Estimation of Corn Leaf Chlorophyll Index Using Hyperspectral Images and a Wavelet-Guided Spectral Index Optimization Framework

Smegnew Moges Mintesinot[1], Charles Cardoso Santana[2], Daniel Marçal de Queiroz[1] , Miquéias Henrique Pereira[1], André Luiz de Freitas Coelho[1]

[1]Department of Agricultural Engineering, Universidade Federal Viçosa (UFV), Viçosa, MG, 36570-900, Brazil

[2]Minas Gerais Agricultural Research Agency – Pitangui Institute of Agricultural Technology, Pitangui, MG, 35650-000, Brazil

Corresponding author: Smegnew Moges Mintesinot (smegnew.mintesinot@ufv.br)

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-15 July 2026

Porto Alegre, Brazil

Abstract.

The corn leaf chlorophyll index (LCI) is a proxy indicator of plant photosynthetic capacity and nitrogen nutritional status, thereby supporting precision nutrient management and data-driven agronomic decision-making. Conventional methods, such as handheld chlorophyll meters, provide accurate leaf-level measurements but are labor-intensive, time-consuming, and inherently limited in their ability to represent field-scale spatial variability due to their point-based sampling nature. Hyperspectral remote sensing provides a scalable, non-destructive approach to field-scale mapping of crop chlorophyll dynamics. This study proposes a wavelet-guided spectral index optimization framework to derive a biophysically interpretable and highly sensitive hyperspectral indicator of LCI from leaf-level hyperspectral reflectance spectra calibrated against chlorophyll meter measurements. Multi-scale wavelet decomposition using the Coiflet wavelet of order 1 (coif1) was applied at decomposition levels 3–5, representing intermediate-to-coarse spectral scales that enhance detection of chlorophyll spectral features. The analysis identified scale-consistent chlorophyll-sensitive domains across the blue–green absorption region (470–497 nm), green reflectance peak (504–620 nm), red absorption shoulder (657–690 nm), and red-edge–near-infrared transition (700–822 nm). Leveraging these biophysically meaningful spectral regions, we developed the Red-Edge Slope Deficit Index (RESDI = (RE_1 − RE) / (B + NIR)). RESDI exhibited the strongest predictive association with SPAD-derived chlorophyll measurements ($R^2 = 0.802$), outperforming conventional leaf chlorophyll indices including the Normalized Difference Red-Edge Index (NDRE; $R^2 = 0.732$), Chlorophyll Index Red Edge (CIre; $R^2 = 0.726$), and Modified Chlorophyll Absorption Ratio Index (MCARI; $R^2 = 0.518$). Across all regression models, RESDI consistently improved the accuracy of LCI estimation compared with conventional chlorophyl indices. Among the evaluated algorithms, partial least squares regression (PLSR) and simple linear regression (SMLR) achieved comparable and the highest predictive performances. In both models, RESDI yielded an $R^2 = 0.7995$ and RMSE = 2.6947 SPAD values, outperforming CIre ($R^2 = 0.7226$, RMSE = 3.1694), NDRE ($R^2 = 0.7288$, RMSE = 3.1339), and MCARI ($R^2 = 0.5086$, RMSE = 4.2185). Similarly, in the support vector machine regression (SVMR) model, RESDI achieved an $R^2 = 0.7733$ and RMSE = 2.8652, exceeding the performance of the other leaf chlorophyll indices. These findings demonstrate the superior predictive performance and stability of RESDI, emphasizing the critical role of red-edge slope deficit normalised by pigment absorption and canopy structural reflectance in governing LCI variability. The framework provides a scalable and transferable approach for hyperspectral monitoring of leaf chlorophyll status in corn production systems, with potential applicability to broader remote sensing platforms.

Keywords.
Proximal Sensing, Crop Monitoring, Vegetation Index, Nitrogen Stress, Precision Agriculture.

1. Introduction

Leaf chlorophyll content is a key physiological indicator for monitoring crop nitrogen status, guiding fertilization strategies, and forecasting yield potential in precision agriculture (Schlemmer *et al.*, 2013). In maize (Zea mays L.), nitrogen availability strongly governs photosynthetic capacity, biomass accumulation, and grain yield, and accurate in-season diagnosis of crop nitrogen status remains crucially important for precision nitrogen management (Dong *et al.*, 2024). Since chlorophyll molecules are nitrogen-containing compounds whose concentration declines predictably under nitrogen deficiency, leaf chlorophyll content serves as a reliable, biologically grounded proxy for plant nitrogen status(Schlemmer *et al.*, 2013). Rapid and non-destructive estimation of this trait is therefore central to site-specific nitrogen management aimed at improving nitrogen use efficiency while minimizing environmental losses.

Traditional chlorophyll quantification methods, including destructive solvent extraction and handheld optical meters such as the SPAD-502, offer accurate point-based leaf-level estimates but are labor-intensive, time-consuming, and inherently limited in spatial coverage(Sims and Gamon, 2002). Destructive sampling prevents repeated monitoring of individual leaves over time, and even non-destructive handheld meters require numerous individual readings to characterize within-field spatial variability. Even non-destructive handheld meters require a high volume of readings, often 30 or more per area of interest, to adequately characterize within-field spatial variability(Daughtry *et al.*, 1993). Furthermore, the SPAD-502 has been shown to lose sensitivity once chlorophyll content exceeds 300 mg m^{-2}, a level typical of healthy, green vegetation, which limits its effectiveness for the early detection of physiological stress(Gitelson and Schepers, 2009). These constraints make conventional methods poorly suited to capturing within-field variability in chlorophyll status, which is critical for variable-rate nitrogen application and other precision agriculture interventions(Daughtry *et al.*, 1993). As a result, there is a growing demand for scalable, non-destructive sensing approaches that provide spatially continuous, temporally responsive chlorophyll information across heterogeneous field conditions.

Hyperspectral remote sensing addresses this need by capturing reflectance across narrow, contiguous spectral bands sensitive to subtle biochemical variations. Leaf and canopy reflectance patterns are primarily shaped by chlorophyll absorption features in the blue and red regions, a green reflectance peak near 550 nm, and a steep red-edge transition (~680–750 nm) that bridges pigment absorption and near-infrared structural scattering(Sims and Gamon, 2002). The red-edge region is particularly valuable for chlorophyll estimation because it remains responsive across a substantially wider range of pigment concentrations than the red band, where reflectance saturates at moderate-to-high chlorophyll levels(Schlemmer *et al.*, 2013). Recent work emphasizes that the shape of the red-edge transition itself, rather than reflectance at any single wavelength, provides a robust indicator of canopy chlorophyll status: in nitrogen-replete maize, intense absorption at the red-edge onset combined with structural scattering in the near-infrared produces a steep rise in reflectance, whereas declining chlorophyll under nitrogen stress weakens this absorption and flattens the spectral profile(Schlemmer *et al.*, 2013). Continuous wavelet transform analysis is effective at extracting such subtle, multi-scale spectral features from corn canopy reflectance for chlorophyll estimation(Zhang *et al.*, 2020), supporting its use as a feature-extraction strategy before index development.

These spectral relationships have motivated the development of numerous chlorophyll indices, including the Normalized Difference Red Edge Index (NDRE), the Chlorophyll Index Red Edge (CIre)(Gitelson and Schepers, 2009), and the Modified Chlorophyll Absorption in Reflectance Index (MCARI)(Daughtry *et al.*, 1993). Although such indices are widely adopted for their computational simplicity and physiological interpretability, they are typically built on fixed band combinations derived empirically from limited datasets. Consequently, their predictive relationships often lack robustness and transferability, as they are frequently site-, species-, and time-specific(Shuai, Fowler and Basso, 2024). Their performance often degrades when applied across different growth stages, cultivars, or environmental conditions, limiting their utility for operational deployment in precision agriculture.

To improve generalizability, optimization-based approaches to spectral index development have gained attention as a means of identifying band combinations that maximize sensitivity to target biochemical traits, rather than relying on predefined formulations(Mulla, 2012). However, hyperspectral data are inherently high-dimensional and characterized by strong inter-band redundancy, meaning that exhaustive search over the full spectral space without prior dimensionality reduction can yield spurious correlations and models that generalize poorly (Fan *et al.*, 2019). Effective feature selection is therefore a prerequisite for robust, optimization-driven index development.

Wavelet transform analysis offers a principled solution by decomposing spectral signals into multi-scale components associated with distinct physiological phenomena, including pigment absorption and red-edge dynamics, while simultaneously suppressing noise and localizing stable, informative spectral regions (Blackburn and Ferwerda, 2008). This approach has been applied to corn canopy hyperspectral reflectance to extract subtle multi-scale features for chlorophyll estimation ((Zhang *et al.*, 2020). However, existing studies typically optimize indices from predefined band combinations or wavelet feature pools, rather than conducting an exhaustive combinatorial band-pair search across wavelet-identified sensitive spectral regions. To address this gap, the present study proposes a wavelet-guided spectral index optimization framework for estimating maize leaf chlorophyll content from hyperspectral reflectance data. Discrete multi-scale wavelet decomposition is used to identify chlorophyll-sensitive spectral regions, from which an exhaustive combinatorial band-pair search is conducted to develop a physiologically interpretable chlorophyl index, evaluated against established indices (NDVI, NDRE, CIre, GNDVI, and MCARI) across multiple regression frameworks.

2. Materials and Methods

2.1 Study Area

Field experiments were conducted during the 2022/2023 summer season at the Teaching, Research, and Extension Unit in Large-Scale Crop and Bioenergy Production (UEPE-GCBE Aeroporto), Department of Agronomy, Federal University of Viçosa (UFV), Viçosa, Minas Gerais, Brazil. The site is characterized by a tropical highland climate (Köppen "Cwa") with rainy summers and dry winters. The soil is a deep Yellow Latosol with a sandy clay loam texture. Average bulk density was 1.46 and 1.45 g cm^{-3} at the 0 to 20 and 20 to 40 cm layers, respectively, with field capacity ranging from 40.9 to 41.3% and permanent wilting point from 25.1 to 26.0%. 2.1

2.2 Experimental Design and Crop Management

Two experiments were conducted to develop hyperspectral indices for assessing maize nitrogen status through leaf chlorophyll measurements under different plant population densities and nitrogen rates. Three plant population densities were assessed: 50,000 (PP50), 60,000 (PP60), and 70,000 (PP70) plants ha^{-1}. Each experimental unit measured 8 m × 3.6 m, with rows spaced 0.6 m apart. The maize hybrid DEKALB 390 Pro 2 was cultivated under a no-tillage system on crop residues. Sowing was performed at a depth of 5 cm using a pneumatic seed drill. Basal fertilization consisted of 400 kg ha^{-1} of an 8:28:16 (N: P_2O_5: K_2O) formulation. Irrigation, pest management, and other cultural practices followed local recommendations for maize production. Data collection was performed 47 days after sowing, corresponding to the V7–V8 growth stage.

2.3 Chlorophyll Reference Measurements

Foliar chlorophyll status was assessed using a SPAD-502 chlorophyll meter (Konica Minolta, Japan). SPAD measurements were used as an indirect indicator of leaf chlorophyll content and associated nitrogen status. Readings were collected from the middle third of leaf C (the uppermost leaf with a visible collar) and/or leaf C–1 (the leaf immediately below). Three measurements were obtained per leaf at each sampling point and averaged to generate a representative SPAD value. The resulting served as ground-truth references for the development and evaluation of hyperspectral chlorophyll indices.

The measured ranged from 26.60 to 64.60, with a mean of 49.57 and a standard deviation of 6.43. The coefficient of variation (CV) was 12.98%, indicating moderate variability in chlorophyll status across the sampled plants. This range of values provided sufficient variability for calibration and validation of the spectral models.

2.4 Hyperspectral Data Acquisition and Preprocessing

2.4.1 Spectral Measurements

Leaf hyperspectral reflectance was measured using an ASD HandHeld 2 portable spectroradiometer (Analytical Spectral Devices, Inc., Boulder, CO, USA), covering the 325–1075 nm spectral range with a spectral resolution of ±1 nm. Measurements were acquired under clear-sky conditions between 10:30 and 14:30 local time to minimize solar zenith angle effects. The fiber-optic probe (25° field of view) was positioned vertically approximately 0.2 m above the leaf surface. Three spectra were collected per plant at consistent sampling positions and averaged to obtain a representative reflectance spectrum. Radiometric calibration was performed using a calibrated white reference panel before and after each measurement session. Subsequent analyses and vegetation-index calculations were restricted to the 400–1075 nm range to exclude edge bands with lower signal-to-noise ratios. Mean smoothed reflectance spectra are presented in Fig.1.

2.4.2 Spectral Smoothing

The raw spectral reflectance data were organized into a matrix $\mathbf{X}_{\mathrm{raw}} \in \mathrm{R}^{\mathrm{n}\times\mathrm{m}}$, where n and m denote the number of samples and wavelength bands, respectively. To reduce high-frequency noise while preserving relevant spectral features, a Savitzky–Golay filter was applied along the wavelength dimension(Chen *et al.*, 2004; Fan *et al.*, 2019). This method fits a polynomial of order p within a moving window of length w to preserve spectral moments that are typically lost in simple smoothing approaches. The selected parameters (p = 2) are consistent with optimized configurations reported for corn canopy analysis (Fan *et al.*, 2019) , while the window length (w = 11) was selected to provide an optimal trade-off between noise suppression and the preservation of narrow spectral features (Chen *et al.*, 2004). The smoothed reflectance at each wavelength position j is computed as shown in Equation (1).

$$X_{SSR} = S_G(X_{raw,} w, p) \tag{1}$$

where X_{SSR} is the Savitzky–Golay smoothed reflectance matrix, X_{raw} is the raw reflectance matrix, $S_G(\cdot)$ denotes the Savitzky–Golay filtering operation, w is the window length, and p is the polynomial order.

The resulting matrix $\mathbf{X}_{\mathrm{SSR}}$represents the smoothed spectral reflectance used in subsequent analyses.

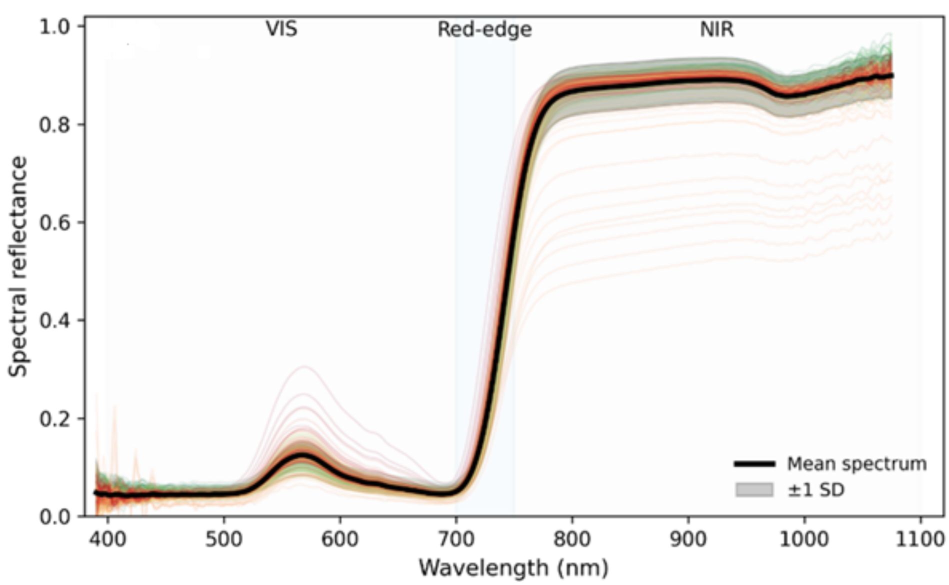

Fig. 1. Smoothed maize canopy reflectance spectra (400–1100 nm); the black line denotes the mean and the shaded area indicates ±1 SD. VIS, red-edge, and NIR regions are highlighted.

2.5 Discrete Multi-Scale Wavelet Sensitivity Analysis

To identify spectral regions most sensitive to chlorophyll variation, a discrete wavelet transform (DWT) was applied to the smoothed reflectance spectra. The Coiflet 1 (Coif1) wavelet was selected as the mother wavelet because of its compact support and ability to capture localized spectral features associated with vegetation reflectance(Blackburn and Ferwerda, 2008). Spectral decomposition was performed at three levels (L3–L5), which characterize broad absorption features and the influence of leaf structural properties. Approximation and detail coefficients from these decomposition levels were combined to provide a multiresolution analysis of spectral sensitivity.

Sensitivity was quantified using the coefficient of determination R^2 between the combined wavelet representation and SPAD measurements at each wavelength position. Spectral regions with $\mathrm{R}^2 > 0.40$ were identified as chlorophyll-sensitive and retained for subsequent vegetation-index development (Zhang *et al.*, 2020). All analyses were performed in Python using the PyWavelets library.

2.6 Spectral Band Definition and Index Construction

Following wavelet sensitivity analysis, the identified chlorophyll-sensitive spectral regions were aggregated into six operational narrow bands for index construction. Reflectance for each band was calculated as the arithmetic mean of spectral reflectance values within the corresponding wavelength interval. The red-edge region (700–751 nm), which exhibited the strongest and most consistent sensitivity to SPAD across multiple wavelet scales, was subdivided into three contiguous sub-bands: $\mathrm{RE_1}$ (700–717 nm), $\mathrm{RE_2}$ (718–734 nm), and $\mathrm{RE_3}$ (735–751 nm). This subdivision was designed to capture variations in the shape and gradient of the red-edge reflectance transition associated with chlorophyll status. The blue band (B) was defined using the chlorophyll-sensitive wavelength intervals identified by wavelet analysis (470–475 nm and 487–497 nm). Additional bands were defined within the green (G) and near-infrared (NIR), as summarized in Table 1.

Table 1: Definition of chlorophyll-sensitive narrow bands derived from discrete wavelet sensitivity analysis, with corresponding wavelength ranges, wavelet levels, and physiological interpretation.

Band	Range (nm)	Wavelet level	Physiological basis
B	470–500	L5 only	Chlorophyll a/b absorption
G	504–620	L4, L5	Carotenoid window
$\mathrm{RE_1}$	700–717	L3, L4, L5	Lower red-edge slope; chlorophyll onset
$\mathrm{RE_2}$	718–734	L3, L4, L5	Mid red-edge; maximum slope inflection
$\mathrm{RE_3}$	735–751	L3, L4, L5	Upper red-edge; chlorophyll saturation
NIR	762–809	L3, L4, L5	Structural scattering; mesophyll cell walls

The resulting band reflectance variables (B, G, R, RE, $\mathrm{RE_1}$, $\mathrm{RE_2}$, $\mathrm{RE_3}$, and NIR) were used as candidate variables for leaf chlorophyll index development.

2.6.1 Red-Edge Slope Deficit Index (RESDI)

The Red-Edge Slope Deficit Index (RESDI) was derived based on red-edge curvature behavior and R^2 performance, physical interpretability, and numerator–denominator independence. The index is defined by Equations (2) and (3).

$$\mathrm{RESDI} = \frac{\mathrm{RE_1} - \mathrm{RE}}{\mathrm{B} + \mathrm{NIR}} \quad (2)$$

which expands explicitly as:

$$\mathrm{RESDI} = \frac{(\mathrm{RE_1} - \mathrm{RE_2}) + (\mathrm{RE_1} - \mathrm{RE_3})}{\mathrm{B} + \mathrm{NIR}} \quad (3)$$

Where $\left[RE = \frac{RE_1 + RE_2 + RE_3}{3}\right]$.

The numerator quantifies the red-edge slope deficit, defined as the deviation of reflectance at the red-edge onset ($\mathrm{RE_1}$: 700–717 nm) from the mid ($\mathrm{RE_2}$) and upper ($\mathrm{RE_3}$) red-edge regions. This captures the progressive flattening of the red-edge slope associated with declining chlorophyll content(Sims and Gamon, 2002).In healthy vegetation, strong chlorophyll absorption in the red region suppresses reflectance at the red-edge onset, driving a steep rise toward the NIR plateau (Gitelson and Schepers, 2009). As chlorophyll declines, this absorption weakens more rapidly at the red-edge onset (nm) than at the NIR plateau, causing the profile to flatten (Sims and Gamon, 2002). The denominator combines blue reflectance (B), sensitive to chlorophyll absorption in the Soret band(Sims and Gamon, 2002), with NIR reflectance, which reflects canopy structural scattering from mesophyll cell walls and is largely independent of pigment concentration(Mulla, 2012). This formulation normalizes the red-edge slope deficit while preserving sensitivity to chlorophyll-driven spectral dynamics."

2.7 Exhaustive Spectral Index Search, Ranking, and Structural Analysis

An exhaustive combinatorial search was performed to identify the optimal normalized-ratio spectral index for SPAD estimation from the 8 spectral variables defined in Section 2.6. All candidate indices followed the general formulation given by Equation (4).

$$SI = \frac{X_i - X_j}{X_k + X_j} \tag{4}$$

where X_i and X_j represent numerator variables and X_k and X_l represent denominator variables, drawn from all pairwise combinations of the spectral feature set.
Numerator terms included both signed contrasts (e.g., RE_1 − B and B − RE_1), retaining both orientations to ensure exhaustive structural coverage without assuming a preferred spectral contrast direction. This resulted in 64 numerator combinations and 36 denominator combinations, yielding 2,304 candidate indices in total. For each candidate index, a simple linear regression model was fitted against reference SPAD measurements using the calibration dataset (n = 333), with R^2 used as the primary performance metric given its direct interpretability as explained variance.

The influence of numerator and denominator composition on predictive performance was evaluated by computing the mean and maximum R^2 for each spectral variable across all combinations in which it appeared. The contribution of NIR-containing denominators was assessed by comparing R^2 distributions between indices with and without NIR in the denominator. Band relevance was quantified using an enrichment ratio (ER), defined by Equation (5).

$$ER = \frac{f_{top}}{f_{bottom}} \tag{5}$$

where f_{top} and f_{bottom} denote the frequency of occurrence of each spectral variable within the top and bottom 20% of performing indices, respectively. ER > 1 indicates preferential association with high-performing formulations, whereas ER < 1 indicates association with lower-performing ones.

The search space structure was further characterized through R^2 distribution analysis, cumulative performance curves, rank–performance relationships, and numerator–denominator interaction heatmaps. The highest-performing structurally non-redundant index, defined here as an index not algebraically equivalent to or a sign-inverted form of a higher-ranked formulation, was designated as the Red-Edge Slope Deficit Index (RESDI), selected based on R^2 performance, physical interpretability, and numerator–denominator independence, as described in Section 2.6.1.

2.7 Reference Chlorophyl Indices

To benchmark the performance of the proposed index against established approaches, five widely used leaf chlorophyl indices were evaluated under identical conditions using the same calibration dataset (Table 2). The selected indices represent different spectral domains and mathematical formulations commonly applied for chlorophyll and nitrogen status estimation.

For all reference indices, band reflectance was derived from the same ASD hyperspectral dataset. When exact wavelength definitions were not directly available from discrete hyperspectral bands, the closest wavelet-informed spectral bands or nearest central wavelengths were used to ensure consistency with the original index formulations.

Table 2. Reference chlorophyl indices used for benchmarking.

Index	Formulation	Reference
NDVI	$(NIR - R)/(NIR + R)$	(Pettorelli *et al.*, 2005)
NDRE	$((NIR - RE)/(NIR + RE)$	(Gitelson and Merzlyak, 1994)
CIre	$(NIR/RE) - 1$	(Gitelson, Gritz and Merzlyak, 2003)
GNDVI	$(NIR - G)/(NIR + G)$	(Gitelson, Kaufman and Merzlyak, 1996)
MCARI	$[(RE - R) - 0.2(RE - G)] \times (RE/R)$	(Daughtry *et al.*, 2000)

2.8 Statistical Analysis and Model Evaluation

The predictive performance of all spectral indices was evaluated using three regression frameworks: simple linear regression (SLR), partial least squares regression (PLSR), and support vector machine regression (SVMR). Given that the indices are univariate predictors, SLR represents the standard least-squares solution; PLSR was included for consistency with the broader index evaluation framework, as it yields an approximately equivalent single-component solution in the univariate case. SVMR was included as a nonlinear benchmark and implemented using a radial basis function (RBF) kernel, with hyperparameters (C and γ) optimized via grid search within an inner cross-validation loop, nested within the outer evaluation framework to prevent data leakage.

Model performance was assessed using leave-one-out cross-validation (LOOCV), which provides an approximately unbiased estimate of generalization error in datasets of limited sample size(Arlot and Celisse, 2010), as in the present study ($n = 333$). The primary evaluation metrics were the cross-validated coefficient of determination (R^2) and root mean square error (RMSE), expressed in SPAD values. All analyses were implemented in Python (version 3.11) using the scikit-learn library for regression modeling, cross-validation, and hyperparameter optimization.

To facilitate interpretation of predictive performance, indices were classified into three accuracy tiers based on cross-validated R^2: high accuracy ($R^2 \geq 0.70$), moderate accuracy ($0.50 \leq R^2 < 0.70$), and low accuracy ($R^2 < 0.50$). These thresholds were adopted as operational benchmarks for comparative interpretation and are used to support the analysis presented in the Results section."

3. Result and Discussion

3.1. Identification of Chlorophyll-Sensitive Spectral Regions

Wavelet-based sensitivity analysis identified discrete wavelength intervals where canopy reflectance co-varied significantly with across decomposition levels L3–L5 (Fig. 2). Five intervals exceeded the $R^2 = 0.40$ threshold: 470–475 nm, 487–497 nm, 504–620 nm, 657–811 nm, and 819–822 nm. Together, these domains spanned 241 nm of the 400–1075 nm range (34.4%), substantially narrowing the spectral search space while retaining the most physiologically informative wavelengths for subsequent band definition and index construction (Section 3.2). A stricter criterion ($R^2 \geq 0.60$) reduced these to four robust clusters of the green reflectance region (530–597 nm), the red absorption shoulder (606–613 nm), the red-edge transition (659–751 nm), and the NIR structural region (762–809 nm).

Sensitivity responses were not uniform across decomposition levels. Blue-related intervals (max $R^2 \approx 0.55$) were detected exclusively at L5, indicating that SPAD-relevant information in this region is only resolvable at coarser spectral scales, consistent with prior reports that certain spectral features emerge only at coarser decomposition levels (Zhang *et al.*, 2020). The green reflectance region (max $R^2 \approx 0.75$) was present at L4 and L5 but absent at L3, suggesting that intermediate-to-coarse decomposition levels were required to capture this broad spectral feature. Among all identified domains, the red-edge transition exhibited the most stable multi-scale response, maintaining strong sensitivity across all three decomposition levels and reaching a maximum R^2 of approximately 0.75. This pattern is consistent with the well-established role of the red-edge as an indicator of chlorophyll concentration and spectral position shifts (Blackburn and Ferwerda, 2008). The persistence of this sensitivity from fine (L3) to coarse (L5) scales suggests that the Coif1 wavelet effectively captures both localized red-edge inflection shifts and their broader spectral context, consistent with the known non-linear relationship between red-edge position and chlorophyll content. The NIR region achieved the highest sensitivity overall (max $R^2 \approx 0.78$), with strong responses persisting across L3–L5 before declining beyond approximately 820 nm. This multi-scale persistence suggests a contribution from canopy structural characteristics that co-vary with chlorophyll status, consistent with leaf optical theory and radiative transfer models(Jacquemoud and Baret, 1990) .

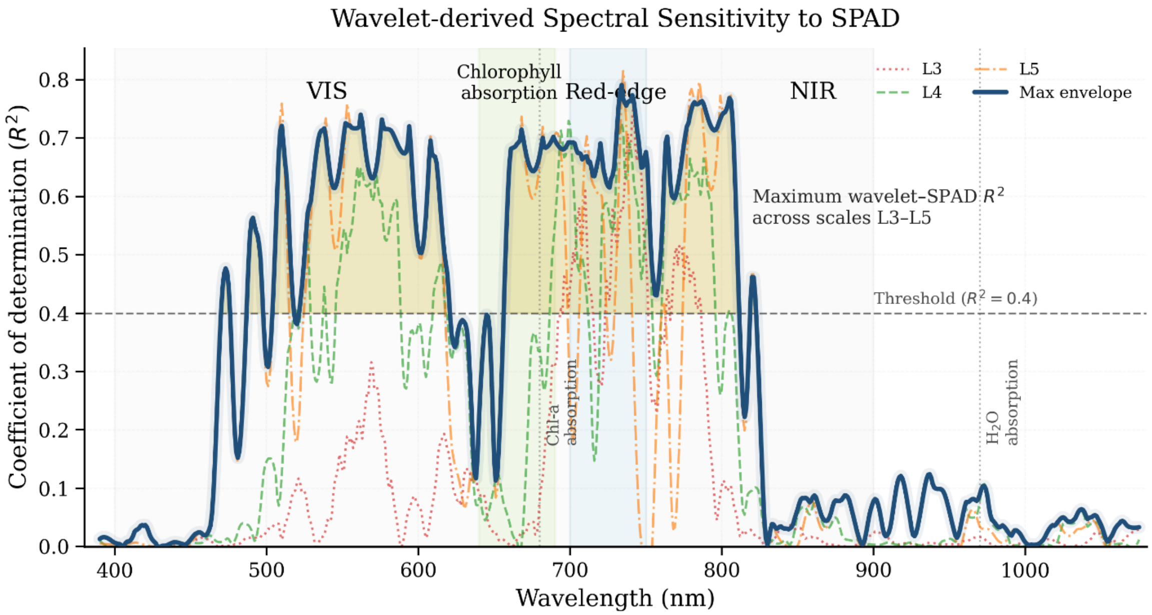

Fig. 2. Multi-scale wavelet sensitivity of hyperspectral reflectance to relative chlorophyll content (SPAD) across the visible (VIS), red-edge, and near-infrared (NIR) spectral domains. Wavelength intervals where continuous wavelet coefficients (CWCs) exceeded the selection threshold ($R^2 > 0.40$) at wavelet decomposition levels L3, L4, and L5 are highlighted.

3.2. Wavelet-Identified Spectral Bands and Mean Reflectance Profile

The mean canopy reflectance spectrum observed in this study displays the standard spectral signature of green vegetation, as shown in Fig. 3, characterized by distinct absorption and reflection features linked to biochemical and structural properties(Mulla, 2012). Low reflectance in the blue and red regions is driven by the strong absorption of photosynthetic pigments, primarily chlorophyll a and b, which absorb strongly across these spectral domains. A local reflectance maximum occurs in the green region (~550–570 nm), representing the green peak where pigment absorption is lower and sensitivity to higher chlorophyll concentrations is maintained(Zhang *et al.*, 2020).

The red-edge transition (700–751 nm) captures the critical shift from pigment-driven absorption in the visible spectrum to structurally driven scattering in the near-infrared. This region was further resolved into three contiguous sub-bands RE_1 (700–717 nm), RE_2 (718–734 nm), and RE_3 (735–751 nm). This finer resolution is expected to improve the ability to capture distinct aspects of red-edge reflectance dynamics, including absorption depth, slope, and inflection point position, each of which has been independently linked to chlorophyll status (Section 3.3). Following this rise, a high-reflectance plateau is reached in the NIR (762–809 nm).

The grey shading representing the ±1 SD envelope illustrates the variability in canopy reflectance among the measured samples. This envelope remained narrow across the visible spectrum due to the relatively uniform nature of pigment-driven absorption features. However, the envelope widened in the NIR plateau, reflecting greater inter-sample variability in canopy structural properties, such as leaf area index (LAI), biomass, and internal mesophyll cell wall scattering, which dominate reflectance in this domain(Mulla, 2012).

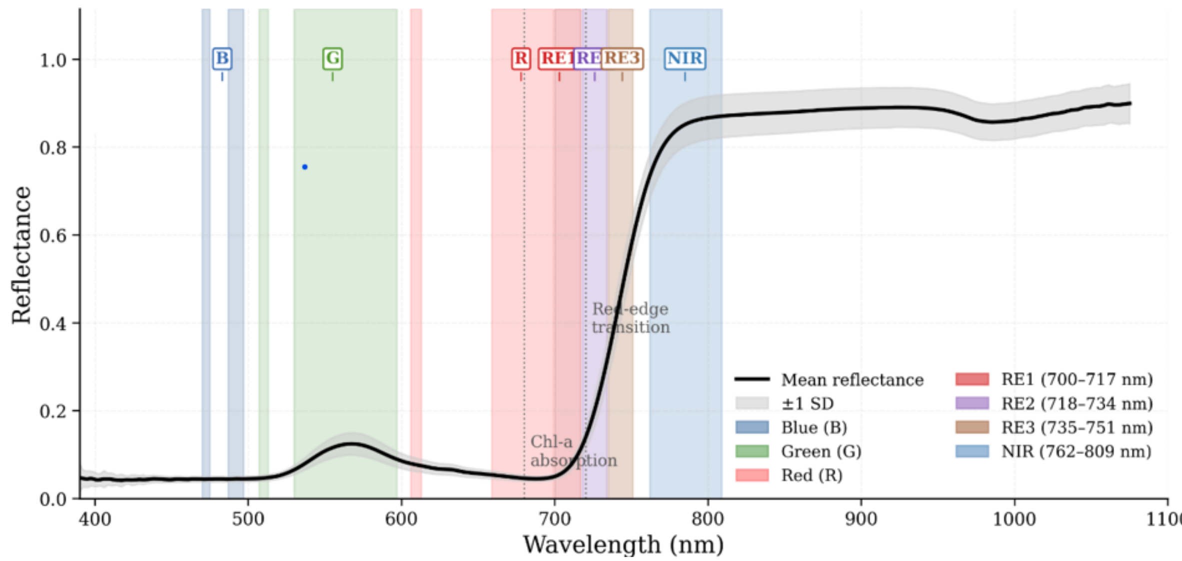

Fig. 3. Mean canopy reflectance spectrum (black line) with ±1 SD envelope (grey shading). Colored regions denote the wavelet-selected spectral bands used for index construction.

3.3 Development and Performance of the Red-Edge Slope Deficit Index (RESDI)

3.3.1 Exhaustive Index Search and Search Space Characterization

To identify the optimal normalized-ratio spectral index for LCI estimation, an exhaustive combinatorial search was performed across all wavelet-selected spectral bands. The search space comprised 8 spectral variables of visible bands (B, G, R), red-edge bands (RE, RE_1, RE_2, RE_3), and NIR, yielding 64 numerator terms and 36 denominator combinations, for a total of 2,304 normalized-ratio indices. Each index was evaluated against measured LCI values using Pearson correlation and ranked by coefficient of determination (R^2). The R^2 ranged from 0.00 to 0.80, with a median of 0.57 and an interquartile range (IQR) of 0.27–0.70. The best-performing index, (RE_1−RE)/(B+NIR), achieved an R^2 of 0.8016.

The performance distribution exhibited moderate negative skewness (skewness = −0.60; Fig. 4a), with R^2 values spanning from near zero to a maximum of 0.8016. The mean and median R^2 values were 0.484 and 0.574, respectively (Fig. 4b), indicating that most index formulations achieved moderate predictive performance, whereas a smaller subset of poorly performing indices generated a tail toward lower R^2 values. Only 74 indices (3.2%) achieved $R^2 \geq 0.79$, demonstrating that high-performing formulations were relatively rare despite the large search space. Conversely, 314 indices (13.6%) exhibited $R^2 < 0.10$, indicating little or no predictive utility.

Although 2,203 indices (95.6%) were statistically significant ($p < 0.05$), 504 of these still exhibited $R^2 < 0.30$. This finding highlights the distinction between statistical significance and practical predictive value, emphasizing that effect size should be prioritized over significance testing when selecting spectral indices for LCI estimation.

3.3.2 Numerator, Denominator, and Band Importance

Denominator composition exerted the strongest influence on index performance. NIR-containing denominators substantially outperformed non-NIR denominators, yielding a mean R^2 of 0.646 versus 0.438 (median 0.717 versus 0.487; $\Delta R^2 = 0.208$). Band enrichment analysis confirmed this pattern: NIR returned the highest enrichment ratio (≈4.3) among all spectral bands, followed by RE_2 (≈2.0) and RE_3 (≈1.1), indicating disproportionate occurrence within top-performing indices.

Among the eight NIR-containing denominator structures, $B + NIR$ranked last by mean R^2 (0.616, 8th of 8) yet produced the single highest-performing index overall (R^2 = 0.8016) when paired with the $RE_1 - RE$ numerator. This reflects the complementary physiological information contributed by chlorophyll-sensitive blue reflectance and canopy-structural NIR reflectance, whose combined effect emerges only under this specific numerator pairing rather than as a general property of $B + NIR$.

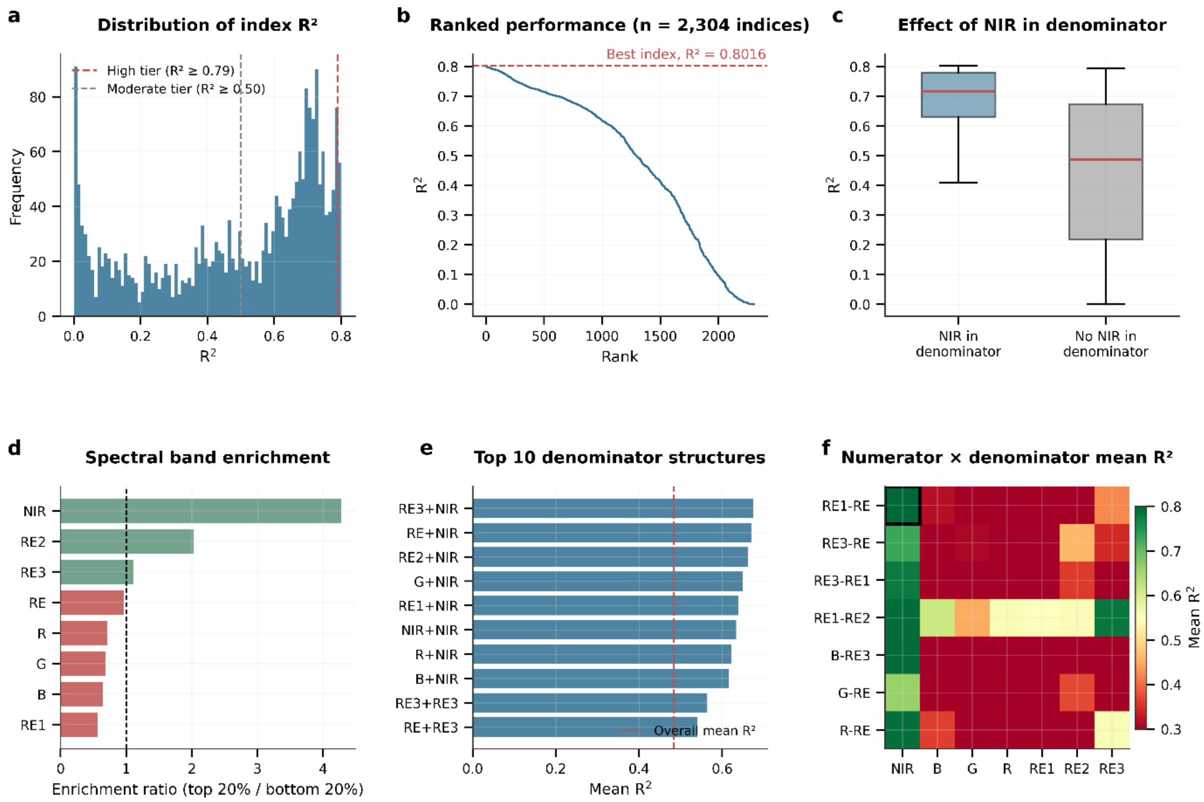

Fig. 4. Exhaustive search of normalized-ratio spectral indices for SPAD estimation (n = 2304). (a) R^2 distribution across all indices, with thresholds for moderate ($R^2 \geq 0.50$) and high ($R^2 \geq 0.79$) performance. (b) Indices ranked by R^2 (best: 0.8016). (c) Effect of NIR inclusion in the denominator on R^2. (d) Spectral band enrichment in top vs. bottom 20% of indices. (e) Top 10 denominator structures by mean R^2 across numerator pairings. (f) Mean R^2 for key numerator–denominator combinations.

Numerator analysis showed that spectral contrast, rather than absolute reflectance, was the primary source of predictive information. The lower red-edge asymmetry numerator $(RE_1 - RE)$ achieved the global maximum R^2 of 0.8016 when paired with $B + NIR$. The numerator–denominator interaction heatmap confirmed that optimal performance requires joint optimization of both components rather than independent selection: $RE_1 - RE$ performs poorly with most denominators except NIR-containing ones, where it reaches its maximum.

Several alternative numerator formulations approached the global maximum when paired with NIR-containing denominators:$B - RE_3$ (R^2 = 0.7983 with $B + NIR$), $RE_3 - RE_1$(R^2 = 0.7979 with $NIR + NIR$), and $RE_1 - RE_2$(R^2 = 0.7973 with $RE_2 + NIR$), each achieved R^2 within 0.004 of RESDI's maximum (Table 3). This convergence across multiple red-edge contrast formulations ($\Delta R^2 < 0.005$) indicates that the predictive information captured by RESDI is robust and not an artifact of a single specific band combination, reinforcing that the red-edge region as a whole, rather than any one narrow sub-band pairing, carries the dominant chlorophyll-related signal.

Table 3. Near-optimal normalized-ratio spectral indices for SPAD estimation, showing alternative red-edge contrast formulations that approached the global maximum R^2 (8-band search space, n = 2,304 total indices).

Rank	Index	Numerator family	R^2	r
1	$(RE_1 - RE)/(B + NIR)$	Lower red-edge asymmetry	0.8016	0.895
2	$((B - RE3)/(B + NIR)$	Blue–upper red-edge contrast	0.7983	0.893
3	$(RE_3\ RE_1)$/(NIR+NIR)	Full red-edge gradient	0.7979	0.893
4	$(RE_1 - RE_2)/(RE_2+NIR)$	Lower-to-mid red-edge slope	0.7973	0.893

3.3.3 RESDI Performance and Search Convergence

The Red-Edge Slope Deficit Index (RESDI), defined in Section 2.6.2, achieved the highest performance among all 2304 evaluated indices, with $R^2 = 0.8016$ ($r = 0.895$, $p = 2.62 \times 10^{-118}$). The physical basis of the numerator, interpreted as a cumulative red-edge slope deficit, is developed in Section 2.6.3. The calibration is given by Equation (6):

$$SPAD = 265.77\ RESDI + 101.70 \tag{6}$$

The regression exhibited a narrow 95% confidence interval and limited residual dispersion across the full SPAD range, indicating a robust and stable linear relationship (Fig. 4). For the $RE_1 - RE$ numerator family, denominator performance followed the hierarchy: B + NIR > NIR + NIR > RE_1 + NIR > R + NIR > RE + NIR > G + NIR, with B + NIR consistently producing the highest R^2, though the numerical advantage over the next-best denominator was modest ($\Delta R^2 = 0.002$–0.006). The exhaustive search converged to the same maximum R^2 as RESDI, confirming that no alternative formulation within the evaluated search space provided superior predictive performance and validating the wavelet-guided optimization framework as a reproducible and theoretically grounded approach for developing crop physiological indices.

3.4 Comparative Assessment of Leaf Chlorophyl Indices

To benchmark the performance of the proposed Red-Edge Slope Deficit Index (RESDI), six widely used chlorophyll- and nitrogen-sensitive vegetation indices were evaluated against measured SPAD values using identical calibration and validation procedures (Fig. 5). Predictive performance varied substantially across indices, reflecting fundamental differences in spectral sensitivity to canopy chlorophyll content.

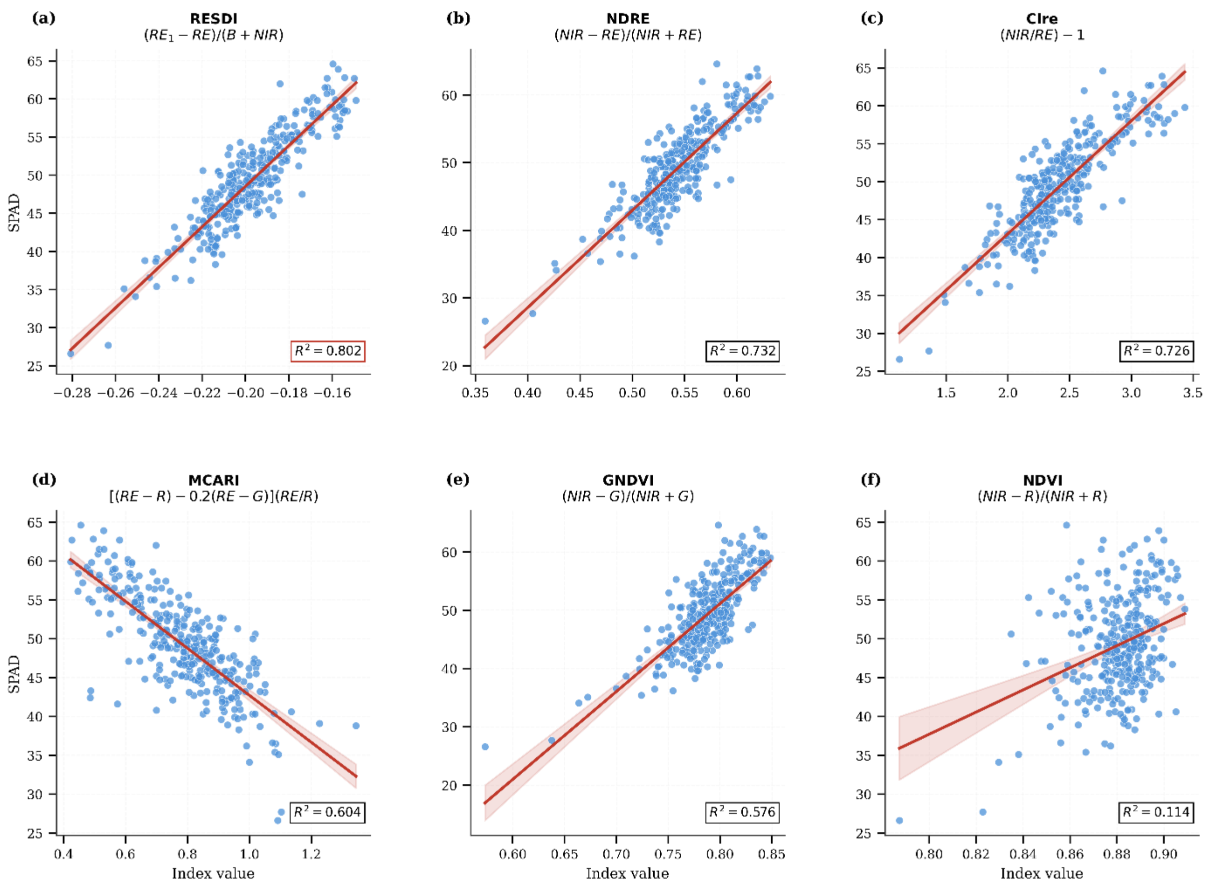

Fig. 5. Relationships between SPAD and six chlorophyll indices: (a) RESDI (b) NDRE, (c) CIre, (d) MCARI, (e) GNDVI, and (f) NDVI, with fitted regression lines and 95% confidence intervals.

RESDI achieved the highest predictive accuracy ($R^2 = 0.802$), exhibiting a linear relationship with SPAD and narrow 95% confidence intervals across the entire sample range. NDRE ($R^2 = 0.732$) and CIre ($R^2 = 0.726$) collectively form a group of indices whose shared reliance on the 700–750 nm red-edge region underlies their superior sensitivity. The success of this tier highlights the critical importance of the 700 to 750 nm region, which wavelet analysis independently identified as the most informative domain for chlorophyll assessment. This spectral domain was independently identified by wavelet analysis (Section 3.1) as the most informative for chlorophyll assessment, and its effectiveness is consistent with Gitelson *et* *al.* (2003), who established that wavelengths near 700 nm avoid the optical saturation characteristic of the main absorption bands.

GNDVI and MCARI showed intermediate performance, with R^2 values of 0.576 and 0.518, respectively. GNDVI exploits the green reflectance peak to extend sensitivity at higher pigment concentrations, but its exclusion of red-edge information limits responsiveness to subtle physiological variation (Mulla, 2012; Schlemmer *et al.*, 2013). MCARI, though designed to correct for soil background, showed increased dispersion at high SPAD levels, suggesting that its internal correction term may attenuate pigment-related signals in dense maize canopies (Daughtry *et al.*, 2000; Gitelson, Gritz and Merzlyak, 2003).

NDVI showed negligible predictive performance ($R^2 = 0.011$), consistent with its well-documented saturation behavior under moderate-to-high chlorophyll conditions, where red-band reflectance becomes increasingly insensitive to further increases in pigment concentration (Shuai, Fowler and Basso, 2024). This result confirms that traditional visible–NIR band combinations are inadequate for chlorophyll discrimination in maize at the growth stages sampled here.
Overall, predictive accuracy was strongly associated with the inclusion of red-edge spectral information. The performance advantage of RESDI over established red-edge indices such as NDRE and CIre further suggests that the wavelet-guided dual-normalization strategy employed in its construction captures additional spectral variance not exploited by existing formulations, while retaining a simple and physically interpretable mathematical structure.

3.5 Cross-Validated Model Performance and Robustness

The robustness of RESDI was evaluated using leave-one-out cross-validation (LOOCV; N = 333) across three regression frameworks: simple linear regression (SMLR), partial least squares regression (PLSR), and support vector machine regression (SVMR) (Table 4). RESDI achieved the highest cross-validated accuracy of all indices under every framework, attaining $R^2 = 0.7995$ and RMSE = 2.6947 SPAD values under both SMLR and PLSR.

SMLR and PLSR produced identical results for all indices, an expected outcome given that each model employed a single predictor. PLSR's latent-variable decomposition confers no advantage over ordinary least squares under these conditions. More substantively, SVMR with an RBF kernel degraded accuracy for the strongest red-edge indices: RESDI declined from $R^2 = 0.7995$ to 0.7733, with analogous reductions for NDRE and CIre. This pattern is consistent with Gitelson et al. (2003) and Schlemmer et al. (2013), who demonstrated that reflectance near the 700 nm red-edge inflection maintains a near-linear relationship with chlorophyll content by avoiding the optical saturation of the primary red absorption band. The linearity of this relationship means that nonlinear kernel complexity introduces overfitting without improving generalization.

An exception was observed for MCARI, where SVMR improved predictive performance ($R^2 = 0.6446$) compared with linear approaches ($R^2 = 0.5086$). This indicates the presence of nonlinear behavior in MCARI's relationship with chlorophyll that is not captured by linear models. Nevertheless, even under nonlinear modeling, MCARI remained substantially less accurate than RESDI. Collectively, these findings demonstrate that the superior performance of RESDI is primarily attributable to its physically meaningful spectral formulation rather than model complexity. The consistency of its performance across linear and nonlinear frameworks further confirms its robustness and suitability for reliable chlorophyll estimation in maize.

The consistency of RESDI across frameworks reflects its physically motivated design rather than statistical convenience. Its use of the 700–750 nm red-edge region targets the spectral domain of highest chlorophyll sensitivity, while the dual-normalization denominator (B + NIR) incorporates established principles for attenuating leaf surface reflectance and canopy structural scattering, effectively decoupling biochemical signals from architectural noise (Sims and Gamon, 2002). Together, these properties position RESDI as a robust and transferable index for non-destructive chlorophyll estimation and precision nitrogen management across a broad range of maize physiological conditions.

Table 4. Cross-validated performance of chlorophyl indices under SMLR, PLSR, and SVMR.

Index	SMLR R^2	SMLR RMSE	PLSR R^2	PLSR RMSE	SVMR R^2	SVMR RMSE
RESDI	**0.7995**	**2.6947**	**0.7995**	**2.6947**	**0.7733**	**2.8652**
NDRE	0.7288	3.1339	0.7288	3.1339	0.6927	3.3357
CIre	0.7226	3.1694	0.7226	3.1694	0.7033	3.2778
MCARI	0.5086	4.2185	0.5086	4.2185	0.6446	3.5875
GNDVI	0.5682	3.9543	0.5682	3.9543	0.5415	4.0748
NDVI	−0.0095	6.0463	−0.0095	6.0463	0.414	4.6065

4. Conclusion

This study developed a wavelet-guided framework for deriving a biophysically interpretable spectral index for non-destructive estimation of corn leaf chlorophyll index (LCI) from hyperspectral reflectance data. Multi-scale Coiflet wavelet decomposition (coif1, levels 3–5) identified four chlorophyll-sensitive spectral domains: the blue–green absorption region (470–497 nm), green reflectance peak (504–620 nm), red absorption shoulder (657–690 nm), and red-edge–near-infrared transition (700–822 nm). Together, these domains spanned 241 nm (34.4%) of the measured spectral range. This substantially narrowed the search space while retaining physiologically informative wavelengths for band definition and index construction.

An exhaustive evaluation of 2,304 normalized-ratio formulations showed that predictive performance depended primarily on denominator composition. NIR-containing denominators outperformed non-NIR alternatives by a mean R^2 of 0.208. The Red-Edge Slope Deficit Index, RESDI = $(RE_1 - RE)/(B + NIR)$, achieved the highest fit with SPAD-derived chlorophyll measurements ($R^2 = 0.8016$), exceeding NDRE ($R^2 = 0.732$), CIre ($R^2 = 0.726$), MCARI ($R^2 = 0.518$), GNDVI ($R^2 = 0.576$), and NDVI ($R^2 = 0.011$). The convergence of near-optimal performance across multiple red-edge contrast formulations ($\Delta R^2 < 0.005$) confirmed that the dominant chlorophyll signal resides in the red-edge region broadly, rather than in any single band pairing.

Under leave-one-out cross-validation (N = 333), RESDI attained $R^2 = 0.7995$ and RMSE = 2.6947 SPAD values under both simple linear and partial least squares regression, and $R^2 = 0.7733$ and RMSE = 2.8652 under support vector machine regression. Its stability across frameworks of varying complexity indicates that performance originates from spectral formulation rather than model flexibility. The blue and NIR normalization denominator reduces the influence of leaf surface reflectance and canopy structural scattering, decoupling the biochemical signal from architectural variability, while the red-edge numerator targets the spectral domain of highest chlorophyll sensitivity.

The wavelet-guided optimization procedure applied here is reproducible and adaptable. Its application to corn LCI estimation suggests potential extension to other crops, growth stages, physiological traits, and airborne or satellite hyperspectral platforms.

5. Acknowledgments

The authors gratefully acknowledge the financial support provided by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for the development of this research and the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Funding Code 001.

6. Reference

Arlot, S. and Celisse, A. (2010) 'A survey of cross-validation procedures for model selection', *Statistics Surveys*, 4, pp. 40–79. Available at: https://doi.org/10.1214/09-SS054.

Blackburn, G.A. and Ferwerda, J.G. (2008) 'Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis', *Remote Sensing of Environment*, 112(4), pp. 1614–1632. Available at: https://doi.org/10.1016/j.rse.2007.08.005.

Chen, J. *et al.* (2004) 'A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter', *Remote Sensing of Environment*, 91(3–4), pp. 332–344. Available at: https://doi.org/10.1016/j.rse.2004.03.014.

Daughtry, C.S.T. *et al.* (1993) 'Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance', 4257(00).

Daughtry, C.S.T. *et al.* (2000) 'Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance', *Remote Sensing of Environment*, 74(2), pp. 229–239. Available at: https://doi.org/10.1016/S0034-4257(00)00113-9.

Dong, L. *et al.* (2024) 'In-season dynamic diagnosis of maize nitrogen status across the growing season by integrating proximal sensing and crop growth modeling', *Computers and Electronics in Agriculture*, 224(July), p. 109240. Available at: https://doi.org/10.1016/j.compag.2024.109240.

Fan, L. *et al.* (2019) 'Hyperspectral-Based Estimation of Leaf Nitrogen Content in Corn Using Optimal Selection of Multiple', (1).

Gitelson, A. and Merzlyak, M.N. (1994) 'Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves', *Journal of Photochemistry and Photobiology, B: Biology*, 22(3), pp. 247–252. Available at: https://doi.org/10.1016/1011-1344(93)06963-4.

Gitelson, A. and Schepers, J. (2009) 'Non-destructive determination of maize leaf and canopy chlorophyll content', 166. Available at: https://doi.org/10.1016/j.jplph.2008.03.004.

Gitelson, A.A., Gritz, Y. and Merzlyak, M.N. (2003) 'Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves', *Journal of Plant Physiology*, 160(3), pp. 271–282. Available at: https://doi.org/10.1078/0176-1617-00887.

Gitelson, A.A., Kaufman, Y.J. and Merzlyak, M.N. (1996) 'Use of a green channel in remote sensing of global vegetation from EOS- MODIS', *Remote Sensing of Environment*, 58(3), pp. 289–298. Available at: https://doi.org/10.1016/S0034-4257(96)00072-7.

Jacquemoud, S. and Baret, F. (1990) 'PROSPECT: A model of leaf optical properties spectra', *Remote Sensing of Environment*, 34(2), pp. 75–91. Available at: https://doi.org/10.1016/0034-4257(90)90100-Z.

Mulla, D.J. (2012) 'Special Issue : Sensing in Agriculture Review Twenty five years of remote sensing in precision agriculture : Key advances and remaining knowledge gaps 5', *Biosystems Engineering*, 114(4), pp. 358–371. Available at: https://doi.org/10.1016/j.biosystemseng.2012.08.009.

Pettorelli, N. *et al.* (2005) 'Using the satellite-derived NDVI to assess ecological responses to environmental change', *Trends in Ecology and Evolution*, 20(9), pp. 503–510. Available at: https://doi.org/10.1016/j.tree.2005.05.011.

Schlemmer, M. *et al.* (2013) 'International Journal of Applied Earth Observation and Geoinformation Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels', *International Journal of Applied Earth Observations and Geoinformation*, 25, pp. 47–54. Available at: https://doi.org/10.1016/j.jag.2013.04.003.

Shuai, G., Fowler, A. and Basso, B. (2024) 'predictor of spatial patterns of Maize (Zea mays L) yields', pp. 963–982.

Sims, D.A. and Gamon, J.A. (2002) ‘Relationships between leaf pigment content and spectral reflectance across a wide range of species , leaf structures and developmental stages’, 81, pp. 337–354.

Zhang, J. *et al.* (2020) ‘Detection of Canopy Chlorophyll Content of Corn Based on Continuous Wavelet Transform Analysis’, (Lcc), pp. 1–20.