

PAPER • OPEN ACCESS

Comparison of Analytical Ability of PLS and SVM Algorithm in Estimation of Moisture Content, Higher Heating Value, and Lower Heating Value of Cassava Rhizome Ground using FT-NIR Spectroscopy

To cite this article: Natrapee Nakawajana and Jetsada Posom 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **301** 012032

View the [article online](#) for updates and enhancements.

Comparison of Analytical Ability of PLS and SVM Algorithm in Estimation of Moisture Content, Higher Heating Value, and Lower Heating Value of Cassava Rhizome Ground using FT-NIR Spectroscopy

Natrapee Nakawajana¹, and Jetsada Posom^{2,3,*}

¹Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

²Applied Engineering for Important Crops of the North East Research Group, Khon Kaen University, Khon Kaen 40002, Thailand

³Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand

*E-mail: jetspo@kku.ac.th

Abstract. FT-NIR spectroscopy coupled with chemometrics analysis was used for non-destructive estimation of moisture content (MC), higher heating value (HHV) and lower heating value (LHV) of cassava rhizome ground. The goal of this study was compared to the analytical ability of both algorithm between PLS and SVM. The purpose was to find the effective modelling technique. The outcome was found that PLS and SVM provided good accuracy in evaluation of energy properties, and could be utilized for quality assurance. PLS algorithm gave slightly higher accuracy than SVM algorithm for the prediction of MC, HHV, and LHV. PLS regression generated no difference between measured and predicted value. PLS and SVM regression showed R^2 between 0.90-0.98 and 0.84-0.90 for all parameters, respectively. The pre-processing of 2nd derivative was suitable for the PLS and SVM regression to the modelling.

1. Introduction

Currently, farmer plants cassava for the commercial purposes, they deliver intact root to the cassava starch factory for producing tapioca flour. After harvesting, cassava root was required for selling, and the remaining part was cassava rhizomes, leaves, and stem. Farmers often burned, leading to greenhouse effect. Nowadays, cassava rhizome has been promoted as renewable resource. Cassava rhizomes could be used as biomass feedstock because they had the high energies content of approximately 17.48 MJ/kg, volatile matter of 75%, ash of 5.6% and sulfur (S) of 0.08% [1].

The important energy characteristic for effective biomass utilization included high heating value (HHV) and the low heating value (LHV), and moisture content (MC). HHV can be measured using bomb calorimetry, but takes a long time for measuring process, high cost of analysis approximately 1,000 baths/sample. HHV can be converted to LHV when its MC was known. Rapid method can be used to estimate energy properties of biomass was near infrared spectroscopy (NIRs) technique. Many research reported the use of NIRs coupled PLSR because of giving a good prediction, such as HV of



straw [2,3], HHV of *Miscanthus* and short rotational coppice Willows [4], HHV of sorghum [5], the LHV of bamboo [6] and HHV of rice husk [7].

Currently, the comparison performance of two regression algorithms was investigated i.e. partial least squares (PLS) and support vector machines (SVM). PLSR was widely utilized for model development, the fact was that it was not complicated, but require lot of sample set. PLS is linear model, however, it was not always the best algorithm, if a complex as nonlinear model was required to model the relationship between the measured spectra and interest [8]. Meanwhile, nonlinear model with a high ability was proposed for the SVM algorithm [9], it could be used where the relationship between spectra and interest is presented as a non-linear. In the field of NIRs to the bioenergy, there were no applications of SVM for regression creation. The comparison of PLS and SVM was studied with raw milk [8], and acerola fruit [10]. Resulting was that both techniques gave good prediction.

The aim of this research was to study the analytical ability of FT-NIR spectrometer in estimating the HHV, LHV, and MC of ground cassava rhizomes. For these object, the spectra-processing coupled with multivariate analysis was not only PLS regression, but also SVM regression, in which algorithm was non-linear regression. These experiment could be appropriate for estimation of the energy properties. The best of our knowledge was that between SVM and PLS with same sample set, wavenumber range, and spectral pre-processing, which algorithm could be more appropriate for the estimation of MC, HHV, and LHV. The results were critically discussed.

2. Materials and methods

2.1. Sample Preparation

A set of 60 cassava root samples were randomly selected from different area in Thailand. Then, each sample was chopped into smaller size and dried it by hot air oven (ULM 500, Memmert, Germany). After that, sample chips were milled using a sieve with a mesh diameter of 2 mm (Pulver-isette14, Fritsch, Germany). The samples were kept in aluminium bag until experiment including the NIR absorbance measurement by FT-NIR spectrometer, the HHV measurement by bomb calorimeter and the moisture content (MC) measurement for reference data using hot air oven.

2.2. NIR Absorbance measurement

Each sample was filled in cup (43 mm in diameter and 50 mm in height) and scanned the absorbance using FT-NIR spectrometer (MPA, Bruker, Germany) at the wavenumber range between 12,500 – 3,600 cm^{-1} with a resolution of 8 cm^{-1} and the average of 32 scans. The NIR absorbance was recorded in $\log(1/R)$ unit, where R is reflection intensity. Then, only sample at the bottom of the quartz cup was used for the HHV measurement and the moisture content measurement.

2.3. Higher heating value measurement

After absorbance measurement, a part of the sample at the bottom of the quartz cup was randomly collected by weighting about 0.5–1.0 g, then put into the pellet press machine to the pelletization. The HHV was measured by bomb calorimeter (C200, IKA, Germany) with 2 repeats per sample. After processing, the HHV was shown in J/g unit.

2.4. The moisture content measurement

Another part of sample at the bottom of the quartz cup about 3 g was obtained and dried at 105 °C for 24 hrs until get a constant weight. MC was measured by hot air oven (ULM 500, Memmert, Germany) with 3 repeats per sample. The moisture content (%MC) of the cassava rhizome based on wet basis was calculated as follows equation 1.

$$\%MC = (A-B)/A \quad (1)$$

where A is the original mass of the cassava root (g) and B is the dried mass of the cassava rhizome ground (g).

2.5. The Lower Heating value (LHV)

LHV is the total energy of biomass released coupled with no water vaporization. It can be converted from HHV form by minus energy of water vaporization, as follows:

$$\text{LHV}_{\text{wb}} = \text{HHV}_{\text{wb}} - 2433 \times \text{MC} \quad (2)$$

where, MC is moisture content (%wb).

2.6. Data analysis

The comparison of the analytical performance was investigated between PLS and SVM algorithm. The PLS was applied because of good ability. SVM algorithm generated for the non-linear model in the high dimensional feature space [11]. Both analytical method was tested by one full cross validation. PLS models were established using OPUS software (version 7.0.129, Germany). SVM algorithm analysis was created by Unscrambler@X 10.3. It was generated using ϵ -SVM (epsilon-support vector machine), with the following kernel: Radian basic function, C value=1, ϵ =0.1. The raw spectra and pre-treated spectra were utilized before establishing the calibration model. Spectral pre-processing techniques used included baseline offset, standard normal variate (SNV), 1st derivative, and 2nd derivative; multiplicative scatter correction (MSC). The statistics for assessment of performance consisted of the determination of the coefficient of determination (R^2), root mean square error of cross validation (RMSECV). The performance of two algorithms (PLS vs SVM) was compared using the statistical parameter of RMSECV by testing a hypothesis between PLS prediction vs SVM prediction; reference value vs PLS prediction; and reference value vs SVM prediction, respectively. The fact was that RMSECV was the accuracy of future sample, then it could be used to test the model performance.

The null hypothesis (H_0) was that the slope (b_1) between measured and predicted value is no significant differences from 1, and intercept (b_0) is no difference significantly from 0, at the 99% confidence levels.

3. Results and Discussion

3.1. NIR Spectra information

Raw spectra and 2nd NIR spectra were shown in figures 1a and b, respectively. The derivative technique helps to separating overlapping peaks and baseline variation, the obvious peak occurring such as 5292 cm^{-1} (1890 nm), 4952 cm^{-1} (2019 nm), 4498 cm^{-1} (2223 nm), and 4142 cm^{-1} (2414 nm), related to O-H hydrogen bonding between water and polyvinyl alcohol OH, O-H stretching from methanol, N-H combination band, C-H aromatic [12], respectively.

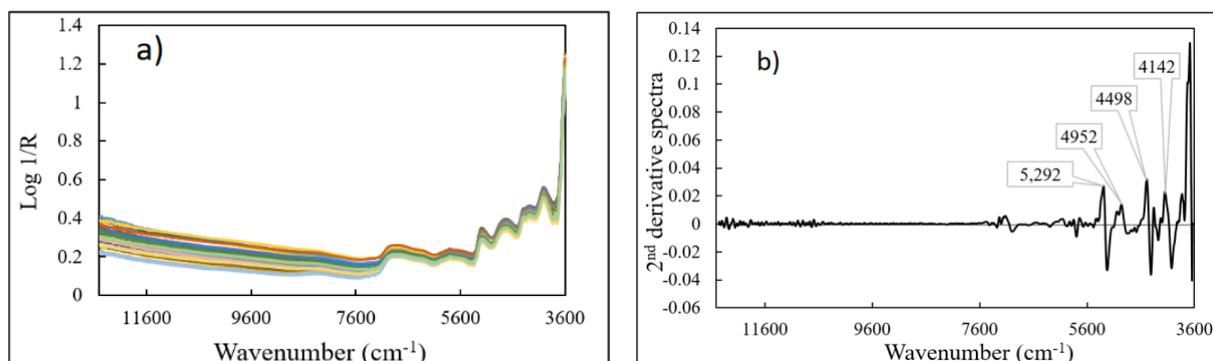


Figure 1. a) Raw spectra, b) average 2nd derivative spectra of ground cassava rhizome.

3.2. Reference data

The statistical data for model development was illustrated in table 1, representing MC, HHV, and LHV. According to the HHV and LHV result of cassava rhizome, it had high energy content comparing to biomass feedstock; 14508.5 to 15728 J/g for rice husk sample [7], 16,891 to 17,662 J/g for *Leucaena leucocephala* pellets [13], 16534-18357 J/g for bamboo biomass [14], 17.8-20.4 MJ/kg for mixing olive pomace and olive tree pruning [8], 15.67–16.99 MJ/kg for sorghum biomass [9]. If biomass representing with high HHV and LHV, it could be suitably used as renewable energy resource, such as biomass pellet, biomass chip and so on.

The repeatability (rep) of reference value and maximum coefficient of determination (R^2_{\max}) of MC and HHV equals to 0.15% wb and 0.89; 127.013 J/g and 0.85, respectively, where $R^2_{\max} = \frac{SD_y^2 - rep^2}{SD_y^2}$, SD_y is standard deviation of measured value, rep was standard deviation of different between duplicates. R^2_{\max} helps developer to decide that whether interest model could be developed or not. If high error in reference method, the first think has to search for the precision and accuracy of reference method. As seen the result, the error from laboratory method was 11% (1-0.89), and 15% (1-0.85), respectively. Consistently, the repeatability of the reference method should be lower than standard deviation. One more think about repeatability, due to the cassava rhizome ground sample may be nonhomogeneous, this may case low R^2_{\max} .

Table 1. Statistical data of MC, HHV, and LHV used for model development.

Parameter	N	Max	Min	mean	SD
MC, %wb	45	6.8041	2.1297	4.5044	1.3825
HHV, J/g	49	17077.5	13583	16173.3	697.4
LHV, J/g	49	16992.2	13453.8	16069.5	714.6

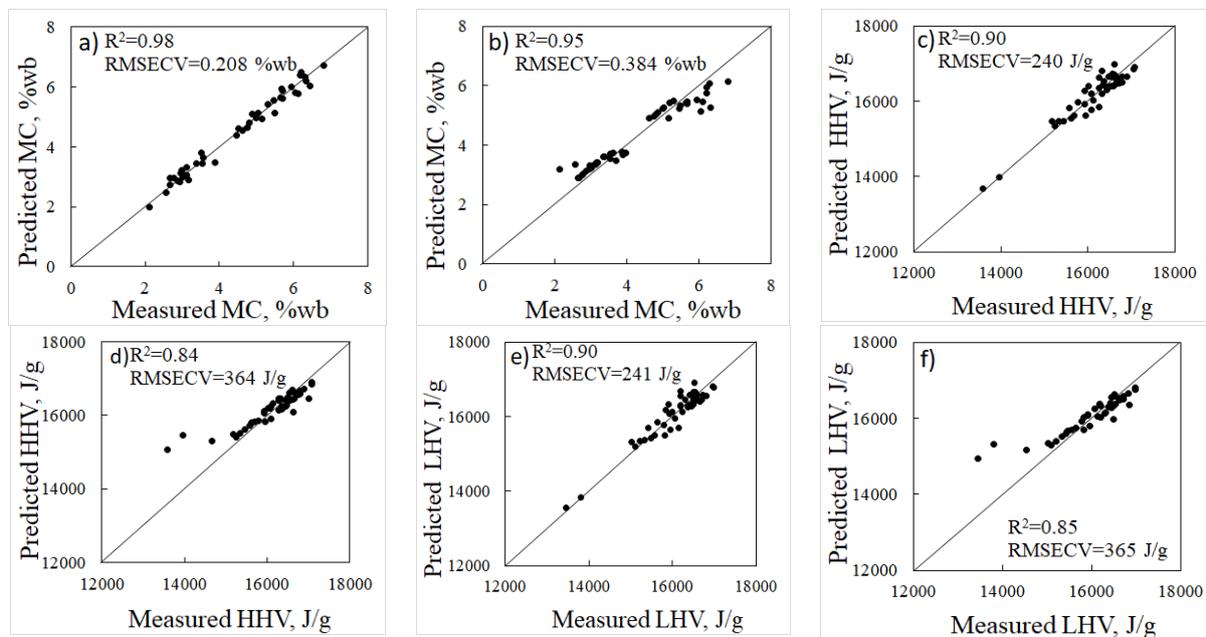


Figure 2. Scatter plot between measured value versus predicted value, a) MC model developed from PLS, b) MC model developed from SVM, c) HHV model developed from PLS, d) HHV model developed from SVM, e) LHV model developed from PLS, and f) LHV model developed from SVM.

3.3. NIR model for estimation of MC, HHV, and LHV

The reference value and its corresponding spectra were applied for model creation. The calibration equation for MC, HHV, and LHV were generated using PLS and SVM and validated by one-full cross validation. The model parameters created from pre-treated spectra were demonstrated in table 2. It was possible to compare between two algorithms in prediction of MC, HHV, and LHV. Figures 2a, b, c, d, e, and f illustrated the measured versus predicted value of MC, HHV, and LHV of PLS and SVM algorithm, respectively. The MC model optimized by PLS and SVM gave R^2 of 0.98 and 0.95; 0.90 and 0.84; 0.90 and 0.85, respectively. R^2 between 0.92-0.96 were good and could be applied for most application and quality assurance [15], R^2 between 0.83 and 0.90 could be used with caution for most applications [15].

PLS provided the better result compared to SVM. They were slightly different on the coefficient of determination and RMSECV. It is underlined that the PLS model is higher performance than SVM model slightly. However, spectral pre-processing method is important, the better result of NIR and SVM model was found where pretreated spectra method of 2nd derivative was applied. The obvious peak finding can be done by the analysis of 2nd derivative, this information benefits the model development in PLS and SVM.

Table 2. Cross validation model for FT-NIR of two algorithms: SVM and PLS.

PLS	Cross validation test			
	Pre-treatment	R^2	RMSECV	Bias
MC, %wb	2 nd derivative	0.98	0.208	-5.3×10^{-6}
HHV, J/g	SNV	0.90	240	0.522
LHV, J/g	SNV	0.90	241	0.224
SVM				
MC, %wb	2 nd derivative	0.95	0.384	-0.0119
HHV, J/g	2 nd derivative	0.84	364	-47.2
LHV, J/g	2 nd derivative	0.85	365	-47.6

To examine that whether measured value and predicted value of MC, HHV, and LHV parameters were different. The coefficient (b_0 : intercept and b_1 : slope) of simple linear regression between measured value versus predicted value was determined. The assumption was that a good prediction should provide b_0 equals 0, and b_1 equals 1. The Z-test method was used to test hypothesis: to examine whether means of two populations were different [16]. As mentioned that null hypothesis (H_0) was: the slope (b_1) was not significantly different from 1, and intercept (b_0) was not significantly from 0. The confidence level of 99% was used. Table 3 shows the result of single factor comparison for both of the PLS and SVM algorithm. The PLS algorithm model provided H_0 accepted for the MC, HHV, and LHV, means that the slope and intercept is not different significantly from 1 and 0, respectively, at 99% confidence levels.

Meanwhile, SVM algorithm model gave H_0 rejected, mean that means that the slope and intercept is different significantly from 1 and 0, respectively, at 99% confidence level. It can be seen that even if SVM model gave rather high R^2 , predicted value and referenced value was also had the different in slope and intercept from 1 and 0, respectively.

Table 3. Single factor comparison for both of the algorithm. A joint test on slope and intercept values of the regression lines, at 99% confidence level, was performed.

Parameter	Z-TEST	b ₁	b ₀
MC	SVM vs PLS	H ₀ rejected	H ₀ rejected
	Ref vs PLS	H ₀ accepted	H ₀ accepted
	Ref vs SVM	H ₀ rejected	H ₀ rejected
HHV	SVM vs PLS	H ₀ rejected	H ₀ rejected
	Ref vs PLS	H ₀ accepted	H ₀ accepted
	Ref vs SVM	H ₀ rejected	H ₀ rejected
LHV	SVM vs PLS	H ₀ rejected	H ₀ rejected
	Ref vs PLS	H ₀ accepted	H ₀ accepted
	Ref vs SVM	H ₀ rejected	H ₀ rejected

Ref is reference value, MC is moisture content, HHV is higher heating value, LHV is lower heating value, b₁ is slope, b₀ is intercept.

4. Conclusions

NIR spectroscopy coupled with linear and non-linear model is could be an interest device to evaluate energy properties of biomass by non-destructive. The PLS regression provided the better results for model development in parameter of MC, HHV, and LHV. However, SVM regression also gives a high R², it was slightly lower than PLS regression. It was found that the SVM should be developed with the pre-processing spectra of 2nd derivative. For the hypothesis test, PLR algorithm gave the result with no difference significant between measured and predicted value. The PLS could be used to the prediction of energy properties of cassava rhizome ground where compared with SVM regression. further study.

References

- [1] Mistry of Energy. Available online: <http://webkc.dede.go.th>. (accessed 18.02.19).
- [2] Huang C, Han L, Yang Z and Liu X 2008 *Energy Convers. Manag* **49** 3433 – 3438
- [3] Huang C Han L, Yang Z and Liu X 2009 *Waste Manage* **29** 1793 – 1797
- [4] Everard C D, McDonnell K P and Fagan C C 2012 *Biomass Bioenergy* **45** 203 – 211
- [5] Zhang K, Zhou L, Brady M, Xu F, Yu J and Wang D 2017 *Energy* **118** 1353– 1360
- [6] Posom J and Sirisomboon P 2017 *Energy* **121** 147–158
- [7] Nakawajana N, Posom J and Paeoui J 2018 *Eng J* **22** 45–56
- [8] Barbanera M, Lascar Near-infrared technology-getting the best out of light. Nanaimo o E, Stanzione V, Esposito A, Altieri R and Bufacchi M 2016 *Renew Energ* **88** 185-191
- [9] Zhang K, Zhou L, Brady M, Xu F, Yu J and Wang D 2017 *Energy* **118** 1353-1360
- [10] Malegori C, Marques E J N, Freitas S T, Pimentel M F, Pasquini C and Casiraghi E 2017 *Talanta* **165** 112-116
- [11] Bassbasi M, Platikanov S, Tauler R and Oussama A 2014 *Food Chem* **146** 250-254
- [12] Workman J and Weyer J R L 2007 Practical Guide to Interpretive Near-Infrared Spectroscopy. Boca Raton, FL: Taylor and Francis 240-262
- [13] Posom J, Shrestha A, Saechua W and Sirisomboon P 2016 *Energy* **107** 464-472
- [14] Posom J and Sirisomboon P 2017 *J Near Infrared Spec.* **25** 301–310
- [15] Williams P 2007 Near-infrared technology-getting the best out of light. Nanaimo, British Columbia, and Winnipeg, Manitoba, Canada: PDK Grain.
- [16] Terms Beginning With Z. Available online: <https://www.investopedia.com/terms/z/z-test.asp>. (accessed 18.02.19).

Acknowledgments

All authors appreciate to thank you the financial support from the King Mongkut's Institute of Technology Ladkrabang Research Fund, Thailand. This study was experimentally supported by Near Infrared Spectroscopy Research Center for Agricultural Product and Food at King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. We thank the Applied Engineering for Important Crops of the North East Research Group (AENE), Khon Kaen university, Khon Kaen, Thailand, for data analysis.