

# Multiplicative Scatter Correction Improves Near-Infrared Spectroscopy–Based PLS Models for Acetic Acid Quantification in Jamblang Vinegar

Noer Octaviani Maliza<sup>1</sup>, Muhammad Ikhsan Sulaiman<sup>2,6\*</sup>, Dewi Yunita<sup>2</sup>, Agus Arip Munawar<sup>3</sup>, Rita Andini<sup>4†</sup>, Safrida<sup>5</sup>

<sup>1</sup>Department of Agricultural Product Technology, Universitas Teuku Umar, 23681 Aceh Barat, Indonesia

<sup>2</sup>Department of Food and Agricultural Product Technology, Universitas Syiah Kuala, 23111 Banda Aceh, Indonesia

<sup>3</sup>Department of Agricultural Engineering, Universitas Syiah Kuala, 23111 Banda Aceh, Indonesia

<sup>4</sup>Research Organization for Environmental and Life Sciences, Research Centre for Applied Botany, National Agency for Research and Innovation, 16911 Cibinong, Indonesia

<sup>5</sup>Department of Biology Education, Universitas Syiah Kuala, 23111 Banda Aceh, Indonesia

<sup>6</sup>Halal Research Centre, Universitas Syiah Kuala, 23111 Banda Aceh, Indonesia

**Abstract.** Jamblang (*Syzygium cumini* L.) vinegar is a potential functional fermented product that requires rapid and reliable quality assessment. This study developed a near-infrared spectroscopy (NIRS)-based method combined with partial least squares regression (PLSR) to quantify acetic acid content in jamblang vinegar. Vinegar samples representing a wide range of acetic acid concentrations were obtained through varied fermentation conditions and analyzed using a reference enzymatic method. NIR spectra were preprocessed using Multiplicative Scatter Correction (MSC) to reduce scattering effects caused by sample heterogeneity. PLSR models constructed from MSC-treated spectra showed improved predictive performance compared with uncorrected spectra, achieving a coefficient of determination ( $R^2$ ) of 0.995 and a residual predictive deviation greater than 2.3. Regression coefficient analysis identified wavelength regions associated with O–H and C–H overtone vibrations as key contributors to acetic acid prediction. The results demonstrate that NIRS coupled with MSC-enhanced PLSR provides a rapid, non-destructive, and accurate approach for acetic acid determination in jamblang vinegar, supporting its application for routine quality control and fermentation monitoring.

## 1 Introduction

*Syzygium cumini* L., known as jamblang, is a tropical fruit renowned for its substantial pharmacological properties, including antioxidant, antidiabetic, anticancer, and anti-inflammatory activities, primarily due to its high concentrations of polyphenols and anthocyanins [1]. Despite its rich bioactive composition, jamblang fruit is severely underutilized owing to its short shelf life of only 2–3 days and astringent flavor profile, which severely restrict large-scale consumption. To maximize its economic value and minimize post-harvest losses, processing into value-added products is crucial. While jamblang wine production has been successfully demonstrated [2], vinegar fermentation provides a superior alternative, delivering substantial functional and commercial benefits.

Vinegar is a globally popular fermented product, traditionally produced across Europe, Asia, and the United States, where apple and balsamic varieties firmly dominate the international market [3]. Beyond its role as a flavor enhancer, vinegar reliably serves as a natural preservative and a vital ingredient in cosmetics and nutraceuticals [4]. The escalating demand for superior product quality and production efficiency has propelled the industry-wide shift from conventional destructive analytical methods to rapid, environmentally sustainable alternatives, such as vibrational spectroscopies [5].

\* Corresponding author: [ikhsan.sulaiman@usk.ac.id](mailto:ikhsan.sulaiman@usk.ac.id)

† Corresponding author: [rita.andini@brin.go.id](mailto:rita.andini@brin.go.id)

Acetic acid, the primary component of vinegar, is primarily responsible for its potent antimicrobial, preservative, and health-promoting properties—including proven antidiabetic effects and blood cholesterol reduction [3,6]. Therefore, its rapid and precise quantification during fermentation is indispensable for ensuring product quality, enabling effective process control [5], and substantiating functional claims. However, traditional analytical methods invariably demand time-consuming and destructive sample preparation [5], highlighting the critical need for efficient spectroscopic solutions. Near-infrared and Fourier-transform infrared spectroscopy, coupled with chemometric techniques such as partial least squares regression, deliver rapid, non-destructive, and highly accurate alternatives for evaluating critical quality parameters in vinegar production [7,8].

This study establishes a highly robust near-infrared spectroscopy-based partial least squares regression model to accurately predict acetic acid content—the pivotal quality parameter—in jamblang vinegar, effectively overcoming limitations of traditional methods [5,7,8] while enabling real-time process control, stringent quality assurance, and reliable authentication for this innovative product. This approach offers a significant advancement over labor-intensive physicochemical detection methods, which have traditionally been the primary monitoring techniques for solid-state fermentation processes in vinegar production [9,10].

## 2 Materials and Methods

### 2.1 Jamblang Fruit Collection and Processing

Fresh fruits of *Syzygium cumini* L. were harvested from Aceh Besar District, Indonesia. To produce vinegar samples spanning a broad range of acetic acid concentrations suitable for chemometric modeling, diverse fermentation protocols were employed. Fruits were processed either intact or as homogenized pulp to investigate the influence of fruit matrix on fermentation kinetics. Alcoholic fermentation was initiated using two distinct yeast inocula: commercial *Saccharomyces cerevisiae* and yeasts derived from traditional fermented rice or cassava. This was followed by acetous fermentation with a palm vinegar starter culture to enhance microbial diversity. Fermentations adhered to the Orléans process under controlled laboratory conditions, with incubation periods extended to 40 days to yield varied acetic acid levels.

### 2.2 Physicochemical Analysis of Jamblang Vinegar

Acetic acid content was determined using a commercial enzymatic assay kit (K-ACETAF, Megazyme, Ireland) following the manufacturer's protocol, a standard reference method for organic acid quantification in fermented products.

### 2.3 Spectroscopic Data Acquisition

Spectra were recorded using an IPTEK T-1516 Fourier-transform infrared spectrometer and processed with Thermo Integration® software. Samples were scanned from 4000 to 400  $\text{cm}^{-1}$  at a resolution of 3  $\text{cm}^{-1}$ , with each spectrum acquired as an average of 10 scans to enhance the signal-to-noise ratio and ensure high data quality for subsequent chemometric analysis. The spectral range of 2500–1000  $\text{cm}^{-1}$  was selected, as it encompasses vibrational modes pertinent to a wide array of functional groups, including those characterizing organic acids like acetic acid present during the fermentation process [7]. This region is particularly informative for the quantitative analysis of vinegar constituents, as demonstrated by previous research utilizing FT-IR spectroscopy for the classification and differentiation of various vinegars [8].

### 2.4 Spectral Processing and Model Development

Spectral preprocessing was performed using Multiplicative Scatter Correction, a widely used technique to minimize light scattering effects arising from sample physical properties such as particle size and surface heterogeneity, which often distort near-infrared and mid-infrared spectra. MSC is based on linear regression

of each spectrum ( $x_i$ ) against a reference spectrum (commonly the mean spectrum,  $x_{ref}$ ), expressed modeled as

$$x_i = a + b \cdot x_{ref} + \epsilon \tag{1}$$

where  $a$  is the offset (baseline shift),  $b$  the slope (multiplicative scatter effect), and  $\epsilon$  the residual error. The corrected spectrum is then computed as

$$x_{i,MSC} = \frac{x_i - a}{b} \tag{2}$$

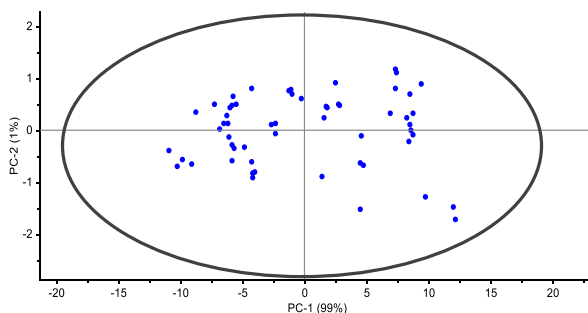
thereby removing additive and multiplicative variations, improving spectral comparability, and enhancing the performance of subsequent chemometric models like Principal Component Analysis and Partial Least Squares Regression. Recent applications in food analysis confirm that MSC substantially improves prediction accuracy for key quality parameters [11,12]. Outliers were identified via PCA using Hotelling's  $T^2$  ellipse. Calibration models were developed with PLS-R and evaluated based on the coefficient of determination ( $R^2$ ) and Root Mean Square Error of Calibration. All data analyses were conducted using Unscrambler® software.

### 3 Results

#### 3.1 Outlier Detection

Hotelling's  $T^2$  statistic identified spectral outliers as deviations beyond the 95% confidence limit of the calibration set, ensuring robust models. No samples exceeded this limit, so none were excluded, preventing bias and maintaining chemical variation representation in jamblang vinegar [13].

Figure 1 shows all samples within the 95% confidence ellipse, confirming no outliers. This indicates consistent, reproducible spectra with minimal noise or errors. In NIR studies, outliers from instrument fluctuations, preparation errors, or fruit heterogeneity can distort calibration and accuracy [14]. No outliers validate MSC preprocessing in reducing scatter and artifacts [15,16], yielding a representative dataset free of confounders [17,18].



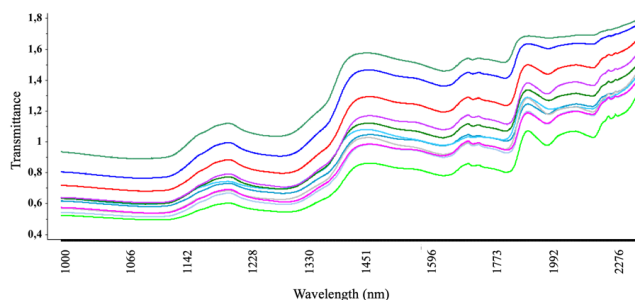
**Fig. 1.** Principal Component Analysis (PCA) score plot with Hotelling's  $T^2$  ellipse showing outlier detection in vinegar spectral data.

#### 3.2 Spectroscopic Profiles of Jamblang Vinegar

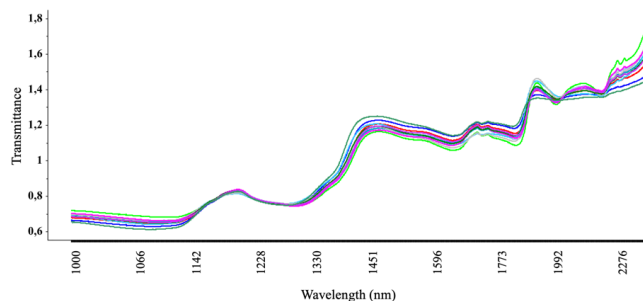
Figure 2 shows transmittance spectra of vinegar samples in the 1000–2500 nm range. Raw spectra had high variability, with rough patterns and baseline shifts due to scattering, heterogeneity, and path length variations, obscuring chemical signals.

Multiplicative Scatter Correction, shown in Figure 3, normalized spectra to a mean reference, reducing scatter and improving uniformity. Post-MSC spectra clustered tightly, revealing chemical variations for quantitative analysis [16].

MSC-treated spectra, while similar to raw ones in patterns, were more reliable for chemometrics. Uniform features reflect consistent fermentation composition. MSC boosts model robustness, as in other food applications using similar preprocessing [10,19,20].



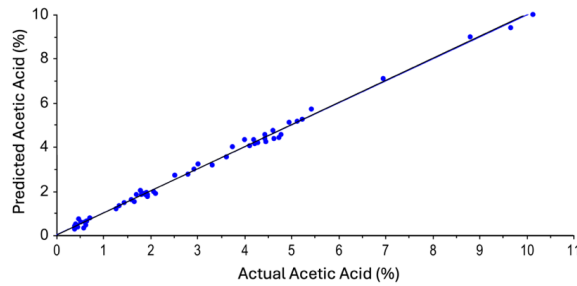
**Fig. 2.** Raw near-infrared (NIR) transmittance spectra of jamblang vinegar samples in the 1000–2500 nm wavelength range.



**Fig. 3.** Multiplicative Scatter Correction (MSC)-processed NIR spectra of jamblang vinegar samples showing improved homogeneity and reduced scattering effects.

### 3.3 PLS-R Model Performance for Acetic Acid Prediction

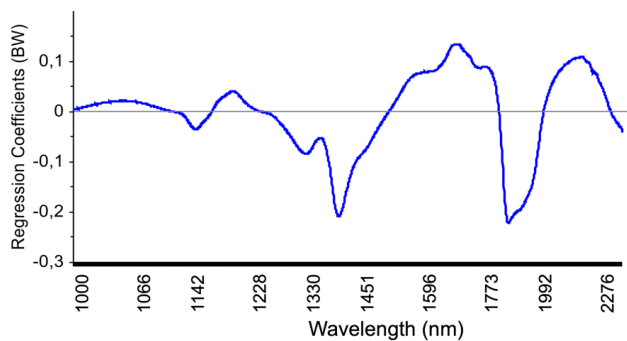
Model performance was assessed using the correlation coefficient ( $r$ ), coefficient of determination ( $R^2$ ), root mean square error of calibration, and residual predictive deviation. For uncorrected spectra, the model achieved  $r = 0.996$ ,  $R^2 = 0.992$ , and  $RMSEC = 0.197$  with 20 latent variables, indicating excellent predictive power. MSC preprocessing further improved results to  $r = 0.997$  and  $R^2 = 0.995$ , with a modest error reduction (Figure 4). These gains show that MSC bolsters PLSR robustness by curbing scattering effects and enhancing the linear spectral-acetic acid relationship. The RPD of 2.34 surpasses thresholds for reliable quantitative prediction, supporting its use in vinegar quality evaluation.



**Fig. 4.** Predicted versus reference values of acetic acid concentration in jamblang vinegar using Partial Least Squares Regression (PLSR) with raw and MSC-corrected spectra.

To interpret the PLSR model, regression coefficients revealed spectral regions most correlated with acetic acid levels. As shown in Figure 5, prominent positive and negative peaks emerged at 1346–1432 nm and 1638–1838 nm, corresponding to O–H and C–H overtone/combination vibrations characteristic of acetic acid and similar compounds. These regions offer chemical insights into the model and serve as benchmarks for future vinegar or fermented product studies.

MSC combined with PLSR delivered a highly accurate, interpretable calibration model. Coefficient analysis verified the chemical significance of spectral features and identified key wavelengths for rapid or portable NIR detection. These results affirm NIRS coupled with chemometrics as a non-destructive, dependable tool for vinegar quality monitoring.



**Fig. 5.** Regression coefficient plot of the PLSR calibration model showing key wavelength regions associated with acetic acid in jamblang vinegar spectra.

## 4 Discussion

The results of this study demonstrate that NIRS combined with chemometric modeling provides an accurate, non-destructive method for quantifying acetic acid concentration in jamblang vinegar. The calibration model exhibited high predictive performance following Multiplicative Scatter Correction, with low RMSEC and an RPD exceeding 2.0—indicating suitability for quantitative analysis [11,12].

These findings align with prior research highlighting NIRS's ability to predict organic acids and other quality parameters in fermented beverages at comparable accuracy levels [7,8]. For example, FT-NIR spectroscopy has successfully predicted titratable acidity and polyphenols in fruit vinegars, achieving  $R^2$  values above 0.98 [10].

The identified wavelength regions, linked to O–H and C–H vibrations, confirm the PLSR model's chemical interpretability. Comparable spectral bands appear in NIRS studies of acetic acid and ethanol in alcoholic and fermented beverages, where O–H overtones dominate [7].

MSC preprocessing improvements emphasize the value of scatter correction for robust calibration. This agrees with studies showing that MSC enhances predictions of fat, protein, acidity, and sugars in complex foods [11,15]. Similarly, scatter correction and derivatives minimize baseline shifts and scattering in alcoholic beverages, yielding reliable models [14,20].

Unlike labor-intensive, time-consuming, and destructive traditional methods, NIRS-PLSR enables rapid quality control. Reviews note that chromatographic and enzymatic assays for organic acids demand extensive preparation and reagents despite their accuracy [7]. Here, NIRS matched this precision more efficiently, favoring routine vinegar monitoring.

This study's novel contribution applies NIRS-PLSR to jamblang vinegar—an underexplored tropical fruit. To our knowledge, it is the first to develop a dedicated chemometric model and pinpoint acetic acid-correlated wavelengths, providing spectral insights for jamblang products and extending NIRS to other underutilized fruits.

The high accuracy also signals NIRS potential for real-time fermentation monitoring, as shown in wine and rice vinegar for tracking kinetics, acidity, and bioactives [10]. Applying this to jamblang vinegar enables value addition, process optimization, consistent quality, and stronger health claims.

## 5 Conclusion

In conclusion, the integration of NIRS with chemometric models, particularly PLSR coupled with MSC preprocessing, offers a highly effective and non-destructive analytical platform for quantifying acetic acid in jamblang vinegar.

## Acknowledgements

The authors gratefully acknowledge the financial support provided by the Research and Community Service Institute of Universitas Syiah Kuala under contract no. 318/UN11.2.1/PT01.03/PNBP/2021. The authors also express sincere appreciation to the National Research and Innovation Agency (BRIN) for the valuable collaboration facilitated through the Implementation Agreements no. 193/V/KS/11/2022 and no. 376/UN11/HK.02.06/2022.

## References

1. M. Qamar, S. Akhtar, T. Ismail, M. Wahid, M. W. Abbas, M. S. Mubarak, Y. Yuan, R. Barnard, Z. M. Ziora, and T. Esatbeyoglu, *Foods* **11**, 378 (2022).
2. T. S. de O. Brandão, L. S. Pinho, A. Ferreira-Silva, J. L. de Souza, C. A. Rosa, E. Teshima, H. N. Brandão, and J. M. David, *BioResources* **12**, 7069 (2017).
3. D. Saha and P. K. Das, in *IntechOpen eBooks* (IntechOpen, 2023).
4. G. Chen, F. Zheng, B. Lin, Y. Yang, X.-C. Fang, K. K. Verma, and L. Yang, *Frontiers in Nutrition* **10**, (2023).
5. C. Menozzi, G. Foca, R. Calvini, L. Catellani, A. Bezzecchi, and A. Ulrici, *Food Analytical Methods* **17**, 1171 (2024).
6. S. Grassi and C. Alamprese, *Food and Bioprocess Technology* **17**, 479 (2023).
7. Ç. Çavdaroglu and B. Özen, *Microchemical Journal* **171**, 106886 (2021).
8. E. Durán, R. Castro, R. N. Marín, M. Palma, and C. G. Barroso, *Journal of the Science of Food and Agriculture* **90**, 712 (2010).
9. E. Kim, K. Cho, S. J. Kwon, S.-H. Seo, S.-E. Park, and S.-H. Seo, *LWT* **135**, 110081 (2020).
10. S. Fan, T. Pan, and G. Li, *International Journal of Food Engineering* **16**, (2020).
11. Ö. Çataltaş and K. Tütüncü, *European Journal of Science and Technology* (2021).
12. X. Chen, C. Li, X. Liu, Y. Su, Z. Sun, L. Yuan, and S. Wang, *Quality Assurance and Safety of Crops & Foods* **15**, 139 (2023).
13. T. Basile, A. D. Marsico, M. F. Cardone, D. Antonacci, and R. Perniola, *Foods* **9**, 98 (2020).

14. P. Mishra, D. N. Rutledge, J. Roger, K. Wali, and H. A. Khan, *Talanta* 229 , 122303 (2021).
15. A. A. Munawar, K. Kusumiyati, Andasuryani, Y. Yusmanizar, and Adrizal, *Acta Technologica Agriculturae* 27 , 23 (2024).
16. K. Kusumiyati, Y. Hadiwijaya, and I. E. Putri, *Jurnal Biodjati* 4 , 89 (2019).
17. R. John, R. Bhardwaj, C. Jeyaseelan, H. Bollinedi, R. Singh, and G. Singh, *Research Square (Research Square)* (2023).
18. S. S. R. Lazim, N. M. Nawi, S. K. Bejo, A. R. M. Shariff, and N. Abdullah, *International Food Research Journal* 29 , 1372 (2022).
19. H. P. Sari, Y. A. Purwanto, and I. W. Budiastira, *Jurnal Agritech* 36 , 294 (2016).
20. R. Hayati, A. A. Munawar, and F. Fachruddin, *Data in Brief* 30 , 105571 (2020).