinoki by dentists, the result shows Hinoki is a strong antibacterial agent and able to carry out bactericidal action on *Porphyromonas gingivalis*. Hinoki antibacterial agent will not cause huge effect on the characteristics of the main outer membrane protein but will interfere with the antioxidant activities on bacteria [9].

Conventional extraction process of Taiwan cypress essential oil will generate the by-product, hydrolat. Hydrolat is the condensate dissolved solution obtained from the distillation of logs, plants, or flowers. During the process of distilling logs, plants, or flowers, it creates oil-water separation due to different density; the essential oil floats on the top and hydrolat water on the bottom. In addition to few amounts of essential oil, hydrolat also contains all the water-soluble substances from logs, plants, or flowers. These hydrolats are currently commonly used to perfume, generic drug, food processing, aromatherapy, and traditional therapy [10]. The Hinoki hydrolat generated from traditional distillation is difficult to control its quality, and there are very few hydrolat studies; therefore, only essential oil will be kept but the Hinoki hydrolat is usually thrown away. If relevant pharmacological research and development application can be carried out as well as the quality of hydrolat after distillation can be controlled, there is a potential opportunity to develop high-value products. It fits the benefits of environments and economics; that is a potential resource recycling business.

Conventionally, there are some testing methods to test chemical elements in a product and can be used to test the elements of hydrolat; for example, gas chromatography-mass spectrophotometer (GC-MS). However, these methods demand high cost and take a long time. The testing fee on hydrolat using conventional methods is too high; if each batch of cypress hydrolat has to be tested using the conventional method, industries will not be able to produce products that are affordable by customers. Therefore, if a technology of rapid testing on distilled cypress hydrolat quality with low cost and high efficiency can be developed, it will be able to greatly reduce the time and cost on the quality of in-coming material of Hinoki hydrolat as well as significantly enhance the development of circular economy on Hinoki hydrolat.

In the recent years, the application of collecting spectrum of an item with spectroscopy and applying various data processing and machine learning has been successfully used in the areas of agriculture, food, drugs, and soil. Some of the applications can be implemented with qualitative analysis while some have already successfully developed quantitative analysis. Compared to the conventional technologies, applying machine learning on substances' spectra owns the several advantages: non-destruction, immediacy, lower cost, and lower labors.

This study focuses on developing Hinoki hydrolat concentration fast-screening technology, using machine learning and spectra.

Literature Review

There are several studies show different parts and substances have great benefits to health and to the environments [5-7, 11-16]. Hinoki essential oils are more popular than Hinoki hydrolat in the consumer market. One reason is not enough studies on Hinoki hydrolat, and the other is the concentration variation of Hinoki hydrolat.

Gas chromatography-mass spectrometry (GC-MS) consists of two main parts: gas chromatography and mass spectrometry. GC-MS can be applied in many fields like environment, biomedicine, and bioinformatics [17]. GC-MS spectrographic method [18], C-NMR spectrum [19] and Headspace solid-phase microextraction (static-HS) [20] can be applied in analyzing the chemical elements of Hinoki substances, but it demands higher cost.

On the other hand, spectra of samples can be easily collected through infra-red spectrum, multi-spectra, and hyperspectral instruments. The process does not compromise the samples, and it is easy to train the operators. Each ingredient inside a sample would generate a different reflection on different spectrum, and these data (or curves) can be used to identify the ingredients.

Near infrared (NIR) spectrum has been used to develop qualitative and quantitative methods with non-invasive modes to monitor wet granulation process [21], in tea [22]. According to the result of analysis in the six types of hydrolats, the experimental outcome revealed better absorption value appears at 290~400nm in Taiwan sandalwood hydrolat and Taiwan cypress hydrolat [23]. Chien et al. [24] have already found out the fast testing on the local representative longan honey and lychee honey through honey spectrum matching with machine learning [25].

Preliminaries on some background technologies and tools

The instrument we use is the FieldSpec4 spectroradiometer from American ASD Company (now Malvern Panalytical Company). FieldSpec4 spectrum ranges from 350nm to 2500nm. It uses graded index InGaAs photodiode SWIR detectors. It provides 3nm spectral resolution in the VNIR (350nm-1000nm) range and 10nm in the SWIR (1001nm-2500nm) range. Field Spec. The light source we use is a 75watt quartz-tungsten-halogen light.

The Principal Component Analysis (PCA) [26] is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Because there are more than hundreds of spectrums in each sample, we apply PCA to reduce the components before the spectra are fed into the machine learning models.

K-means clustering [27] is a method that aims to partition n sample observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. Here, because most of the Hinoki hydrolat we got does not have been tested under the conventional chemical testing like GC-MS tests. We try to apply k-means to cluster these samples to see whether we can correctly cluster and classify the samples, according to Hinoki hydrolat concentration.

Partial Least Squares Regression (PLSR) [28] finds a linear regression model by projecting the predicted variables (Y) and the observable variables (X) to a new space in such a way that it finds the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. We speculate that PLSR is suited in our cases because there are many spectra data in each sample and there might be multicollinearity among the X values.

The design of Hinoki hydrolat concentration experiments and the samples

To confirm wood genes, Taiwanese cypress samples were sent to the Forestry Bureau laboratory. After determining the source of the cypress wood, the Taiwanese cypress wood was subjected to a water distillation method. This method is to mix and heat the

plant body with water. The essential oil of the plant body can be distilled out along with the water vapor. After the essential oil and hydrosol vapor are cooled, the specific gravity of the essential oil is relatively high. The hydrosol is small and remains in the upper layer. The essential oil and hydrosol can be obtained by separating the two. This study is to take out the cypress hydrosol for analysis and research.

We totally got 9 batches of Hinoki hydrolat samples, where each batch represents all the samples from the same day, and each batch consists of several Hinoki hydrolat samples from the different distillation phases at a specific date; the number of sample for each day ranges from 5 to 12. We pick up one sample from each day batch and label them sequentially A to J. We do not have necessary pre-knowledge (concentration) of these concentrations of these Hinoki hydrolat samples and we have no way to apply supervised machine learning from the raw data.

Therefore, we designed the following two-phase experiment to tackle the challenges (depicted in Figure 1). The experiment design consists of two-phases. In Phase 1, we select one sample from each batch, and label them A to J. For each sample from A to J, we dilute it with different concentration (0%, 10%, 30%, 50%, 70%, 90%, 100%) with pure water, and label them accordingly; A1 represents 100%A, A0.1 represents 10%A with 90% pure water, A0.3 represents 30%A with 70% pure water, etc. We then use ASD field spec 4 to scan the samples and get the spectra. Following that, we identify potential bands, and then various pre-processing mechanisms are applied before the data are fed into machine learning models (here, we test k-means and PLS). Finally, we select a suitable model for phase 2. In Phase 2, we apply the selected model on all the rest samples and evaluate its prediction.

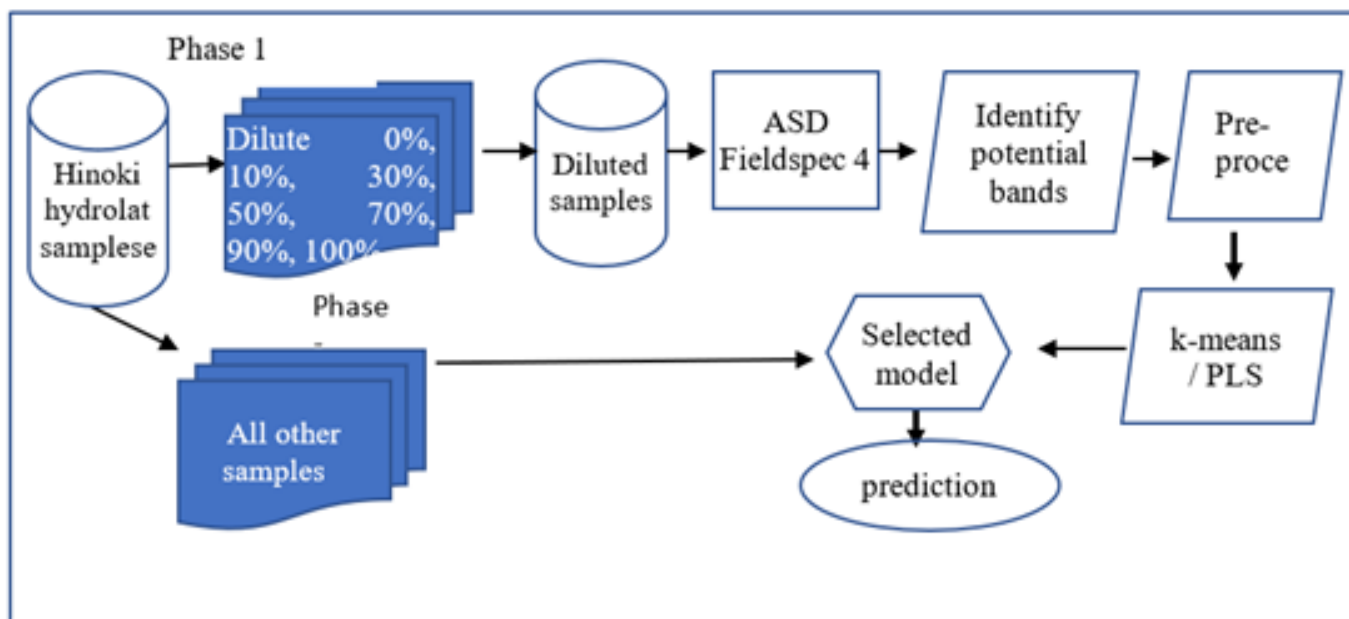


Figure 1: Hinoki hydrolat concentration experiments design.

Result and Discussion

Identifying potential bands for the experiments

We visually identify the potential bands for the classification. We also use Orange tool [29] to identify the potential bands for classification, and the results agree with our visual inspection.

Figure 2 shows the Hinoki hydrolat samples' spectra and

the water's spectra in different bands. Figure 2a shows the spectra in the 350~2500nm. Here, we can see that we cannot differentiate the Hinoki hydrolat samples from the water in some bands. The identified three potential bands are 350nm~400nm, 750nm~850nm, and 1200nm~1300nm respectively. Figure 3 show the diluted A-batch Hinoki hydrolat samples' spectra and the water's spectra in different bands.

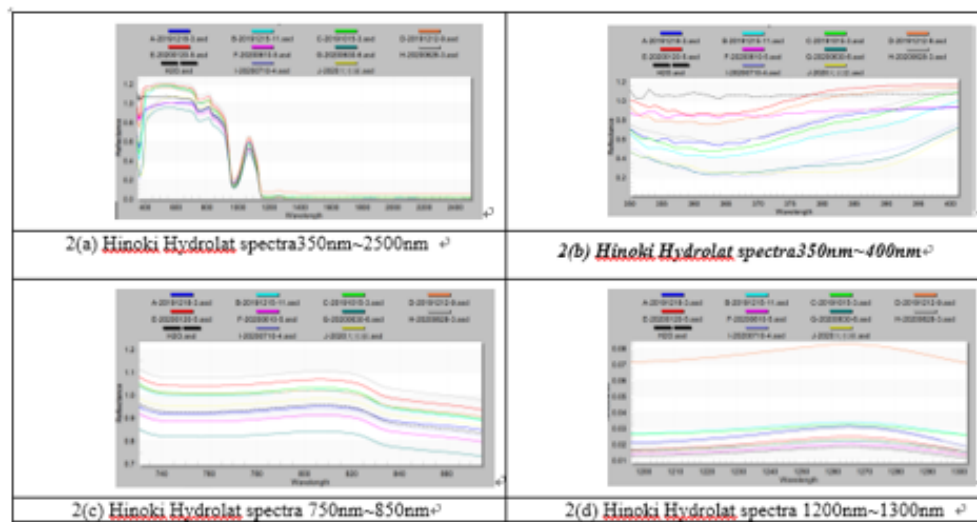


Figure 2: Hinoki hydrolat spectra and water spectra in different bands.

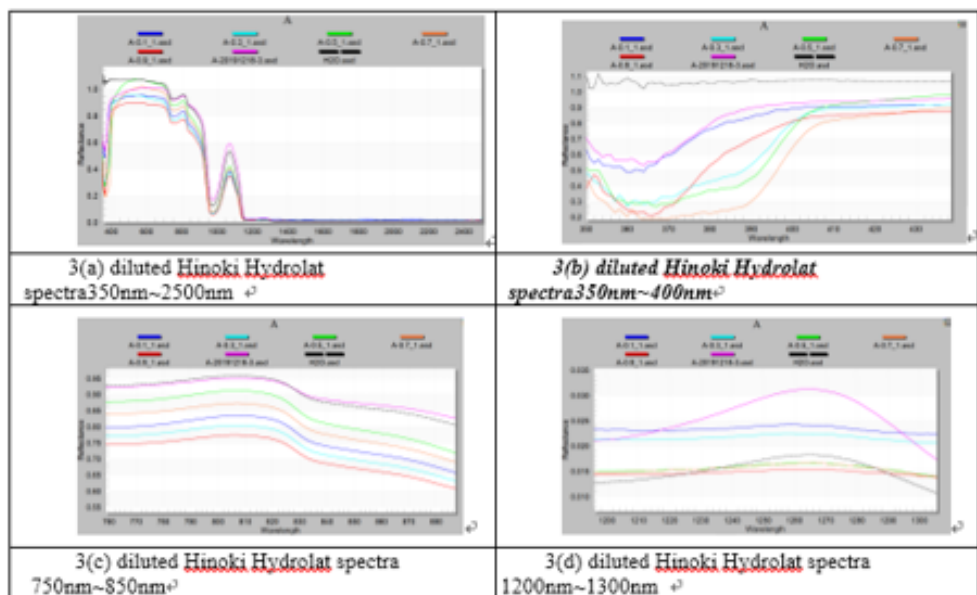


Figure 3: Diluted A-batch Hinoki hydrolat spectra and water spectra in different bands.

Even though we can differentiate different Hinoki hydrolat spectra and the water, we cannot visually identify any linear relation of the relative positions of the diluted samples, according to the diluted concentration.

The results of K-means clustering

Based on the identified three bands, we apply k-means clustering on the spectra data and the PCA-processed spectra data to see whether k-means can effectively differentiate different Hinoki hydrolat spectra and the water spectra. The clustering patterns from the k-means clustering are shown in Figures 4&5.

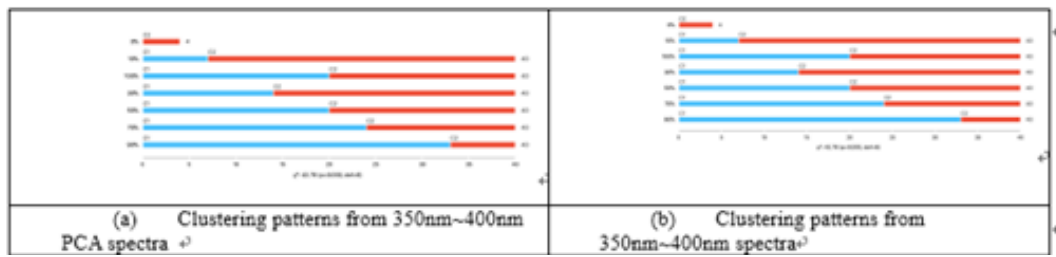


Figure 4: The k-means clustering based on 350~400nm diluted spectra.

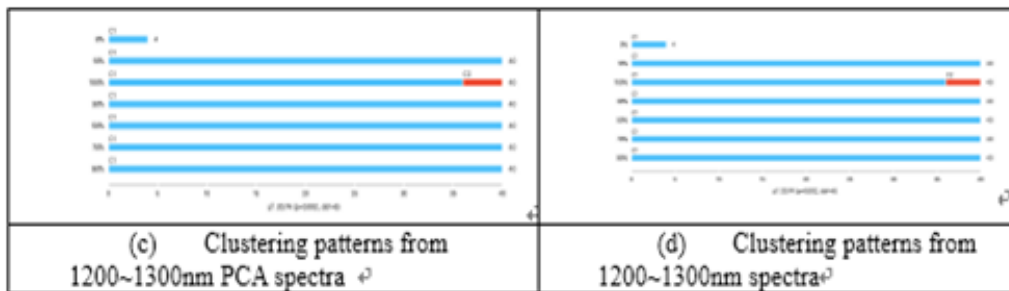


Figure 5: The k-means clustering based on 1200~1300nm diluted spectra.

From these figures, we can see that the k-means clustering cannot effectively classify the diluted spectra based on the concentration. For example, in Figure 4a, pure water (0%) is labeled as C2 class; however, there are more samples from the 100% spectra being classed C2 class than the samples from the 90% spectra. That

is, the k-means clustering apparently cannot classify the samples, according to their concentration. The k-means clustering, based on other bands, show similar phenomenon. In a short summary, the k-means clustering cannot effectively classify the diluted spectra based on the concentration.

The results of PLSR experiments on the diluted samples

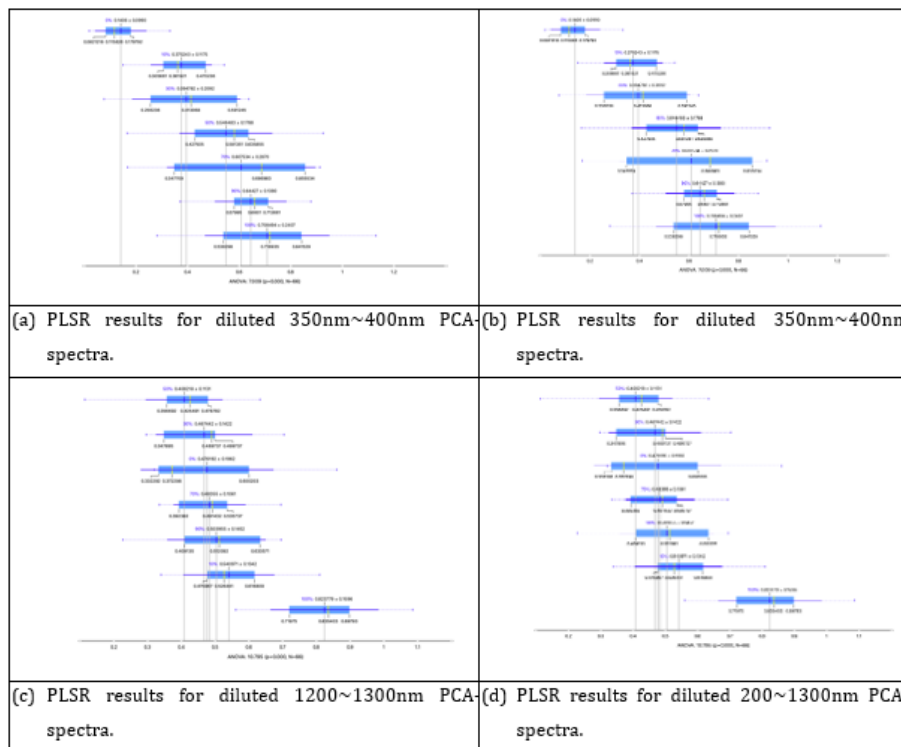


Figure 6: PLSR result for diluted samples in different bands.

Figure 5 shows the PLSR results for diluted Hinoki hydrolat samples at different bands, where (a) and (b) show that for the 350~400nm band, and (c) and (d) show that for the 1200~1300nm band. The results for the 750~850 band are similar to that for the 1200~1300nm band. In Figure 5a&b, we can see that PLSR can correctly differentiate the diluted concentration from 0% (top-left) to 100% (bottom-right); that is, the PLSR applied on the 350~400nm band linearly predicts the diluted samples according to the diluted concentration. On the contrary, in Figure 6c&d, the results do not linearly predict the diluted samples, according to the concentration; that is, the PLSR for the 1200~1300nm (also the 750~850nm) band cannot linearly predict the diluted samples, according to the concentration. TPLSR results for the batch samples

using the diluted samples PLSR model at 350~400nm.

As the PLSR results for the diluted samples at 350~400nm band can linearly predict the concentration at the 350~400nm band, we now apply the model to evaluate the concentration of the batch samples (A~J) using the 350~400 nm band. The results are shown in Figures 7a&b. The results show that the concentration of the batch sample ranging from low to high is water (0%), D, F, E, H, B, C, G, J, A and I. Among the batch samples, the batch-J samples are the mix of all other batch samples, and the batch-J samples' concentration does sit between other batches. For the application of this model and technology, we will plan to use this model to fast filter out some batches of which the PLSR predictions are close to the water.

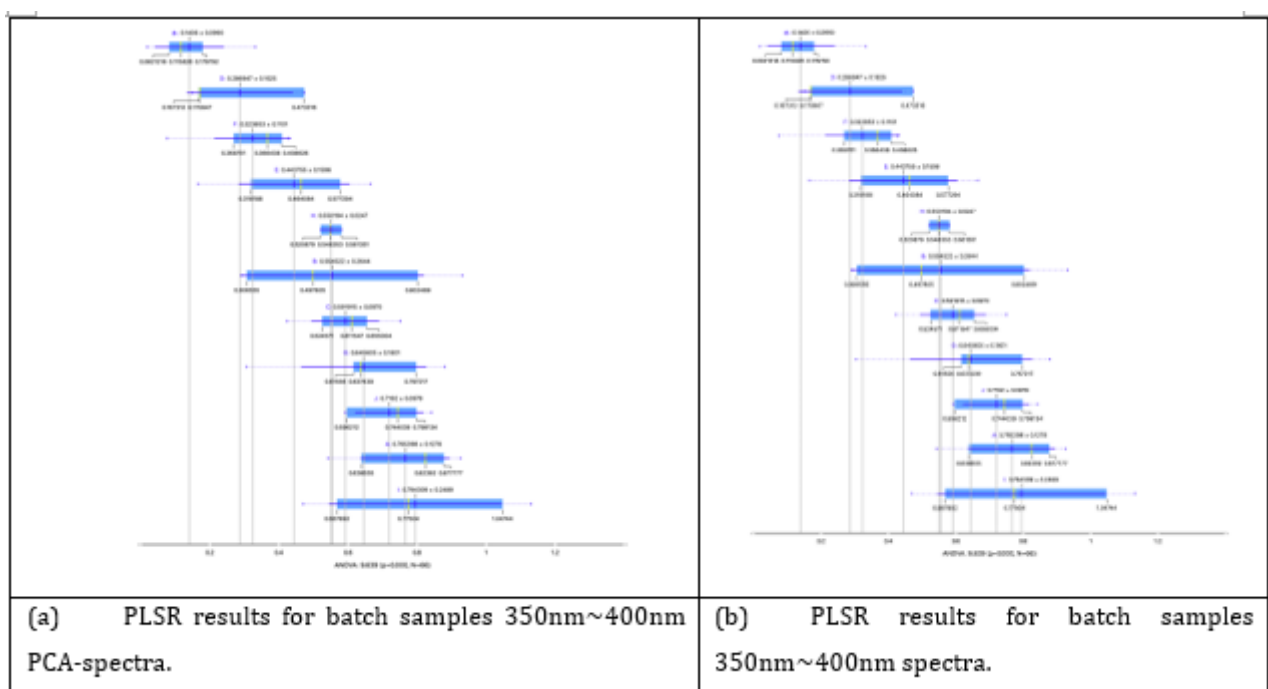


Figure 7: PLSR results for the batch samples, using the model from 350~400nm diluted-sample PLSR model.

Conclusion

In this paper, to differentiate the concentration of Hinoki hydrolat samples, we have designed an experiment that scans the samples' spectra, pre-processes the spectra, and then applies two machine learning algorithms (k-means and PLSR). The results show that the PLSR applied on the 350~400nm band does linearly predict diluted samples, according to their concentrations. We also use the PLSR model at 350~400nm band to fast classify the concentrations of the batch samples. The model and the technology could be applied to fast filter out low-concentration Hinoki hydrolat samples. In the future, we plan to incorporate GC-MS experiments to get the more precise concentration data to train our PLSR models so that the refined PLSR model might further improve its quantitative prediction accuracy.

Conflict of Interest

The authors declare there is no conflict of interest.

Funding

This study did not receive any funding.

References

1. Lu YC (2016) The yield model and wood utilization of Taiwan Red Cypress Plantations.
2. Chen YJ, Chang ST (2017) Distribution and characteristic comparisons of the endemic cypresses in Taiwan. *Taiwan Journal of Forest Science* 32: 71-86.
3. Quitain AT, Sato N, Daimon H, Fujie K (2003) Qualitative investigation on hydrothermal treatment of hinoki (*Chamaecyparis obtusa*) bark for production of useful chemicals. *J Agric Food Chem* 51(27): 7926-7929.

4. Matsubara Y, Sawabe A, Iba H, Iizuka Y (1990) Structure of terpenoid glycosides in the leaf of hinoki (*Chamaecyparis obtusa* sieb, et zucc.). *Agric Biol Chem* 54(2): 555-556.
5. Yu J, Komada H (1999) Hinoki (*Chamaecyparis obtusa*) bark, a substrate with anti-pathogen properties that suppress some root diseases of tomato. *Sci Hortic* 81(1): 13-24.
6. Ikei H, Song C, Miyazaki Y (2018) Physiological effects of touching hinoki cypress (*Chamaecyparis obtusa*). *J Wood Sci* 64: 226-236.
7. Ikei H, Song C, Miyazaki Y (2015) Physiological effect of olfactory stimulation by Hinoki cypress (*Chamaecyparis obtusa*) leaf oil. *J Physiol Anthropol* 34: 44.
8. Hieda T, Tazaki M, Morishita Y, Aoki T, Nagahama S (1996) Sesquiterpene alcohols from *Chamaecyparis obtusa* leaf oil. *Phytochemistry* 42(1): 159-162.
9. Kim SQ, Shin MK, Auh Q, Lee JY, Hong JP, et al. (2007) Effect of phytoncide on *Porphyromonas gingivalis*. *J Oral Med Pain* 32: 137-150.
10. Govil J, Bhattacharya S (2013) *Essential Oils I*, Studium Press, Studium Press LLC, USA.
11. Ohgaku A, Endo A, Hasegawa S, Hirose Y (1984) Diterpene production by callus of some plants belonging to Cupressaceae. *Agric Biol Chem* 48(10): 2523-2527.
12. Lee SH, Do HS, Min KJ (2015) Effects of essential oil from Hinoki cypress, *Chamaecyparis obtusa*, on physiology and behavior of flies. *PLoS One* 10(12): e0143450.
13. MS Bae, DH Park, CY Choi, GY Kim, JC Yoo, et al. (2016) Essential oils and non-volatile compounds derived from *Chamaecyparis obtusa*: broad spectrum antimicrobial activity against infectious bacteria and MDR (multidrug resistant) strains. *Nat Prod Commun* 11(5): 693-694.
14. Daimon H, Faisal M (2010) Preliminary investigation on the useful chemicals obtained from high-temperature and high-pressure water treatment of Hinoki (*Chamaecyparis obtusa*) bark. *J Chem Eng Environ* 7(4): 196-203.
15. Chung MJ, Cheng SS, Lin CY, Chang ST (2018) Profiling of volatile compounds from five interior decoration timbers in Taiwan using TD/GC-MS/FID. *J Wood Sci* 64: 823-835.
16. Lee M, Park S, Lee S, Lee H, Kil D (2014) Emission characteristics of volatile organic compounds by humidifier with using hinoki cypress extracts. *J Korean Wood Sci Technol* 42(6): 747-757.
17. Beale DJ, Pinu FR, Kouremenos KA, Poojary MM, Narayana VK, et al. (2018) Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics* 14(11): 152.
18. Ohtani Y, Ninomiya A, Shibayama Z, Sameshima K (2002) Chemical distinction of hinoki [*Chamaecyparis obtusa*] clones made by multivariate analysis of essential oil components. *Wood Ind* (2002): 246-250.
19. Emami SA, Massoomi H, Moghadam MS, Asili J (2009) Identification of volatile oil components from aerial parts of *Chamaecyparis lawsoniana* by GC-MS and ¹³C-NMR methods. *J Essent Oil Bear Plants* 12(6): 661-665.
20. Chen YJ, Lin CY, Cheng SS, Chang ST (2015) Rapid discrimination and feature extraction of three *chamaecyparis* species by static-HS/GC-MS. *J Agric Food Chem* 63(3): 810-820.
21. Alcalà M, Blanco M, Bautista M, González JM (2010) On-line monitoring of a granulation process by NIR spectroscopy. *J Pharm Sci* 99(1): 336-345.
22. Bian M, Skidmore AK, Schlerf M, Fei T, Liu Y, et al. (2010) Reflectance spectroscopy of biochemical components as indicators of tea (*Camellia sinensis*) quality. *Photogramm Eng Rem S* 76(12): 1385-1392.
23. Syu LL, Luo SH, Hsu JC (2013) Composition analysis and antioxidant activity of six hydrolats. *Hungkuang Acad Rev* 71(2013): 77-92.
24. Chien HY, Shih AT, Yang BS, Hsiao VK (2019) Fast honey classification using infrared spectrum and machine learning. *Math Biosci Eng* 16(6): 6874-6891.
25. Scikit-Learn (2018).
26. Ringnér M (2008) What is principal component analysis? *Nat Biotechnol* 26(3): 303-304.
27. K Means Clustering (2018).
28. Martens H, Martens M (2000) Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual Prefer* 11(1-2): 5-16.
29. Orange (2020).