

Acta Technologica Agriculturae 2
Nitra, Slovaca Universitas Agriculturae Nitriae, 2026, pp. 98–105

PREDICTION OF NITROGEN CONTENT IN WETLAND AND DRYLAND SOILS OF JAVA ISLAND USING NEAR-INFRARED SPECTROSCOPY AND MULTIVARIATE ANALYSIS

Adnan ADNAN*, Yaya SURYANA, Abdul AZIZ, Taslim ROCHMADI, Arie Rakhman HAKIM, Fahrodji FAHRODJI, Amrullah KAMARUDDIN, Wenny OKTAVIANI, Nizam GHAZALI, Adim HADI, Ardani Cesario ZUHRI, Galang Ilman ISLAMI

National Research and Innovation Agency, Research Organisation for Energy and Manufacture, South Tangerang, Indonesia, yaya002@brin.go.id (Y.S.), abdu002@brin.go.id (A.Z.), tasl001@brin.go.id (T.R.), arie035@brin.go.id (A.R.H.), fahr003@brin.go.id (F.F.), amru001@brin.go.id (A.K.), wenn001@brin.go.id (W.O.), niza001@brin.go.id (Z.G.), adim002@brin.go.id (A.H.), arda004@brin.go.id (A.C.Z.), gala001@brin.go.id (G.I.I.)

*correspondence: adna002@brin.go.id

This study evaluated the potential of near-infrared spectroscopy (NIRS) combined with multivariate analysis to predict soil nitrogen (N) content in both wetland and dryland systems of Java Island, Indonesia. A total of 145 soil samples were collected across four provinces and analysed using a Frontier FT-NIR spectrometer in the 750–2500 nm range. Spectral data were pre-processed using several methods, and partial least squares regression (PLSR) models were calibrated and validated against reference Kjeldahl measurements. Among all approaches, asymmetric least-squares (ALS) baseline correction coupled with PLSR achieved the best predictive performance, with $R^2 = 0.882$, RMSE = 0.0401% N, and RPD = 2.84 on the independent validation set. Characteristic wavelengths were identified around 1400, 1900, and 2200–2330 nm, corresponding to overtone and combination bands of N–H and O–H vibrations. The results confirm that NIRS offers a rapid, non-destructive, and cost-effective alternative for soil nitrogen assessment under diverse agroecosystems. This approach can support precision nutrient management in tropical smallholder farming systems and provides a foundation for developing portable, field-ready detection tools.

Keywords: near-infrared spectroscopy; soil nitrogen; wetland and dryland soils; multivariate analysis; Java Island, chemometrics

Nitrogen (N) is a key macronutrient essential for plant growth and soil fertility. It plays a crucial role in photosynthesis and the synthesis of amino acids, directly influencing crop yield and quality (Fathi, 2022). In tropical agricultural systems such as those in Indonesia, where soil fertility varies significantly across regions, accurate assessment of nitrogen content in soils is critical for effective nutrient management and sustainable farming practices (Susanti et al., 2024).

Conventional methods for determining soil nitrogen, such as the Kjeldahl or Dumas methods, are accurate but labour-intensive, time-consuming, and often require hazardous chemicals. These limitations hinder their widespread application, particularly in field-based, real-time monitoring scenarios. Consequently, there is a growing interest in developing rapid, cost-effective, and non-destructive alternatives for soil nitrogen analysis (Saha et al., 2023).

Near-infrared spectroscopy (NIRS) has emerged as a promising tool for soil analysis due to its speed, minimal sample preparation, and capacity for multi-constituent prediction (Beć et al., 2021). NIRS captures the molecular overtones and combination vibrations of chemical bonds (mainly C–H, O–H, and N–H), which can be indirectly related to nitrogen content (Bec et al., 2022). However, the spectral response is often influenced by factors such as soil moisture, texture, and organic matter, requiring robust multivariate calibration models to extract meaningful information (Seidel et al., 2022).

NIRS has shown promise in predicting soil nitrogen content across various soil types and climates, but developing a universal model remains challenging due to several factors. NIRS, combined with machine learning algorithms like random forest (RF), has demonstrated high prediction accuracy for soil nitrogen content, as evidenced by a correlation coefficient of 0.909 in a study using 143 soil samples (Zheng et al., 2022). However, the inherent heterogeneity of soils and environmental variations poses significant challenges to achieving a universal model with high precision and robustness (Huai et al., 2025). The Pair-Soil-Spectra approach, which contrasts different soil sample pairs, has improved the performance of NIRS-based models, particularly in small sample cases, suggesting that feature metrics can enhance model accuracy across diverse datasets (Wang et al., 2025).

In the context of Indonesia – particularly Java Island, the country's most densely populated and agriculturally intensive region – the ability to rapidly assess soil nitrogen levels in both wetland and dryland farming systems is of strategic importance (Pravitasari et al., 2021; Priatama et al., 2022). Here, wetland farming refers to paddy cultivation systems conducted in flooded or irrigated fields (Tirtalistyani et al., 2022). In contrast, dryland farming encompasses rainfed cropping systems such as dryland rice, maize, legumes, or tubers grown. Java's agro-ecological diversity, coupled with frequent rainfall and high humidity, leads to rapidly changing soil conditions (Rejekiningrum et al., 2022).

Farmers require rapid nitrogen assessment throughout the cropping cycle. However, conventional laboratory analyses, as commonly practiced, are time-consuming, which delays the availability of results and thereby hinders timely fertiliser recommendations. Moreover, these methods are impractical for frequent, field-based applications, particularly for smallholder farmers with limited resources (Grzebisz and Niewiadomska, 2024; Udvardi et al., 2021). The development of rapid, accessible soil nitrogen analysis techniques applicable to both wetland and dryland environments is essential to support sustainable nutrient management in the region.

Therefore, the development of a fast and reliable method to predict soil nitrogen content across different agricultural land types – namely wetland and dryland systems – can significantly support precision agriculture, optimise fertiliser use, reduce environmental impact, and ultimately enhance crop productivity in Java. The novelty of this study lies in its comparative evaluation of NIRS performance across both wetland and dryland soils in a tropical setting. This area has received limited attention in previous studies. By establishing robust multivariate calibration models tailored to the heterogeneous soil conditions of Java Island, this research provides new insights into the feasibility of NIRS for site-specific nutrient management in smallholder farming systems.

This study, therefore, aims to evaluate the performance of NIRS combined with multivariate analysis in predicting soil nitrogen content in samples collected from both wetland and dryland fields across various locations on Java Island, Indonesia.

Material and Methods

Soil sampling

Soil samples were collected from 145 locations across Java Island, Indonesia, covering four provinces: West Java (Bogor City, Subang Regency), Central Java (Solo City, Magelang

Regency, Banyumas Regency), Yogyakarta Special Region (Gunung Kidul Regency, Sleman Regency), and East Java (Batu City, Malang Regency) (Fig. 1). Within each location, representative agricultural fields – including both wetland and dryland agroecosystems – were purposively selected. A composite soil sample was then obtained by thoroughly mixing five sub-samples randomly collected from distinct points within the site.

Sampling was conducted using a hoe, with the aid of a crowbar or fork in harder soils. Before sampling, the soil surface was cleared of grasses and debris. A V-shaped pit was dug to a depth of approximately 30 cm, and soil slices of approximately 5 cm thickness were taken from the surface downward using a hoe. The five sub-samples from each site were homogenised to form a single composite sample.

Each composite sample (~2 kg) was cleaned of plant residues, roots, stones, and coarse organic materials, and then placed into a double zip-lock plastic bag labelled with the corresponding sampling location. All tools were cleaned between sampling events to avoid cross-contamination. The soil samples were subsequently stored for further processing, including conventional wet chemistry analysis and near-infrared spectral measurement.

Laboratory analysis

Total soil nitrogen was determined at the Laboratory of the Department of Soil Science and Land Resources (Lab. DITSL), IPB University, Dramaga, Indonesia, with accreditation number LP-1067 IDN. The analysis followed the standard Kjeldahl digestion–distillation–titration procedure, which has been widely used for determining soil nitrogen.

Approximately 1 g of air-dried, sieved soil (< 2 mm) was digested with concentrated sulfuric acid (H_2SO_4) in the presence of a mixed catalyst (K_2SO_4 – CuSO_4 –Se), which accelerates the conversion of organic nitrogen into ammonium sulphate ($(\text{NH}_4)_2\text{SO}_4$). After digestion, the solution was made alkaline with sodium hydroxide (NaOH), releasing ammonia (NH_3), which was subsequently

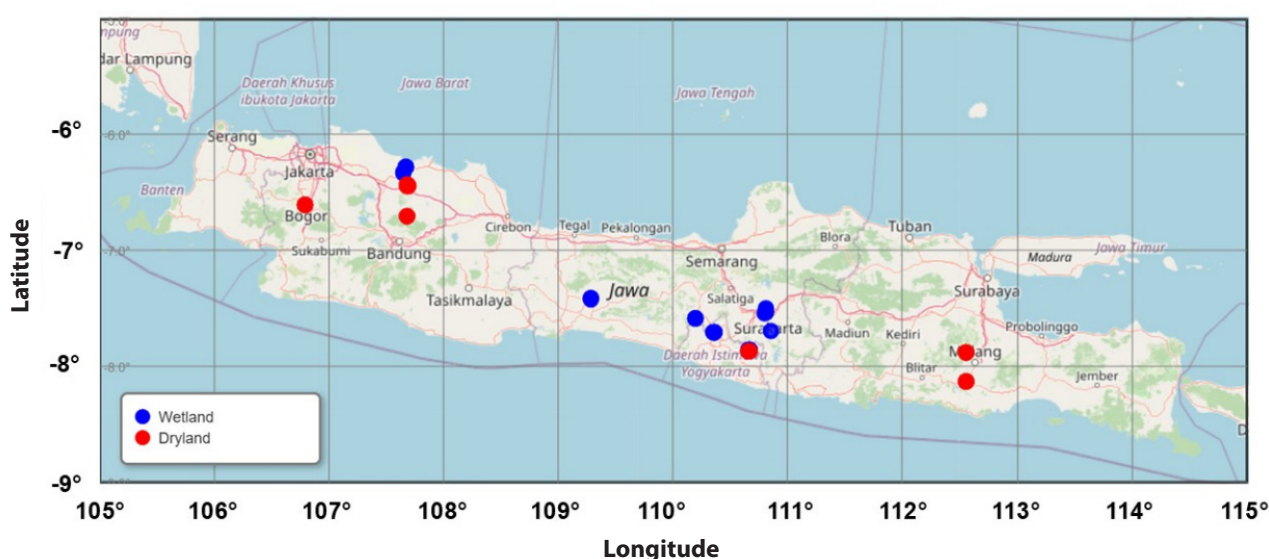


Fig. 1 Sampling sites in Java Island, Indonesia

distilled and absorbed into a boric acid solution. The captured NH_3 was then quantified by titration with standardised hydrochloric acid (HCl). The total nitrogen content was calculated from the volume of titrant consumed, expressed as a percentage of the soil's dry weight.

Spectral data acquisition

Spectral measurements were conducted using a Frontier FT-NIR spectrometer (PerkinElmer, USA) at the Integrated Laboratory of Bioproducts (iLaB), National Research and Innovation Agency (BRIN), Cibinong, Indonesia. Soil spectra were acquired in absorbance mode across the spectral range of 750–2500 nm, with a spectral resolution of 16 cm^{-1} and averaged over 32 scans to improve the signal-to-noise ratio. Data acquisition was conducted using the PerkinElmer Spectrum IR software (version 10.6.2.1159), and the resulting spectra were exported in .csv format for subsequent processing and analysis. The final spectral matrix consisted of 145 samples and 4665 wavelength variables (145×4665), which was then used for calibration and validation of multivariate models to predict soil nitrogen content.

Before measurement, soil samples were air-dried, finely ground, and sieved through a 60–80 mesh sieve to ensure uniform particle size. These preparations minimised the confounding influence of moisture and standardised sample presentation across sites. Approximately 4 mL of each prepared sample was placed in a glass vial (diameter 1.5 cm, height 4.5 cm). The vial was then inserted into the FT-NIR sample holder for spectral acquisition. To prevent cross-contamination, glass vials were thoroughly cleaned between successive samples.

Principal component analysis (PCA)

Raw absorbance spectra in the range of 750–2500 nm were examined to provide an initial overview of spectral variation between soil types. For each land type (dryland and wetland), the mean spectrum was calculated and visualised together with shaded bands representing ± 1 standard deviation (SD) from the mean. This approach allowed both the central spectral trend and within-group variability to be clearly displayed, facilitating comparison of spectral patterns before subsequent multivariate analysis.

Principal component analysis (PCA) was applied to examine the multivariate structure of the spectral dataset and to identify potential outliers before calibration modelling (Rodionova et al., 2021). Before computation, the spectral matrix (X) was mean-centred and scaled to unit variance to minimise differences in signal magnitude across wavelengths and to ensure that all variables contributed equally to the principal components. The distribution of samples was then evaluated using a score plot of the first two principal components (PC1 vs. PC2), overlaid with a 95% Hotelling's T^2 ellipse as a multivariate confidence limit. Samples positioned outside this boundary were considered potential outliers.

Spectral preprocessing

To minimise non-chemical sources of variation and enhance the robustness of calibration models, the raw NIR spectra were subjected to individual preprocessing techniques,

each applied in a single step rather than in combination. Each preprocessing method was evaluated and applied separately, ensuring that the resulting spectral datasets retained chemical interpretability while reducing specific physical or instrumental artifacts. For multiplicative scatter effects, standard normal variate (SNV) and multiplicative scatter correction (MSC) were considered. SNV normalises each spectrum relative to its mean and standard deviation, whereas MSC aligns each spectrum to a reference by removing additive and multiplicative distortions. Extended MSC (EMSC) provides an advanced alternative by modelling additional systematic effects, such as temperature- or matrix-induced variability (Roger et al., 2022).

For baseline and additive variations, polynomial detrending was employed to eliminate low-order drifts, while asymmetric least squares (ALS) correction was used as a robust approach to remove fluorescence-like backgrounds without altering true spectral features. For resolution enhancement and noise suppression, the Savitzky-Golay (SG) algorithm was applied, combining polynomial smoothing with derivative computation to improve peak separation and reduce high-frequency noise. Finally, mean centring and autoscaling were used independently to standardise variable scales and prevent the dominance of high-intensity regions during calibration (Roger et al., 2022).

Partial least squares regression

Partial least squares regression (PLSR) was employed to model the relationship between the NIR spectral predictors and total nitrogen concentration. Spectral data in the range of 750–2500 nm were arranged as the predictor matrix (X), while reference total nitrogen values served as the response variable (Y).

Models were fitted using the `pls` function in R with mean-centring applied to the predictor variables (Mevik and Wehrens, 2021). Calibration was performed on 70% of the dataset, selected by stratified sampling of the reference nitrogen values, while the remaining 30% was reserved for independent external validation. Model complexity was determined within the calibration set using 10-fold random-segment cross-validation, and the optimal number of latent variables was selected according to the one-standard-error rule, ensuring parsimonious solutions.

Model performance evaluation

The performance of the PLSR models was evaluated using both cross-validation predictions (from the calibration set) and external predictions (from the independent prediction set). Several statistical indices were calculated to quantify model accuracy, precision, and robustness (Hau et al., 2024):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y - \hat{y})^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (2)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}) \quad (3)$$

$$RPD = \frac{SD_y}{RMSE} \quad (4)$$

$$RPIQ = \frac{IQR_y}{RMSE} \quad (5)$$

where: n – number of samples; y – reference soil nitrogen content; $\hat{y}$ – predicted soil property; $\bar{y}$ – mean of measured values; RMSE – root mean square error; R^2 – coefficient of determination; Bias – mean difference between predicted and measured values (positive – underestimation, negative – overestimation); SD_y – standard deviation of reference values; IQR_y – interquartile range of reference values ($Q3 - Q1$); RPD – ratio of performance to deviation; RPIQ – ratio of performance to interquartile distance

A well-performing model is characterised by low RMSE values, an optimal number of latent variables (ncomp) that balances parsimony and performance, and high R^2 , RPD, and RPIQ values, while maintaining minimal Bias (Eq. 1–5).

The predictive performance of the models was interpreted based on widely adopted empirical guidelines in the spectroscopy literature. In general, RPD and RPIQ values below 1.5 are considered inadequate for practical application, values between 1.5 and 2.0 suggest limited utility for preliminary screening, and values of 2.0–2.5 indicate approximate or semi-quantitative prediction. Values in the range of 2.5–3.0 are typically regarded as sufficient for quantitative applications, whereas values exceeding 3.0 are often classified as excellent, reflecting highly reliable predictive ability.

Results and Discussion

Raw NIR spectral characteristics of dryland and wetland soils

The average NIR spectra of dryland and wetland soils revealed differences across the 750–2500 nm range, with dryland soils generally exhibiting higher absorbance values (Fig. 2). The shaded ± 1 standard deviation (SD) regions illustrate within-group variability, which remained relatively small across most of the spectrum but broadened around several key absorption regions. Importantly, although some spectral overlap between land types was observed, the mean spectra remained distinctly separated, highlighting systematic compositional differences.

Dashed vertical lines at 1400, 1900, and 2200 nm indicate regions of strong absorption, corresponding to regions commonly reported as diagnostic in soil NIR spectra (Fig. 2). These bands are associated with the first overtone of O–H stretching around 1400 nm, the combination of O–H stretching and H–O–H bending near 1900 nm, and organic or mineral-related functional groups such as CH_2 , CH_3 , NH_3 , Si–OH, and cation–OH around 2200 nm. The presence of these spectral features is consistent with previous findings that reported similar absorption valleys in soil reflectance spectra (Yu et al., 2023).

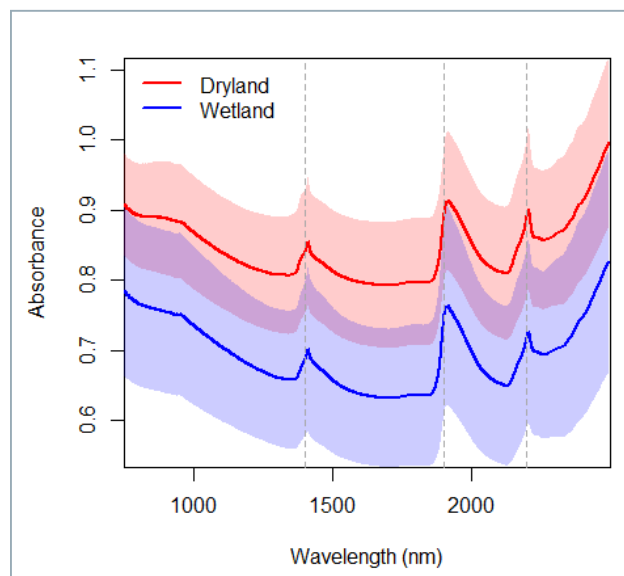


Fig. 2 Average near infrared spectra with ± 1 SD

As all samples were air-dried before spectral acquisition, these bands are unlikely to be dominated by free water absorption. Instead, the observed features are more plausibly attributable to molecular vibrations associated with organic matter and nitrogen-containing functional groups in soil constituents. The higher absorbance values in dryland soils may indicate a greater proportion of nitrogen relative to wetland soils.

Principal component analysis (PCA)

The eigenvalue decomposition showed that the first principal component (PC1) captured 97.95% of the total variance, while the second component (PC2) explained an additional 1.38% (Fig. 3). Together, PC1 and PC2 accounted for 99.33% of the spectral variance, indicating

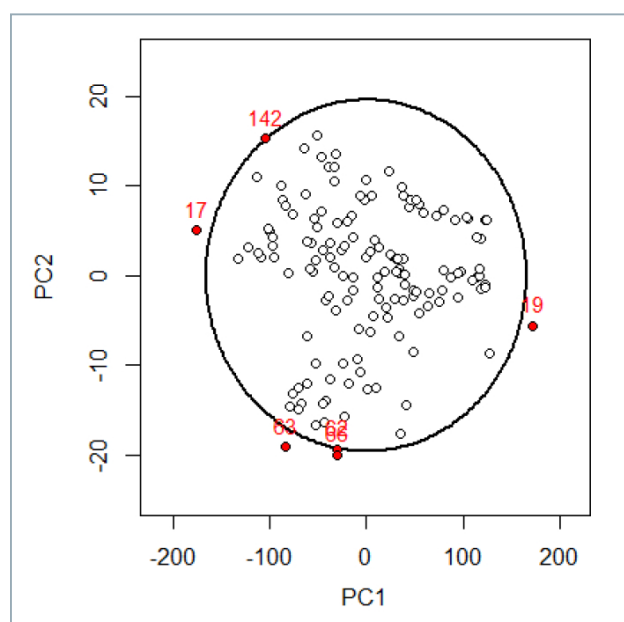


Fig. 3 Principal component analysis with 95% Hotelling's T2 ellipse

that the majority of systematic information in the dataset is concentrated in the first two components. Accordingly, the score plot based on PC1 and PC2 was used to represent the multivariate structure of the samples and to evaluate potential outliers using the 95% Hotelling's T^2 ellipse.

Several samples (No. 17, 19, 62, 63, 66, and 142) were flagged as potential outliers. However, their positions were only slightly outside the confidence boundary and still relatively close to the main cluster of observations. This suggests that the deviations are more likely to reflect natural variability in soil composition rather than instrumental or measurement errors. Considering that soil NIR spectra are inherently heterogeneous, excluding such samples without strong justification could result in an overly optimistic calibration model that underestimates real-world variability. Furthermore, outliers may contain important information for field prediction, as extreme cases are often the most critical to detect in practice. For these reasons, all samples were retained for subsequent preprocessing and calibration modelling to preserve the representativeness and robustness of the dataset.

Outliers were retained in the calibration dataset to preserve the representativeness of soil variability, as their exclusion may lead to overly optimistic estimation accuracies and reduce the generalisability of the models. This approach is consistent with recommendations from previous studies, which emphasised that removing spectral outliers can mask the true limitations of infrared evaluations and undermine the robustness of soil property prediction (Ludwig et al., 2016). In addition, it has been highlighted that the cause of spectral dissimilarity should be clearly identified before deciding on exclusion, since outliers may also reflect genuine sample heterogeneity rather than measurement error (Wadoux et al., 2021).

Preprocessing and PLSR

Applying the a priori selection rule (lowest external-validation RMSE as the primary criterion, with higher R^2 , fewer latent variables, smaller bias, and higher RPD/RPIQ as tie-breakers), baseline correction by asymmetric least squares (ALS) emerged as the best preprocessing (Table 1). On the independent prediction set, ALS achieved the lowest RMSE (0.0401), the highest R^2 (0.8816), and the highest RPD/RPIQ (2.84/2.29), with a model complexity of 10 latent variables. The method exhibited a relatively small negative bias (-0.0103), indicating that the model tended to slightly underestimate the reference values on average.

Two close competitors were EMSC and MSC. EMSC delivered a prediction RMSE of 0.0406 (+1.3% relative to ALS) with similarly strong R^2 and RPD, making it an attractive second choice – particularly when instrument/batch transfer is anticipated (Table 1). MSC yielded RMSE 0.0419 (+4.5% relative to ALS), high R^2 , and low bias (-0.0009) with a slightly more parsimonious model (9 components). While MSC is operationally simple and robust, ALS provided consistently better external accuracy in this dataset and was therefore selected. Overall, ALS achieved the best balance between external predictive accuracy and reasonable model complexity and was adopted as the final preprocessing method for soil-N modelling in this study.

Another single-step preprocessing performed less favourably (Table 1). SNV improved upon raw spectra but did not match ALS/EMSC/MS for prediction error. Detrending matched the raw performance in cross-validation yet underperformed on the external set, underscoring the need for selection to be based on independent validation rather than calibration/CV alone. Savitzky-Golay derivatives, especially the 2nd derivative, degraded accuracy – likely due

Table 1 Performance metrics of calibration and prediction models under different spectral preprocessing methods

No	Pre-processing	Set	Ncomp	RMSE	R^2	Bias	RPD	RPIQ	Quality RPD	Quality RPIQ
1	raw	calibration	15	0.0352	0.7914	0.0021	2.16	2.95	approx.	good
		prediction	15	0.0460	0.8329	-0.0035	2.47	1.99	approx.	screening
2	SNV	calibration	10	0.0421	0.7048	0.0004	1.81	2.47	screening	moderate
		prediction	10	0.0421	0.8635	-0.0006	2.70	2.18	good	moderate
3	MSC	calibration	9	0.0408	0.7194	-0.0001	1.86	2.55	screening	good
		prediction	9	0.0419	0.8669	-0.0009	2.71	2.19	good	moderate
4	Savitzky-Golay 1 st derivative	calibration	4	0.0496	0.5712	0.0013	1.53	2.10	screening	moderate
		prediction	4	0.0473	0.8282	0.0044	2.40	1.94	approx.	screening
5	Savitzky-Golay 2 nd derivative	calibration	1	0.0837	0.0019	0.0002	0.91	1.24	poor	poor
		prediction	1	0.1086	0.0645	0.0006	1.05	0.84	poor	poor
6	Detrending (baseline polynomial)	calibration	14	0.0353	0.7949	0.0027	2.15	2.95	approx.	good
		prediction	14	0.0452	0.8407	-0.0056	2.52	2.03	good	moderate
7	Baseline correction ALS	calibration	10	0.0363	0.7748	0.0010	2.09	2.87	approx.	good
		prediction	10	0.0401	0.8816	-0.0103	2.84	2.29	good	moderate
8	EMSC	calibration	10	0.0441	0.6992	0.0015	1.73	2.36	screening	moderate
		prediction	10	0.0406	0.8767	-0.0037	2.80	2.26	good	moderate

RMSE – root mean square error; R^2 – coefficient of determination; RPD – ratio of performance to deviation; RPIQ – ratio of performance to interquartile distance; Raw – unprocessed spectra; SNV – standard normal variate; MSC – multiplicative scatter correction; ALS – asymmetric least squares; EMSC – extended multiplicative scatter correction

to noise amplification in these dried-soil spectra – despite reducing the number of latent variables.

Spectral preprocessing in near-infrared spectroscopy (NIRS) is essential for improving data quality, reducing noise, and minimising interferences such as light scattering, baseline drift, and instrumental variation. Preprocessing enhances the reliability of calibration models by eliminating irrelevant spectral information that can obscure the chemical attributes of samples (Munawar et al., 2022). Common techniques include Savitzky-Golay smoothing, de-trending, baseline correction, standard normal variate (SNV), and multiplicative scatter correction (MSC), which standardise spectra and improve the signal-to-noise ratio. The application of MSC, when coupled with partial least squares regression (PLSR), has been shown to yield the highest predictive accuracy ($R^2 = 0.97$; RPD = 4.52) for chlorogenic acid determination in intact coffee beans (Munawar et al., 2024).

Prediction accuracy on the final preprocessing

External validation of the model using asymmetric least squares (ALS) baseline correction as the final preprocessing step showed strong predictive performance: RMSE = 0.0401 (absolute error in N, % units), $R^2 = 0.8816$ (88.16% of the variance explained), and a relatively small negative bias = -0.0103, indicating a slight overall under-prediction (Fig. 4). The scatter falls close to the 1 : 1 line, while the fitted regression line sits marginally below it, consistent with the negative bias. Visual inspection suggests errors are roughly uniform across the concentration range, with a mild tendency for larger under-prediction at the upper end.

Dryland samples (red dots) cover a wider and overall higher nitrogen range than wetland samples (blue dots), which are concentrated at the lower end (Fig. 4). The elevated absorbance values of dryland soils observed in Fig. 1 align with their comparatively higher nitrogen concentrations, highlighting the spectral-chemical relationship. Despite this shift in concentration domain, the point cloud for both groups remains well aligned with the 1 : 1 trend, suggesting that the ALS-pre-processed PLSR calibration generalises across land types without noticeable group-specific distortion.

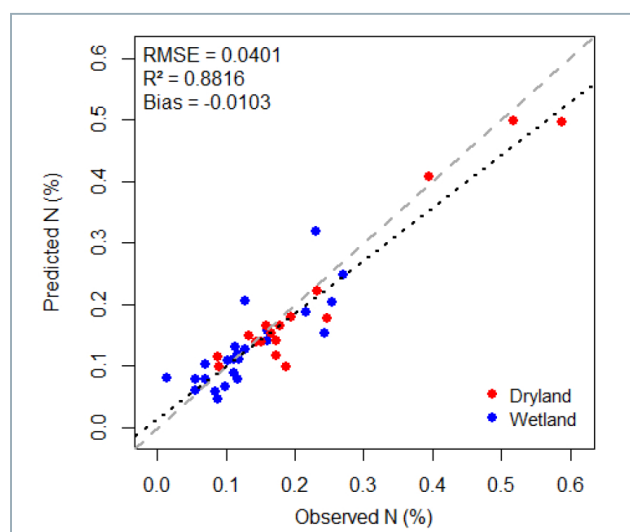


Fig. 4 Prediction performance of soil nitrogen under ALS preprocessing and PLSR

A study on soil nitrogen prediction using near-infrared spectroscopy analysed 143 soil samples collected from the Xunsi River Basin in Hubei Province, China. Among several preprocessing methods, loess smoothing was identified as the most effective. The optimised Random Forest regression model achieved strong performance, with $R^2C = 0.921$ and RMSEC = 0.115 in calibration, and $R^2P = 0.83$ with RMSEP = 0.141 in external validation, outperforming SVM and BP neural network models (Tan et al., 2022). Similarly, a large-scale investigation involving 7014 grassland canopy samples collected across multiple farms in New Zealand over eight years (2009–2016) demonstrated that spectral preprocessing with a Savitzky-Golay first derivative, combined with a 1D-CNN model, yielded the highest accuracy ($R^2 = 0.84$, nRMSE% = 11 in calibration) and maintained robust performance on independent datasets ($R^2 = 0.62$ – 0.75 ; nRMSE% = 13–14.5) (Pullanagari et al., 2021).

In the present study, soil samples were air-dried before spectral scanning. Removing moisture substantially improves the accuracy of near-infrared spectroscopy (NIRS) models for soil composition because water introduces strong and variable absorption features that can mask the actual chemical signatures of nutrients such as total nitrogen (TN) and soil organic carbon (SOC) (Munnaf and Mouazen, 2022; Tang et al., 2023). Several correction strategies – notably external parameter orthogonalisation (EPO) and orthogonal signal correction (OSC) – have been shown to mitigate moisture effects and thereby enhance prediction performance. For example, combining EPO with machine-learning models increased the coefficient of determination for SOC prediction to $R^2 = 0.76$ compared with $R^2 = 0.55$ for uncorrected models (Nawar et al., 2020).

In the present study, a benchtop FT-NIR spectrometer was employed rather than a portable device to maximise spectral sensitivity and stability and to minimise season-dependent environmental confounding. Seasonal shifts can induce distinct spectral feature spaces, necessitating domain-adaptation strategies to transfer models between seasons (Suarin and Chia, 2022). Studies using NeoSpectra spectrometers likewise indicate that, although portable units can estimate SOC and TN, predictive accuracy declines when training and testing data originate from different seasons or devices (Mitu et al., 2024). Portable NIRS, while valuable for real-time assessment, generally provides lower sensitivity and stability than benchtop instruments, with additional reports underscoring the need for careful calibration and validation under seasonal variability (Semella et al., 2022).

Wavelength assignment

After ALS baseline correction and PLS regression, the PLSR coefficient curve across wavelength can be read as an importance profile. Local positive maxima (peak) mark wavelengths where higher spectral intensity is associated with higher predicted total N, while local negative minima (trough) mark wavelengths where higher intensity relates to lower predicted N. Coefficient magnitude reflects the relative contribution of each wavelength.

In the present analysis, influential spectral features associated with soil nitrogen clustered at chemically

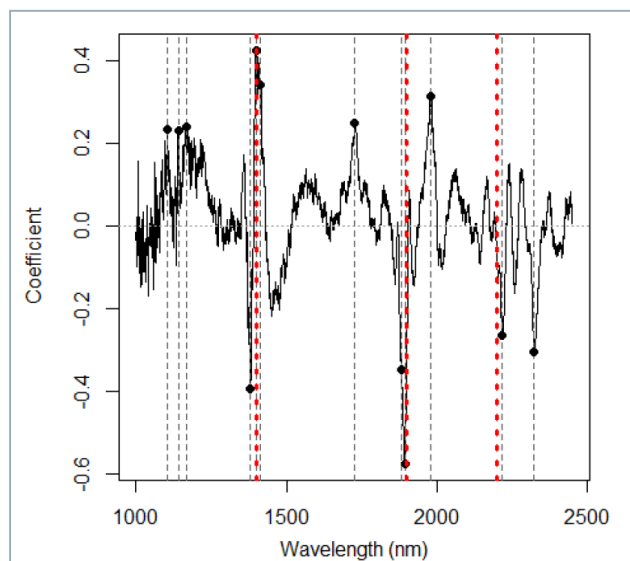


Fig. 5 Regression coefficients of ALS-PLSR model

meaningful bands (Fig. 5). A pronounced peak was observed around 1400 nm at 1402.1 nm, with adjacent support at 1414.5 nm and a flanking trough at 1381.1 nm. Near 1900 nm, the most significant negative feature occurred at 1893.0 nm, accompanied by a trough at 1881.4 nm. Additional troughs were detected in the 2200–2330 nm region (2217.4 and 2321.6 nm), while peaks also emerged at 1103.6, 1143.0, 1168.5, 1727.6, and 1979.6 nm.

Similar nitrogen-related absorptions have been widely reported in the near-infrared region, particularly around 1400–1500 nm, 1900–2100 nm, and 2300–2499.5 nm, with highly influential bands such as 1447, 1448.5, 2065, 2071, 2443, and 2444.5 nm identified as critical for soil total nitrogen prediction (Wang et al., 2021). For practical field applications, a reduced set of sensitive wavelengths at 1070, 1130, 1245, 1375, 1550, and 1680 nm was shown to provide a reliable estimation of soil nitrogen concentration through the development of an on-the-go detector (Zhou et al., 2021).

The differences between the current results and those reported in the literature may be attributed to variability in soil type, sample origin, and acquisition conditions. Large and diverse datasets, as used in global-scale studies, can produce more generalised but less site-specific characteristic bands (Wang et al., 2021), while instruments designed for operational feasibility emphasise narrower selections of wavelengths optimised for detector hardware and field stability (Zhou et al., 2021).

By comparison, the present study, based on localised soil samples and full-spectrum modelling, highlighted bands closely associated with N–H overtones and combination vibrations, particularly near 1400, 1900, and 2200–2330 nm. These differences emphasise that both dataset diversity and intended application – laboratory calibration versus field-ready sensing – play a decisive role in the identification of characteristic nitrogen-related wavelengths.

Conclusion

This study demonstrated that near-infrared spectroscopy, when combined with robust preprocessing and multivariate calibration, can effectively predict soil nitrogen content

across both wetland and dryland soils in Java. The ALS–PLSR model provided the strongest external validation results (RMSE = 0.0401% N, $R^2 = 0.8816$, RPD = 2.84), indicating good quantitative accuracy. Key absorption features were concentrated near 1400 nm, 1900 nm, and 2200–2330 nm, aligning with molecular vibrations of N–H and O–H groups. Importantly, the model generalised well across land types, with dryland soils exhibiting higher nitrogen ranges but following the same predictive trend as wetland soils. The findings highlight the feasibility of NIRS as a rapid, non-destructive method for soil nitrogen monitoring in tropical agroecosystems. However, the current work was limited to dried laboratory samples and a single instrument. Future research should extend validation to field-moist soils, multiple instruments, and cross-seasonal datasets, while also exploring advanced preprocessing and machine learning algorithms. Such efforts are essential to ensure robustness, transferability, and eventual deployment of portable NIRS-based tools for site-specific nutrient management in smallholder farming systems.

Acknowledgement

This work was funded by the Research and Innovation for Advanced Indonesia (RIIM) Programme, Batch 2, through the Indonesia Endowment Fund for Education Agency (LPDP) and the National Research and Innovation Agency (BRIN), under contract numbers B-1737/II.7.5/FR/11/2022 and B-14905/III.3/KS.00.00/11/2022. Additionally, the authors acknowledge the facilities and scientific and technical support from the Advanced Characterisation Laboratories Cibinong – Integrated Laboratory of Bioproduct, National Research and Innovation Agency, through E-Layanan Sains (ELSA).

References

- Bec, K.B., Grabska, J., Huck, C.W., 2022. In silico NIR spectroscopy – A review. *Molecular fingerprint, interpretation of calibration models, understanding of matrix effects and instrumental difference. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 279, 121438. <https://doi.org/10.1016/j.saa.2022.121438>
- Beć, K.B., Grabska, J., Huck, C.W., 2021. Principles and applications of miniaturized near-infrared (NIR) spectrometers. *Chemistry – A European Journal* 27(5), 1514–1532. <https://doi.org/10.1002/chem.202002838>
- Fathi, A., 2022. Role of nitrogen (N) in plant growth, photosynthesis pigments, and N use efficiency: A review. *Agrisost* 28, 1–8. <https://doi.org/10.5281/zenodo.7438164>
- Grzebisz, W., Niewiadomska, A., 2024. Nitrogen cycle in farming systems. *Agronomy* 14(1), 89. <https://doi.org/10.3390/agronomy14010089>
- Hau, N.-X., Tuan, N.-T., Trung, L.-Q., Chi, T.-T., 2024. Estimation of soil organic carbon content using visible and near-infrared spectroscopy in the Red River Delta, Vietnam. *Chemometrics and Intelligent Laboratory Systems* 255, 105253. <https://doi.org/10.1016/j.chemolab.2024.105253>
- Huai, S., Zhang, Q., Jin, Y., Yu, W., Meersmans, J., Wang, S., Colinet, G., Lu, C., 2025. Infrared spectroscopy for soil NPK estimation: Advances, challenges, and future directions in predictive modelling. *TrAC Trends in Analytical Chemistry* 118142. <https://doi.org/10.1016/j.trac.2025.118142>
- Ludwig, B., Linsler, D., Höper, H., Schmidt, H., Piepho, H.-P., Vohland, M., 2016. Pitfalls in the use of middle-infrared spectroscopy:

- representativeness and ranking criteria for the estimation of soil properties. *Geoderma* 268, 165–175. <https://doi.org/10.1016/j.geoderma.2016.01.010>
- Mevik, B.-H., Wehrens, R., 2021. Introduction to the pls Package. <https://cran.r-project.org/web/packages/pls/vignettes/pls-manual.pdf>
- Mitu, S.M., Smith, C., Sanderman, J., Ferguson, R.R., Shepherd, K., Ge, Y., 2024. Evaluating consistency across multiple NeoSpectra (compact Fourier transform near-infrared) spectrometers for estimating common soil properties. *Soil Science Society of America Journal* 88(4), 1324–1339. <https://doi.org/10.1002/saj2.20678>
- Munawar, A.A., Devianti, Satriyo, P., Bahari, S.A., 2022. Near infrared spectroscopy: Rapid and simultaneous approach to predict the fixed carbon, volatile matter and ash contents in biochar produced from agricultural residues. *Acta Technologica Agriculturae* 25(1), 1–6. <https://doi.org/10.2478/ata-2022-0001>
- Munawar, A.A., Kusumiyati, Andasuryani, Yusmanizar, Adrizal, 2024. Near infrared technology coupled with different spectra correction approaches for fast and non-destructive prediction of chlorogenic acid on intact coffee beans. *Acta Technologica Agriculturae* 27(1), 23–29. <https://doi.org/10.2478/ata-2024-0004>
- Munnaf, M.A., Mouazen, A.M., 2022. Removal of external influences from on-line vis-NIR spectra for predicting soil organic carbon using machine learning. *Catena* 211, 106015. <https://doi.org/10.1016/j.catena.2022.106015>
- Nawar, S., Munnaf, M.A., Mouazen, A.M., 2020. Machine learning based on-line prediction of soil organic carbon after removal of soil moisture effect. *Remote Sensing* 12(8), 1308. <https://doi.org/10.3390/RS12081308>
- Pravitasari, A.E., Rustiadi, E., Priatama, R.A., Murtadho, A., Kurnia, A.A., Mulya, S.P., Saizen, I., Widodo, C.E., Wulandari, S., 2021. Spatiotemporal distribution patterns and local driving factors of regional development in Java. *ISPRS International Journal of Geo-information* 10(12), 812. <https://doi.org/10.3390/ijgi10120812>
- Priatama, R.A., Rustiadi, E., Widiatmaka, W., Pravitasari, A.E., 2022. Physical geographical factors leading to the disparity of regional development: The case study of Java Island. *Indonesian Journal of Geography* 54(2), 195–205. <https://doi.org/10.22146/ijg.66729>
- Pullanagari, R.R., Dehghan-Shoar, M., Yule, I.J., Bhatia, N., 2021. Field spectroscopy of canopy nitrogen concentration in temperate grasslands using a convolutional neural network. *Remote Sensing of Environment* 257, 112353. <https://doi.org/10.1016/j.rse.2021.112353>
- Rejekiingrum, P., Apriyana, Y., Sutardi, Estiningtyas, W., Sosiawan, H., Susilawati, H.L., Hervani, A., Alifia, A.D., 2022. Optimising water management in drylands to increase crop productivity and anticipate climate change in Indonesia. *Sustainability* 14(18), 11672. <https://doi.org/10.3390/su141811672>
- Rodionova, O., Kucheryavskiy, S., Pomerantsev, A., 2021. Efficient tools for principal component analysis of complex data – A tutorial. *Chemometrics and Intelligent Laboratory Systems* 213, 104304. <https://doi.org/10.1016/j.chemolab.2021.104304>
- Roger, J.-M., Mallet, A., Marini, F., 2022. Preprocessing NIR spectra for aquaphotomics. *Molecules* 27(20), 6795. <https://doi.org/10.3390/molecules27206795>
- Saha, D., Senthilkumar, T., Singh, C.B., Manickavasagan, A., 2023. Application of near-infrared (NIR) hyperspectral imaging system for protein content prediction in chickpea flour. In Saini, M.K., Goel, N., Shekhawat, H.S., Mauri, J.L., Singh, D. (eds), *Agriculture-Centric Computation. ICA 2023. Communications in Computer and Information Science*, vol. 1866. Springer, Cham, Switzerland, pp. 141–153. https://doi.org/10.1007/978-3-031-43605-5_11
- Seidel, M., Vohland, M., Greenberg, I., Ludwig, B., Ortner, M., Thiele-Bruhn, S., Hutengs, C., 2022. Soil moisture effects on predictive VNIR and MIR modeling of soil organic carbon and clay content. *Geoderma* 427, 116103. <https://doi.org/10.1016/j.geoderma.2022.116103>
- Semella, S., Hutengs, C., Seidel, M., Ulrich, M., Schneider, B., Ortner, M., Thiele-Bruhn, S., Ludwig, B., Vohland, M., 2022. Accuracy and reproducibility of laboratory diffuse reflectance measurements with portable VNIR and MIR spectrometers for predictive soil organic carbon modeling. *Sensors* 22(7), 2749. <https://doi.org/10.3390/s22072749>
- Suarin, N.A.S., Chia, K.S., 2022. Transferring near infrared spectroscopic calibration model across different harvested seasons using joint distribution adaptation. In Wahab, N.A., Mohamed, Z. (eds), *Control, Instrumentation and Mechatronics: Theory and Practice. Lecture Notes in Electrical Engineering*, vol. 921. Springer, Singapore, pp. 707–716. https://doi.org/10.1007/978-981-19-3923-5_61
- Susanti, W.I., Cholidah, S.N., Agus, F., 2024. Agroecological nutrient management strategy for attaining sustainable rice self-sufficiency in Indonesia. *Sustainability* 16(2), 845. <https://doi.org/10.3390/su16020845>
- Tan, B., You, W., Tian, S., Xiao, T., Wang, M., Zheng, B., Luo, L., 2022. Soil nitrogen content detection based on near-infrared spectroscopy. *Sensors* 22(20), 8013. <https://doi.org/10.3390/s22208013>
- Tang, R., Jiang, K., Li, C., Li, X., Wu, J., 2023. Modeling to correct the effect of soil moisture for predicting soil total nitrogen by near-infrared spectroscopy. *Electronics* 12(6), 1271. <https://doi.org/10.3390/electronics12061271>
- Tirtalistyani, R., Murtiningrum, M., Kanwar, R.S., 2022. Indonesia rice irrigation system: Time for innovation. *Sustainability* 14(19), 12477. <https://doi.org/10.3390/su141912477>
- Udvardi, M., Below, F.E., Castellano, M.J., Eagle, A.J., Giller, K.E., Ladha, J.K., Liu, X., Maaz, T.M., Nova-Franco, B., Raghuram, N., Robertson, G.P., Roy, S., Saha, M., Schmidt, S., Tegeder, M., York, L.M., Peters, J.W., 2021. A research road map for responsible use of agricultural nitrogen. *Frontiers in Sustainable Food Systems* 5, 660155. <https://doi.org/10.3389/fsufs.2021.660155>
- Wadoux, A.M.J.C., Malone, B., Minasny, B., Fajardo, M., McBratney, A.B., 2021. Similarity between spectra and the detection of outliers. In *Soil Spectral Inference with R. Progress in Soil Science*. Springer, Cham, Switzerland, pp. 115–141. https://doi.org/10.1007/978-3-030-64896-1_7
- Wang, Y., Li, M., Ji, R., Wang, M., Zheng, L., 2021. A deep learning-based method for screening soil total nitrogen characteristic wavelengths. *Computers and Electronics in Agriculture* 187, 106228. <https://doi.org/10.1016/j.compag.2021.106228>
- Wang, Y., Zhao, C., Xing, Z., Zhu, M., Hao, L., Wang, K., Bai, J., Tian, H., Dong, D., 2025. Pair-soil-spectra: An approach for NIRS-based soil total nitrogen content detection with feature metrics in cases of small sample sizes. *Analytical Chemistry* 97, 454–463. <https://doi.org/10.1021/acs.analchem.4c04548>
- Yu, B., Yan, C., Yuan, J., Ding, N., Chen, Z., 2023. Prediction of soil properties based on characteristic wavelengths with optimal spectral resolution by using Vis-NIR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 293, 122452. <https://doi.org/10.1016/j.saa.2023.122452>
- Zheng, B., Xiao, T., Wang, M., Tian, S., Tan, B., 2022. Soil nitrogen detection based on random forest algorithm and near infrared spectroscopy. In *2022 International Conference on Computation, Big-Data and Engineering (ICCBDE)*, pp. 286–289. IEEE, Yunlin, Taiwan. <https://doi.org/10.1109/ICCBDE56101.2022.9888195>
- Zhou, P., Li, M., Yang, W., Yao, X., Liu, Z., Ji, R., 2021. Development and performance tests of an on-the-go detector of soil total nitrogen concentration based on near-infrared spectroscopy. *Precision Agriculture* 22, 1479–1500. <https://doi.org/10.1007/s11119-021-09792-0>