Principal component analysis of UAV-derived vegetation indices and laboratory tissue nutrients for crop health assessment

Oluibukun Gbenga Ajayi[1, 2] and Fatima Abiola Ogunlesi[3]

[1] Department of Land and Spatial Sciences, Namibia University of Science and Technology, Windhoek- oajayi@nust.na
[2] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa - oluibukun.ajayi@up.ac.za
[3] Department of Surveying and Geoinformatics, Federal University of Technology, Minna- ayodejitima@gmail.com

Keywords: Remote sensing, drone mapping, mixed-crop farming, field-scale crop assessment, nutrient mapping, precision agriculture

Abstract

Remote sensing and laboratory assays can improve field-scale crop assessment and management. This exploratory pilot study analyses relationships between leaf tissue nutrients and UAV-derived normalised difference vegetation index (NDVI) using seventeen paired samples collected across a mixed crop trial. Tissue measures for nitrogen, phosphorus and potassium were standardised and entered into principal component analysis to reduce pairwise correlation and extract orthogonal nutrient axes. The first principal component explained 54.79% of variance, the second explained 34.10%, together accounting for 88.9%. Principal component scores for the first two axes were used in linear and polynomial regression models to predict NDVI. Model skill was assessed on training data and with leave-one-out cross-validation, and bootstrap resampling produced empirical confidence intervals for component loadings. Linear models built on principal components provided the most stable cross-validated performance, while polynomial expansions improved training fit but generalised poorly. These findings indicate that a low-dimensional nutrient representation can predict NDVI with reasonable stability and that combining spectral and biochemical data supports spatially explicit nutrient assessment. The study recommends expanded and stratified sampling, reflectance calibration and targeted spectral bands for follow-up studies, and external validation before wider applications.

1. Introduction

Remote sensing has become an essential tool for field-scale crop monitoring because multispectral and multisource imagery deliver timely spatial information on canopy condition and vegetation vigour (Ahmad et al., 2022; Omia et al., 2023; Ajayi et al., 2024a). Vegetation indices, notably the normalised difference vegetation index (NDVI), remain widely used proxies for green biomass and canopy chlorophyll (Sarvakar & Thakkar, 2024). Improvements in sensor miniaturisation, flight control, and image processing now make small unmanned aerial vehicles (UAVs) a practical platform for plot- and field-scale workflows that include radiometric correction and orthomosaic generation for precise index extraction at the plot level (Guebsi, 2024, Ajayi et al., 2024b).

Linking spectral indices to laboratory measures of plant biochemistry increases the agronomic value of remote observations (Hatfield et al., 2008). A growing body of UAV research shows that narrowband indices and red-edge-based metrics often outperform broad-band NDVI for nitrogen retrieval because they target chlorophyll absorption features more directly (Walsh et al., 2018). Hyperspectral imaging and data fusion with machine learning can further improve trait estimation when adequate calibration data are available, but predictive skill remains conditional on sensor band selection, crop type, and growth stage (Li et al., 2024; Pranga et al., 2021). Phosphorus and potassium are more difficult to detect remotely because their spectral imprints are weaker, and successful retrieval often depends on indirect structural signals or combination predictors rather than on direct absorption features (Pimstein et al., 2011; Shen et al., 2021).

A persistent challenge in linking spectral and biochemical data is statistical redundancy among nutrient measures (Jain et al., 2024). Tissue nitrogen, phosphorus, and potassium often covary under shared soil type management and fertilisation regimes. Multicollinearity inflates parameter variance and complicates interpretation when raw nutrient variables are entered together into regression models. Principal component analysis (PCA) is a straightforward chemometric technique that transforms correlated variables into orthogonal components representing dominant modes of variation and is therefore commonly used to reduce predictor redundancy before regression or clustering analysis (Crista et al., 2024). Using component scores stabilises coefficient estimates in modest-sized calibration sets and yields interpretable predictors that reflect joint nutrient processes rather than single elements. Bootstrap resampling of the PCA solution provides empirical confidence intervals for component directions and helps quantify uncertainty in the derived axes, which is valuable when sample counts are limited.

Model selection and validation are central for operational mapping (Ajayi, et al., 2022). Flexible model forms, such as polynomial expansions or complex machine learning algorithms, can achieve strong calibration fit yet fail to generalise when the calibration set is small or narrowly sampled. Transparent reporting of both in-sample and out-of-sample performance, with cross-validation and bootstrap uncertainty estimates, has, therefore, become standard practice in precision agriculture studies (Hastie et al., 2009; Efron & Tibshirani, 1993). Recent applied UAV work reiterates this point and recommends simple models or penalised regression when sample size is limited because these approaches reduce overfitting and improve stability of predictions (Shukla et al., 2025).

https://doi.org/10.5194/isprs-annals-XI-3-2026-705-2026

Several recent studies report strong relationships between targeted spectral indices and leaf nitrogen across crops (Walsh et al., 2018; Pranga et al., 2021; Fuentes et al., 2025). Large-scale collated datasets and meta-analyses show that red-edge and red-edge NIR combinations often yield or produce higher R-squared values for leaf nitrogen than NDVI alone across multiple crop types (Fuentes et al., 2025). Work that integrates UAV multispectral data with texture and structure measures reports improved nitrogen and potassium mapping when structural variation is explicitly modelled, which is important because potassium influences turgor and canopy architecture rather than chlorophyll content directly (Zhang et al., 2025a; Zhang et al., 2025b). Studies focused on methodological robustness show that accounting for phenological stage shadow and soil background improves model transferability, and they recommend stratified sampling or explicit covariates for canopy architecture when the crop mixture is heterogeneous (Hernández-Ochoa et al., 2022; Khechba et al., 2024).

This paper builds on those findings and addresses two practical questions relevant to UAV-based nutrient mapping. First, can a low-dimensional representation of leaf tissue nutrients obtained with principal component analysis capture the dominant nutrient variation that relates to NDVI across a mixed crop experimental farm? Second, do component scores used as predictors produce regression models for NDVI that generalise more reliably than models based on raw nutrient variables or modest nonlinear expansions? To answer these questions, an exploratory pilot study that emphasises method development and feasibility rather than definitive performance claims was conducted. Collocated laboratory assays for nitrogen, phosphorus and potassium were paired with UAV-derived NDVI. Nutrient variables were standardised and PCA was applied, with bootstrap resampling used to quantify uncertainty in component loadings. The first two principal component scores served as predictors in linear and second-degree polynomial regression models for NDVI. Model skill was evaluated using training metrics and leave-one-out cross-validation.

For the purposes of this paper, crop health refers to canopy vigour and biochemical status that are commonly used in agronomic assessment. Operationally, we treat NDVI as a remote proxy for green biomass and canopy chlorophyll, while tissue nitrogen, phosphorus and potassium describe key biochemical dimensions of plant nutrient status. Thus, crop health in this study is the joint condition described by canopy spectral indices and leaf nutrient composition.

2. Materials and Methods

2.1 Study area

The study was conducted on an experimental farm with mixed cropping plots that included maize, groundnut, rice, yam, cassava and soybean located in Gidan Kwano, Minna, Nigeria. The sampling locations lie within a contiguous cultivation area where management practice and soil context are broadly consistent across plots. Seventeen paired field samples were collected during a single campaign, and each sample includes laboratory tissue tests and a collocated UAV-derived NDVI values. The spatial coordinates for each sample were recorded in metric units using a handheld GNSS receiver.

2.2 Data collection and acquisition

Field sampling followed standard tissue test protocols used Crop Production Laboratory of the Federal University of Technology, Minna, Nigeria. At each sample point, a leaf tissue sample was taken and processed for nitrogen, phosphorus, and potassium content using laboratory methods. Ground coordinates were logged for each sample to extract values from the UAV orthomosaic.

UAV imagery was acquired over the study area on the same day as field sampling to ensure the same meteorological conditions using DJI Mavic Pro UAV deployed at 30 m flight height, with 75% forwardlap and 65% sidelap (Ajayi & Onifade, 2023). Multispectral image frames covering the study area were processed with photogrammetric software to generate a georeferenced orthomosaic, from which NDVI maps were derived. Radiometric corrections and georeferencing were applied according to standard photogrammetric procedures. NDVI was computed from the orthomosaic using the standard band ratio formula, and the value for each sample location was obtained from the georeferenced NDVI layer. The final merged dataset, therefore, contains one NDVI value and three nutrient measurements per sample.

2.3 Data preprocessing

Nutrient variables were inspected for range and any transcription errors. No missing values were present in the analysis set. Pairwise covariance and Pearson correlation matrices were computed for the raw nutrient measures. Nutrient variables were standardised variable-wise to allow comparability and to prepare the matrix for multivariate analysis. Standardisation follows the z-score transformation as presented in equation (1)

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

where x_{ij} is the original measurement for sample i and nutrient j, $\bar{x}_j$ is the sample mean for nutrient j and s_j is the sample standard deviation for nutrient j.

2.4 Principal component analysis

Principal component analysis (PCA) was applied to the standardised nutrient matrix to obtain orthogonal axes that summarise joint variation in nitrogen, phosphorus and potassium. PCA was implemented using singular value decomposition, such that the standardised data matrix X is decomposed as

$$X = U\Sigma V^{\top} \quad (2)$$

The matrix V contains the loading vectors and the diagonal of Σ contains singular values related to the component variances. The eigenvalues λ_k are computed from the squared singular values and the proportion of variance explained for the component k is as presented in equation (3)

$$\text{ExplainedVar}_k = \frac{\lambda_k}{\sum_r \lambda_r} \quad (3)$$

Component scores for each sample were computed as

$$\text{PCscore}_{ik} = X_{i\cdot} \cdot v_k \quad (4)$$

where v_k is the loading vector for the component k. The first two component scores were retained for regression modelling

because they accounted for the majority of the variance in the nutrient matrix. Bootstrap confidence intervals for PCA loadings were computed using 1000 bootstrap replicates, resampling observation rows with replacement; the arbitrary random seed, which does not influence the substantive outcome, was set to 42 to ensure reproducibility. Sensitivity tests with 500 replicates gave consistent estimates. Bootstrap resampling with 1000 replicates was used to obtain empirical 95% confidence intervals for each loading coefficient, following standard practices for assessing sampling variability. The bootstrap percentile intervals were computed from the empirical distribution of bootstrap loadings. For a concise methodological reference on PCA and interpretation of loadings we refer to Jolliffe and Cadima (2016).

2.5 Regression modelling and validation

The principal component scores for PC1 and PC2 were used as predictors in a linear regression model of NDVI of the form

$$\mathrm{NDVI}_i = \beta_0 + \beta_1 \mathrm{PC1}_i + \beta_2 \mathrm{PC2}_i + \varepsilon_i \quad (5)$$

where $\varepsilon_i$is the error term. As a modest nonlinear alternative, we fitted a second-degree polynomial model, which includes squared terms for PC1 and PC2 and the interaction term PC1 times PC2. The explicit model is given in equation 6. Equivalent linear and polynomial regression models were fit using the raw nutrient variables for comparison.

Model performance was evaluated using three summary statistics. Root mean squared error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Error (MAE) are given in equations 6-8, respectively.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (6)$$

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (7)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | \hat{y}_i - y_i | \quad (8)$$

where $y_i$is the observed NDVI and $\hat{y}_i$is the model prediction.

To provide conservative estimates of predictive performance for this small sample, leave-one-out cross-validation (LOOCV) was applied. The LOOCV procedure iteratively leaves a single observation out of the training set, fits the model to the remaining $n-1$observations, predicts the held-out observation and accumulates prediction residuals across all folds. Cross-validated RMSE and MAE summarise generalisation error under this protocol. Model coefficients and fit statistics are reported for both the training fit and LOOCV.

All analyses were performed in Python using standard scientific libraries for linear algebra, statistics and plotting.

3. Results

The results are structured as follows. First, the NDVI image of the study area and the N-P-K values of the crop samples are presented, followed by the principal component analysis results, model performance, and diagnostic and agreement statistics.

3.1 NDVI and N-P-K values

Figure 1 presents the NDVI image of the study area, while the NDVI values obtained from the UAV orthomosaic and the N-P-K values from the tissue tests conducted for each of the crop samples are presented in Table 1. The Leaf nutrient concentrations are reported as percent dry weight, and the values reported by the laboratory are as unitless mass fractions (that is, fraction of one).

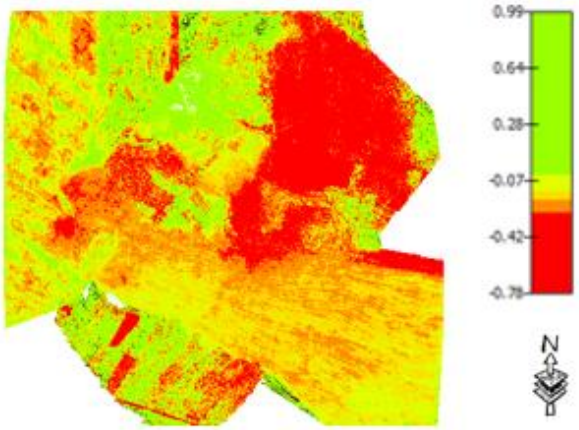

Figure 1. NDVI image of the study area.

NDVI typically ranges from -1 to 1, with positive values indicating live green vegetation. Figure 1 shows the full mosaic NDVI range used for visualisation; sample extraction used canopy mask pixels only, and extracted sample NDVI values ranged from 0.32 to 0.97, which indicates areas with little or no vegetation as well as patches of healthy vegetation.

Sample Description	Eastings (m)	Northings (m)	NDVI	N	P	K
AGP1-334(Maize)	217321	1047389	0.65	0.184	0.06	0.5
AGP1-334(Groundnut)	217321	1047389	0.65	0.181	0.06	0.7
AGP2-335(Maize)	217360	1047446	0.48	0.148	0.08	0.6
AGP2-335(Groundnut)	217360	1047446	0.96	0.146	0.02	0.9
AGP4-337(Yam)	217265	1047452	0.96	0.140	0.03	0.7
AGP6-339(Rice)	217171	1047618	0.32	0.158	0.08	0.3
AGP8-341(Rice)	217145	1047675	0.97	0.192	0.07	0.7
AGP9-342(Rice)	217243	1047529	0.35	0.129	0.02	0.5
AGP10-343(Yam)	217174	1047476	0.64	0.149	0.05	0.6
AGP11-344(Yam)	217171	1047494	0.76	0.187	0.08	0.6
AGP12-345(Cassava)	217151	1047360	0.34	0.175	0.02	0.5
AGP13-346(Soyabeans)	217108	1047394	0.67	0.218	0.11	0.6
AGP14-347(Soyabeans)	217126	1047422	0.70	0.123	0.03	0.7
AGP16-349(Soyabeans)	217232	1047388	0.79	0.177	0.03	0.6
AGP17-350(Cassava)	217205	1047394	0.72	0.180	0.05	0.8
AGP19-352(Groundnut)	217174	1047385	0.44	0.172	0.03	0.7
AGP11-344(Maize)	217171	1047494	0.93	0.133	0.04	0.5

Table 1. NDVI and corresponding N-P-K values of crops

3.2 Principal Component Analysis

All nutrient values were standardised prior to principal component analysis (PCA) and regression to remove scale effects. The PCA reduced the three nutrient variables to two dominant axes, with the following explained variances: PC1 = 54.79%, PC2 = 34.10%, PC3 = 11.10%. The combination of PC1 and PC2 explains 88.9%of the variance, so these two components capture the dominant nutrient variation in the dataset, with PC1 aligned with nitrogen and phosphorus and PC2 aligned with potassium. Tables 2 and 3 present the PCA loadings and explained variance with eigenvalues, respectively. Figures 2 and 3 present the Correlation matrix heatmap of tissue nutrients showing pairwise Pearson correlation coefficients, and the Covariance matrix heatmap of tissue nutrients showing pairwise covariance in original measurement units, respectively.

Figures 4 and 5 summarise the outcomes of the principal component analysis. Figure 4 presents the scree plot, showing that the first two components capture nearly 89% of the total variance and therefore provide an adequate low-dimensional representation of the nutrient data. Figure 5 displays the biplot of these two components, illustrating how sample scores and nutrient vectors are distributed. Nitrogen and phosphorus load strongly along the first axis, while potassium defines the second, confirming the orthogonal nutrient structure identified in the PCA. Note that PCA component signs are arbitrary and may flip under different implementations; that convention does not change component interpretation because a component and its sign-inverted score describe the same axis of variation.

Variable	PC1 Loading	PC1 95% CI Lower	PC1 95% CI Upper	PC2 Loading	PC2 95% CI Lower	PC2 95% CI Upper
N	0.6536	-0.7683	0.8013	-0.3937	-0.8521	0.8455
P	0.7129	-0.7023	0.7848	0.0335	-0.6799	0.7737
K	-0.2541	-0.9164	0.8781	-0.9186	-0.9702	0.9905

Table 2. Principal component loadings with 95% bootstrap confidence intervals.

PC	Eigenvalue	Explained variance percent	Cumulative percent
PC1	1.6436	54.7927	54.7927
PC2	1.0231	34.1036	88.8963
PC3	0.3333	11.1036	100

Table 3. PCA eigenvalues and explained variance.

Table 2 reports the component loadings with empirical 95% bootstrap confidence intervals. Key findings from the loading table are:

i. PC1 loads positively on nitrogen and phosphorus, with bootstrapped 95% intervals indicating the joint N P axis is robust under resampling. This suggests PC1 represents a combined N and P availability axis.
ii. PC2 loads strongly on potassium with a negative sign in the loading matrix, and the bootstrap interval is consistent with a distinct potassium axis.
iii. PC3 captures the remaining variance and is not used for predictive modelling.

Table 3 shows a clear break after the second principal component, with the first two components explaining about 89% of the total variance, which supports a low-dimensional representation of the nutrient data

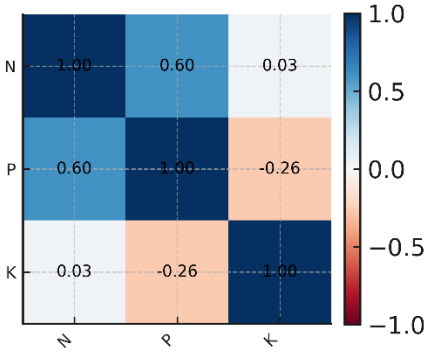

Figure 2. Correlation matrix heatmap of tissue nutrients.

The correlation heatmap shows a moderate positive correlation between nitrogen and phosphorus, r = 0.322, and weak negative correlations between nitrogen and potassium, r = -0.086, and between phosphorus and potassium, r = -0.057. The covariance heatmap (Figure 3) displays the raw covariation among nutrient measures; values are small because nutrient percentages are on a similar scale and because data were subsequently standardised for PCA.

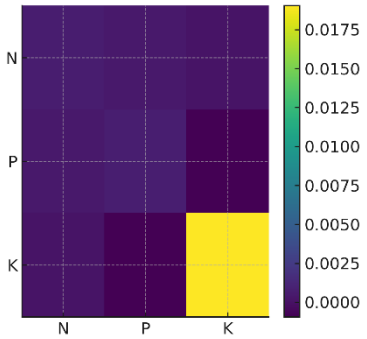

Figure 3. Covariance matrix heatmap of tissue nutrients.

The correlation and covariance diagnostics indicate that nitrogen and phosphorus have a moderate positive association, whereas potassium is largely independent of them. The N P positive correlation of about 0.32 supports the interpretation that a joint nitrogen-phosphorus axis will emerge from PCA. The weak or slightly negative correlations between potassium and the other nutrients explain why potassium projects strongly on a separate principal component; this pattern motivates retaining multiple components to represent distinct nutrient processes, and it clarifies why standardization prior to PCA is appropriate.

The biplot in Figure 4 visualises those two axes, with nitrogen and phosphorus loading strongly and positively on the first component, while potassium aligns with the second, indicating that N and P covary under similar soil or management conditions, while K represents an independent source of variation.

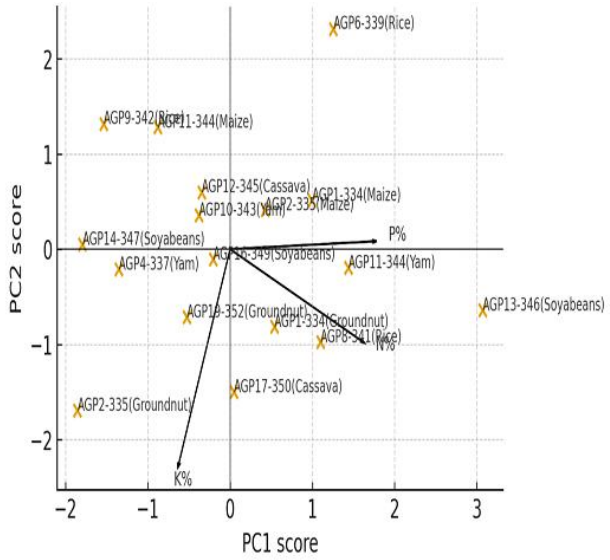

Figure 4. Biplot of PC1 and PC2 showing sample scores and nutrient loading vectors for N, P and K.

Sample scores cluster along the PC1 and PC2 axes, indicating distinct nutrient patterns across plots and crop types, and supporting the use of the first two component scores as compact, orthogonal predictors in the subsequent NDVI models.

3.3 Model Performance

A linear regression using PC1 and PC2 provides the most appropriate balance between goodness of fit and generalisation among the candidate models. The training RMSE of 0.1785 shows that the model reproduces the observed NDVI with moderate accuracy under calibration conditions, while the LOOCV RMSE of 0.2101 indicates only a modest reduction in predictive skill when evaluated out of sample. The small difference between these two error measures suggests that the model is relatively stable and not strongly affected by overfitting. Table 4 compares these metrics across alternative formulations and confirms that more flexible specifications do not yield improved cross-validated performance.

Model	Training RMSE	Training R2	Training MAE	LOOCV RMSE	LOOCV R2	LOOCV MAE
Linear PC1 PC2	0.1785	0.3018	0.1399	0.2101	0.0331	0.1661
Linear raw N P K	0.1717	0.3543	0.1440	0.2204	-0.0643	0.1870
Polynomial PC1 PC2	0.1770	0.3140	0.1406	0.3085	-1.0854	0.2525

Table 4. Model performance summary, training and leave-one-out cross-validation metrics

The coefficient estimates for the selected model are presented in Table 5. The intercept is highly significant and represents the expected NDVI at average nutrient conditions in the PCA space. PC1 has a small and statistically non-significant coefficient, indicating that the combined nitrogen-phosphorus gradient captured by this component does not exert a consistent linear influence on NDVI. In contrast, PC2 is statistically significant and negative, indicating that variation along the potassium-dominated axis is associated with a systematic decrease in NDVI. Because the component scores are standardised, this coefficient reflects the expected NDVI change for a one-standard-deviation shift in that nutrient pattern.

Term	Coefficient	StdErr	t	pvalue
const	0.6665	0.0477	13.9689	0.0000
PC1	-0.0211	0.0372	-0.5671	0.5796
PC2	-0.1129	0.0472	-2.3940	0.0312

Table 5. Coefficients for linear regression of NDVI on PC1 and PC2.

Figure 5 shows the relationship between observed and predicted NDVI for the linear PC model. Most points cluster around the one-to-one line, confirming that the model captures the overall trend in the data. However, the dispersion increases toward the lower and higher ends of the NDVI range, with a few observations departing from the diagonal. This spread is consistent with the reported RMSE values and indicates that the model is more suitable for describing general patterns than for precise point-level prediction.

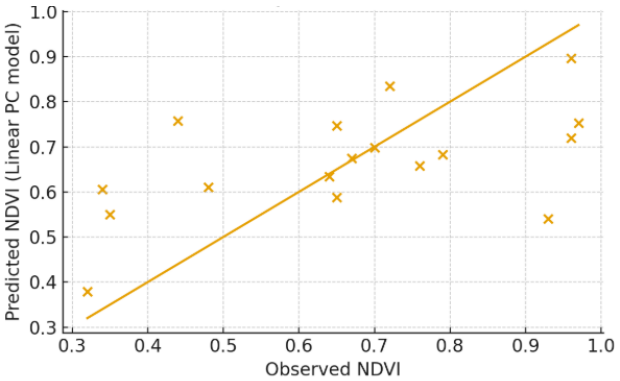

Figure 5. Observed NDVI versus NDVI predicted from the linear model using PC1 and PC2. The 1 to 1 line is shown for reference.

3.4 Diagnostics and agreement

Agreement and residual diagnostics show small mean bias but a few outlying samples with large residuals. Figure 6 presents a Bland-Altman plot that quantifies agreement between observed and predicted NDVI. The mean difference is close to zero, which indicates little systematic bias, while the 95% limits of agreement are wide enough to reflect moderate prediction uncertainty; observations that fall outside those limits should be inspected for sampling error, phenology mismatch, or mixed pixels.

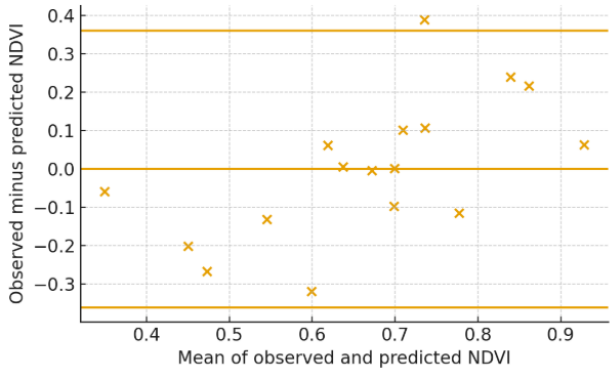

Figure 6. Bland Altman plot showing the difference between observed and predicted NDVI for the linear PC model, with mean difference and 95% limits of agreement.

4. Discussion

4.1. Principal component axes and spectral response

Principal component analysis reduced the three tissue nutrients to two orthogonal axes that capture most of the sample variation. PC1 explains 54.8% of the variance and groups nitrogen and phosphorus, while PC2 explains 34.1% and is dominated by potassium. Interpreting these axes in plant physiological terms, PC1 corresponds to biochemical processes that influence chlorophyll content and green biomass, which helps explain the observed sensitivity of NDVI to PC1 in the scatter and fitted plots (Figure 4). PC2 isolates potassium-related variation, which tends to affect leaf turgor and structure rather than chlorophyll concentration. The fitted linear model shows a measurable contribution from PC2 after accounting for PC1 (Table 3), which means potassium differences in this field produce detectable NDVI contrasts once joint nitrogen phosphorus effects are removed.

These axis interpretations align with recent field-level and UAV studies that report strong nitrogen signals in common vegetation indices, while other macronutrients may require different spectral bands or targeted sampling to retrieve reliably (Crista et al., 2024; Hamdane et al., 2023). Using the two-component scores as predictors, therefore, presents a compact and interpretable predictor set that removes collinearity and is convenient when calibration samples are limited.

4.3 Model comparison and generalisation behaviour

Model comparison shows the linear regression on PC1 and PC2 best balances fit and generalisation (Table 3). Polynomial expansions and models fit on raw N-P-K produce lower training error but substantially worse leave-one-out cross-validated error, which signals overfitting in this small, mixed crop sample. In practical terms, this means extra model flexibility captures idiosyncratic structure in the calibration data rather than relationships that hold under out-of-sample testing.

Statistically, the reduction in R-squared from training to LOOCV quantifies how much the calibration fit capitalises on sample-specific patterns. The decline in R-squared from training to LOOCV reduces confidence in precisely mapped values and motivates additional sampling before using predictions for site-level prescriptions. For screening and mapping where robustness of spatial patterns is primary, the linear PC model is the most defensible option.

4.4. Diagnostics and likely error sources

Agreement and residual diagnostics show a small mean bias overall, but a handful of outlying residuals that deserve attention. Possible causes or explanations include a mismatch in timing between tissue sampling and image capture, mixed pixels that combine crops and bare soil, and differences in canopy architecture across crops that affect the proportion of sunlit leaf area. These non-biochemical sources of variation can dominate spectral signals when sampling density is low, which is consistent with recent analyses of UAV reflectance errors (Feng et al., 2024).

Field verification is recommended for flagged samples to check labelling, canopy condition and image metadata. If particular crop groups or growth stages consistently produce larger residuals, stratified modelling or inclusion of crop and phenology indicators will likely reduce error.

4.5. Implications for operational mapping and sampling design

For operational use, it is important to treat the linear PC model as a screening tool that highlights candidate zones for targeted follow-up sampling. The LOOCV RMSE of around 0.21 indicates moderate uncertainty relative to the NDVI range observed in this study, so direct prescription without confirmatory tests would be imprudent. To improve transferability, three complementary actions are recommended. First, expand calibration sampling across crops, management regimes, and phenological stages so that models can capture broader variation. Secondly, enriching predictors with red edge or narrowband indices and canopy structural metrics often increases sensitivity to chlorophyll and reduces unexplained variance, which is important. Finally, evaluation of regularised and latent-variable methods, such as ridge regression and partial least squares, is required because they handle correlated predictors and small samples more robustly than unpenalised polynomial expansions (Lu et al., 2025).

5. Conclusion

This study applied principal component analysis to relate nutrient composition from tissue samples with NDVI derived from UAV imagery. The PCA reduced the three correlated macronutrients (nitrogen, phosphorus, and potassium) into two orthogonal axes that explained nearly 89% of total variation. The first component, dominated by nitrogen and phosphorus, reflected biochemical factors linked to chlorophyll and biomass, while the second component, associated with potassium, captured variation related to canopy structure.

Regression analysis using these components as predictors showed that NDVI variation can be partially explained by nutrient composition, with the linear model demonstrating the best trade-off between fit and generalisation under leave-one-out cross-validation. Although the model performance indicates moderate predictive capability, the results confirm that principal component scores provide a simplified and interpretable representation of nutrient effects on spectral response.

Residual analysis revealed some outliers likely caused by crop heterogeneity, small intra-day differences between tissue sampling and image capture, and canopy or soil background effects. These factors highlight the need for larger and more stratified calibration data and for improved synchronisation between field and UAV observations.

The findings demonstrate the potential of combining UAV remote sensing with nutrient component analysis for early detection of crop stress and for guiding precision agriculture practices. Future work should integrate additional spectral bands, such as red-edge indices, increase sample diversity across crop types and growth stages, and test regularised or latent-variable regression techniques to enhance model stability and predictive accuracy. Incorporating uncertainty visualisation in nutrient maps will further improve their interpretability for practical field applications.

5.1 Limitations and Future Work

The small number of paired observations limits the statistical analysis and the ability to generalise the findings beyond the sampled plots, and single samples can have a strong influence on model coefficients and predictive metrics. To quantify

uncertainty, we report influence diagnostics and bootstrap confidence intervals. Mixed crop types and differing phenological stages within the sample set may confound relationships between nutrient status and canopy reflectance, and future trials should adopt stratified sampling to separate crop and stage effects. Radiometric control was limited because no absolute reflectance panels were used during image acquisition, which means NDVI values should be treated as relative and can be sensitive to illumination and camera exposure; future campaigns should include reflectance calibration and co-located field spectroradiometer measurements. The consumer UAV used provides less spectral separation than dedicated multispectral sensors that include red edge bands, and targeted testing with sensors that capture near infrared and red edge information will help isolate nutrient-related spectral signals. Temporal mismatch between tissue sampling and image capture can introduce additional error, so synchronising sampling and imaging or recording precise timestamps for use as covariates is recommended. From a modelling perspective, the small sample size constrains flexible methods and increases overfitting risk; therefore, subsequent work should test regularised approaches such as ridge regression, partial least squares and principal component regression, and should report sensitivity to the number of bootstrap replicates.

References

Ahmad, U., Nasirahmadi, A., Hensel, O. and Marino, S., 2022. Technology and data fusion methods to enhance site-specific crop monitoring. *Agronomy*, *12*(3), p.555. https://doi.org/10.3390/agronomy12030555

Ajayi, O.G., Ibrahim, P.O., Adegboyega, S.O., 2024. Effect of hyperparameter tuning on the performance of Yolov8 for multi-crop classification on UAV images. *Applied Sciences, 14*(13):5708. https://doi.org/10.3390/app14135708

Ajayi, O.G., Iwendi, E., Adetunji, O.O., 2024. Optimizing crop classification in precision agriculture using AlexNet and UAV Imagery. *Technology in Agronomy. 4*: e011. https://doi.org/10.48130/tia-0024-0009

Ajayi, O.G., & Olufade, O.O., 2023. Drone-based crop type identification with convolutional neural networks: an evaluation of the performance of RESNET architectures. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci*., X-1/W1-2023, 991–998, https://doi.org/10.5194/isprs-annals-X-1-W1-2023-991-2023

Ajayi, O.G., Opaluwa, Y.D., Ashi, J. & Zikirullahi, W.M., 2022. Applicability of artificial neural network for automatic crop type classification on UAV-based images. *Environmental Technology and Science Journal*. 13(1), 57-72. https://doi.org/10.4314/etsj.v13i1.5

Crista, L., Todorovic, M., & Varga, B. 2024. Utilizing principal component analysis to assess the effects of foliar fertilization on maize quality. *Agriculture,* 14(8), 1428. https://doi.org/10.3390/agriculture14081428

Efron, B., & Tibshirani, R. J., 1993. An introduction to the bootstrap. Chapman & Hall/CRC.

Feng, G., Gu, Y., Wang, C., Zhou, Y., Huang, S., & Luo, B., 2024. Wheat Fusarium Head Blight Automatic Non-Destructive Detection Based on Multi-Scale Imaging: A Technical Perspective. *Plants*, *13*(13), 1722. https://doi.org/10.3390/plants13131722

Fuentes, I., Al-Shammari, D., Al-Nasrawi, A.K., Wang, Y., Wang, J., Lebrini, Y., Chen, Y., Jones, B.G. and Bishop, T.F., 2025. The normalised difference vegetation index as an analytic tool for wheat crop yield prediction: A review and meta-analysis. *Precision Agriculture*, *26*(4), p.55. https://doi.org/10.1007/s11119-025-10247-z

Guebsi, R. 2024. Drones in precision agriculture: A review of recent advances and operational applications. *Drones*, 8(11), 686. https://doi.org/10.3390/drones8110686

Hamdane, Y., Segarra, J., Buchaillot, M.L., Rezzouk, F.Z., Gracia-Romero, A., Vatter, T., Benfredj, N., Hameed, R.A., Gutiérrez, N.A., Torró Torró, I. and Araus, J.L., 2023. Using ground and UAV vegetation indexes for the selection of fungal-resistant bread wheat varieties. *Drones*, *7*(7), p.454. https://doi.org/10.3390/drones7070454

Hastie, T., Tibshirani, R., & Friedman, J. 2009. The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

Hatfield, J.L., Gitelson, A.A., Schepers, J.S. and Walthall, C.L., 2008. Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, *100*, pp.S-117. https://doi.org/10.2134/agronj2006.0370c

Hernández-Ochoa, I.M., Gaiser, T., Kersebaum, K.C., Webber, H., Seidel, S.J., Grahmann, K. and Ewert, F., 2022. Model-based design of crop diversification through new field arrangements in spatially heterogeneous landscapes. A review: Model-based design of crop diversification through new field arrangements in spatially heterogeneous landscapes. A review. *Agronomy for Sustainable Development*, *42*(4), p.74. https://doi.org/10.1007/s13593-022-00805-4

Jain, S., Sethia, D. and Tiwari, K.C., 2024. A critical systematic review on spectral-based soil nutrient prediction using machine learning. *Environmental Monitoring and Assessment*, *196*(8), p.699.

Jolliffe, I. T., & Cadima, J., 2016., Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A, *374*(2065), 20150202.

Khechba, K., Laamrani, A., Belgiu, M., Stein, A., Dong, Q. and Chehbouni, A., 2024. Design and Use of a Stratum-Based Yield Predictions to Address Challenges Associated with Spatial Heterogeneity and Sample Clustering in Agricultural Fields Using Remote Sensing Data. *Sustainability*, *16*(21), p.9196. https://doi.org/10.3390/su16219196

Li, X., Wu, J., Lu, S., Li, D., & Lu, D., 2024. Integration of Handheld and Airborne Lidar Data for *Dicranopteris Dichotoma* Biomass Estimation in a Subtropical Region of Fujian Province, China. *Remote Sensing*, *16*(12), 2088. https://doi.org/10.3390/rs16122088

Lu, F., Gao, W., & Chen, X., 2025. Data integration based on UAV multispectral and proximal sensors for crop trait retrieval. *Remote Sensing,* 17(8), 1411. https://doi.org/10.3390/rs17081411

Omia, E., Bae, H., Park, E., Kim, M.S., Baek, I., Kabenge, I. and Cho, B.K., 2023. Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sensing*, *15*(2), p.354. https://doi.org/10.3390/rs15020354

Pimstein, A., Karnieli, A., Bansal, S.K. and Bonfil, D.J., 2011. Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. *Field Crops Research*, *121*(1), pp.125-135. https://doi.org/10.1016/j.fcr.2010.12.001.

Pranga, J., Borra-Serrano, I., Aper, J., De Swaef, T., Ghesquiere, A., Quataert, P., Roldán-Ruiz, I., Janssens, I. A., Ruysschaert, G., & Lootens, P., 2021. Improving Accuracy of Herbage Yield Predictions in Perennial Ryegrass with UAV-Based Structural and Spectral Data Fusion and Machine Learning. *Remote Sensing*, *13*(17), 3459. https://doi.org/10.3390/rs13173459

Sarvakar, K. and Thakkar, M., 2024. Different vegetation indices measurement using computer vision. In *Applications of computer vision and drone technology in agriculture 4.0* (pp. 133-163). Singapore: Springer Nature Singapore.

Shen, J., Qiao, W., Chen, H., Zhou, J., & Liu, F., 2021. Application of Visible/Near Infrared Spectrometers to Quickly Detect the Nitrogen, Phosphorus, and Potassium Content of Chemical Fertilizers. *Applied Sciences*, *11*(11), 5103. https://doi.org/10.3390/app11115103

Shukla, P., Shukla, S., and Kumar Singh, A., 2025. Trajectory-Prediction Techniques for Unmanned Aerial Vehicles (UAVs): A Comprehensive Survey, in *IEEE Communications Surveys & Tutorials*, 27(3), 1867-1910, doi: 10.1109/COMST.2024.3471671.

Walsh, O.S., Shafian, S., Marshall, J.M., Jackson, C., McClintick-Chess, J.R., Blanscet, S.M., Swoboda, K., Thompson, C., Belmont, K.M. and Walsh, W.L., 2018. Assessment of UAV based vegetation indices for nitrogen concentration estimation in spring wheat. *Advances in Remote Sensing*, *7*(2), pp.71-90.

Zhang, W., Zhu, S., Zhong, Y., Li, H., Sun, A., Zhang, Y., & Zeng, J., 2025a. UAV Remote Sensing for Integrated Monitoring and Model Optimization of Citrus Leaf Water Content and Chlorophyll. *Agriculture*, *15*(21), 2197. https://doi.org/10.3390/agriculture15212197.

Zhang, H., Wang, G., Song, F., Wen, Z., Li, W., Tong, L. and Kang, S., 2025b. Improving chili pepper LAI prediction with TPE-2BVIs and UAV hyperspectral imagery. *Computers and Electronics in Agriculture*, *235*, p.110368. https://doi.org/10.1016/j.compag.2025.110368