Development and Selection of a Visible-Range Spectral Index to Differentiate Sampling Positions in Stenocereus queretaroensis

Desarrollo y selección de un índice espectral basado en el intervalo visible para diferenciar posiciones de muestreo en Stenocereus queretaroensis

Rosa Janette Pérez Chimal[1], Jorge Ulises Muñoz Minjares[1], Claudia Angélica Rivera Romero[1], Julián Moisés Estudillo Ayala[2], Alejandro Barrientos García[3]

[1]Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica, Zacatecas, México. C.P. 99601.
[2]Universidad de Guanajuato, División de Ingenierías Campus Irapuato-Salamanca, Departamento de Optoelectrónica, Guanajuato, México. C.P. 36855.
[3] Universidad Politécnica de Guanajuato, Ingeniería Robótica, Óptica no lineal y Procesamiento de Señales, Guanajuato México, 38483
r.perez@uaz.edu.mx

Abstract

Spectral characterization of vegetation using reflectance indices commonly relies on relationships between spectral bands selected according to their ability to discriminate conditions or characteristics of interest. However, formulations operating exclusively within the visible range are relevant when spectral acquisition is restricted to 400–700 nm. This study developed and evaluated a procedure for selecting a spectral relationship capable of differentiating sampling positions P0 and P150 in Stenocereus queretaroensis. Thirty-one spectral bands at 10 nm intervals were analyzed, and 1,860 candidates were generated from four formulations: difference, ratio, normalized difference, and spectral slope. Candidates were evaluated using paired P0–P150 comparisons by campaign, Benjamini–Hochberg false discovery rate correction, generalized estimating equation models, and plant-grouped cross-validation. The 520–540 nm relationship showed one of the most consistent behaviors among the candidate set and was significant in all 13 campaigns. The normalized ND520–540 formulation, defined as (R520 − R540)/(R520 + R540), showed a mean P150−P0 difference of −0.0271 and a standardized paired effect size of d_z = −0.850; the GEE model yielded β = −0.0308 (p = 2.59 × 10^{-33}). Five-fold plant-grouped cross-validation showed consistent selection of the proposed formulation across held-out plant partitions. The results indicate that the 520–540 nm relationship represents a consistent member of a central spectral family, without demonstrating unequivocal statistical superiority over all nearby formulations. ND520–540 is therefore proposed as the primary normalized formulation for differentiating P0 and P150 under the experimental conditions studied. Independent validation is required before assessing its generalization to new plants, campaigns, or acquisition conditions.

Keywords: Stenocereus queretaroensis; spectral index; reflectance; visible range; 520–540 nm; sampling-position differentiation.

1. Introduction

Stenocereus queretaroensis is a columnar cactus of ecological and productive relevance in arid and semi-arid environments. Its fruits are part of traditional use and represent a relevant component of regional agroecological systems. Studies of the genus Stenocereus have documented its physiological adaptations to environments with limited water availability and its potential as a plant genetic resource for dry regions (Pimienta-Barrios & Nobel, 1994, 1998; Nobel & Pimienta-Barrios, 1995; García-Cruz *et al.*, 2022).

Non-destructive optical methods allow quantitative characterization of plant tissues without altering plant development. Spectral reflectance represents the optical response of a material to incident radiation and provides the basis for spectral signatures capable of describing differences among plant surfaces. These signatures can be acquired using contact spectrophotometry or multispectral systems, providing repeatable measurements under defined experimental protocols (Hernández & Montaner, 2009; Jacquemoud & Ustin, 2001; Klančnik & Gaberščik, 2016; Acevedo-Correa *et al.*, 2024).

Vegetation spectral response depends on the interaction between electromagnetic radiation and the biophysical properties of plant tissue. Within the visible range, photosynthetic pigments contribute substantially

to absorption and reflectance, while internal structure and other tissue components modify the shape of the spectral signature. Consequently, reflectance should be interpreted in relation to both plant-material characteristics and the acquisition conditions (Lichtenthaler, 1987; Roy, 1989; Jacquemoud & Ustin, 2001).

Vegetation indices are widely used to synthesize spectral information into quantitative, low-dimensional variables. Their development is based on mathematical relationships among two or more reflectance bands to emphasize differences associated with vegetation vigor, canopy structure, photosynthetic activity, or pigment content. Since the formulation of the normalized difference vegetation index (NDVI), spectral-band relationships have become a widely used approach for representing vegetation behavior from optical measurements (Rouse *et al*., 1974; Huete, 1988; Qi *et al.*, 1994).

Subsequent formulations expanded this concept through narrow-band ratios, normalized differences, and indices designed for specific vegetation conditions. Reviews of vegetation indices show that band selection depends on the biological objective and the acquisition system, while spectral relationships may respond differently across species and experimental conditions (Carter, 1994; Huete, 2012; Xue & Su, 2017; Zeng *et al.,* 2022). Therefore, development of a new index should be grounded in spectral evidence from the material under study rather than relying solely on transfer of existing formulations.

A frequent limitation is that many widely used indices incorporate red-edge or near-infrared bands. Although these regions provide valuable information for numerous applications, their use depends on the measurement system having access to those wavelengths. When spectral acquisition is restricted to the visible range of 400–700 nm, it is necessary to determine whether relationships among the available visible bands contain sufficient information to differentiate conditions or sampling positions (Huete, 2012; Xue & Su, 2017). This need is particularly relevant for spectrophotometric systems designed exclusively for the visible range.

In Stenocereus queretaroensis, spectral signatures acquired directly from the stem provide an opportunity to examine optical differences among positions within the same plant using a non-destructive protocol. Development of a species-specific spectral formulation should be based on systematic evaluation of relationships among visible bands and on the stability of their behavior across acquisition campaigns, without assigning a particular band in advance to a specific physiological mechanism. Studies using in-situ multispectral measurements in other plant systems demonstrate the value of spectral features and vegetation indices for discrimination tasks (Cao *et al.*, 2022).

Against this background, the objective of this study was to develop and select a spectral formulation based exclusively on the 400–700 nm visible range to differentiate experimental sampling positions P0 and P150 in Stenocereus queretaroensis. Different reflectance relationships and mathematical transformations were evaluated systematically to identify a compact, reproducible spectral variable supported by statistical evidence under the experimental conditions of the study.

2. Materials and Methods

The analysis was conducted using spectral signatures of Stenocereus queretaroensis acquired within the visible range of 400–700 nm. The experimental database comprised 109 plants evaluated across 13 sampling campaigns. Each spectral signature was represented by 31 bands at 10 nm intervals. The analysis was restricted to experimental positions P0 and P150, with plant identity retained as the clustering unit for analyses involving repeated measurements. Spectral measurements were obtained using a YS45 spectrophotometer under D65 illumination with an 8 mm aperture. Measurements were acquired at the P0 and P150 positions of each plant. The resulting signatures were organized by plant, sampling position, campaign, and wavelength to preserve the experimental structure and support subsequent paired, temporal, and repeated-measures analyses.

The overall methodological workflow used for the development and selection of the spectral index is presented in Figure 1.

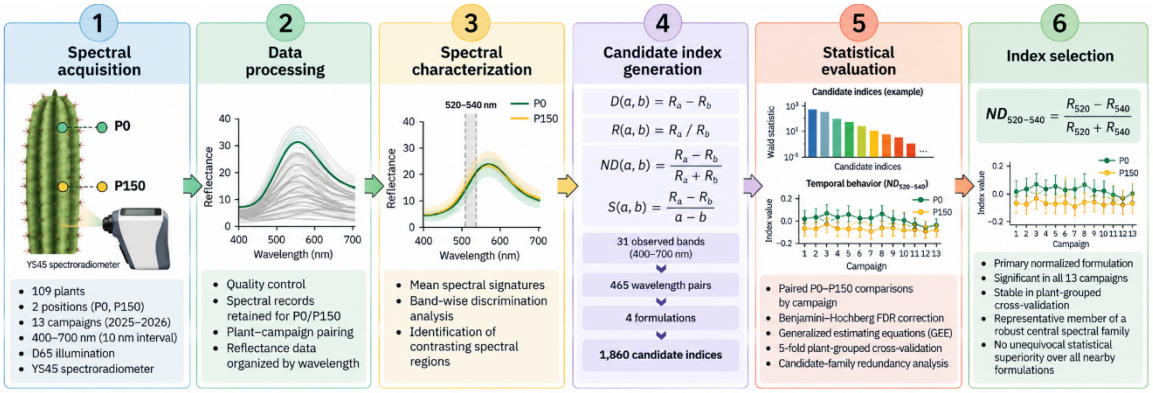


Figure 1. Methodological workflow for the development and selection of $ND_{520-540}$ in Stenocereus queretaroensis. Source: Authors' own work.

2.1 Experimental design and spectral database

Spectral comparisons were established between reflectance values recorded at P0 and P150 for each wavelength. These measurements were used to construct a search space for systematic evaluation of relationships among the available visible bands, without incorporating spectral information outside 400–700 nm.

The design was structured as an intra-plant comparison between sampling positions. This structure preserved the correspondence between P0 and P150 measurements from the same plant and campaign, which was subsequently incorporated into paired analyses and repeated-measures models. Across 109 plants and 13 campaigns, 1,417 plant–campaign combinations were theoretically possible. Four combinations lacked a P150 record, and two additional pairs were excluded from the $ND_{520-540}$ analysis because the index could not be calculated, resulting in 1,411 valid paired comparisons. For paired comparisons, the standardized effect size was calculated as $d_z = \frac{\overline{D}}{SD_D}$, where $D = X_{P150} - X_{P0}$, $\overline{D}$ is the mean paired difference, and SD_D is its standard deviation.

$$d_z = \frac{\overline{D}}{SD_D} \quad (1)$$

where $D = X_{P150} - X_{P0}$, $\overline{D}$ is the mean paired difference, and SD_D is its standard deviation. The mean spectral reflectance profiles of P0 and P150 across the visible range are shown in Figure 2.

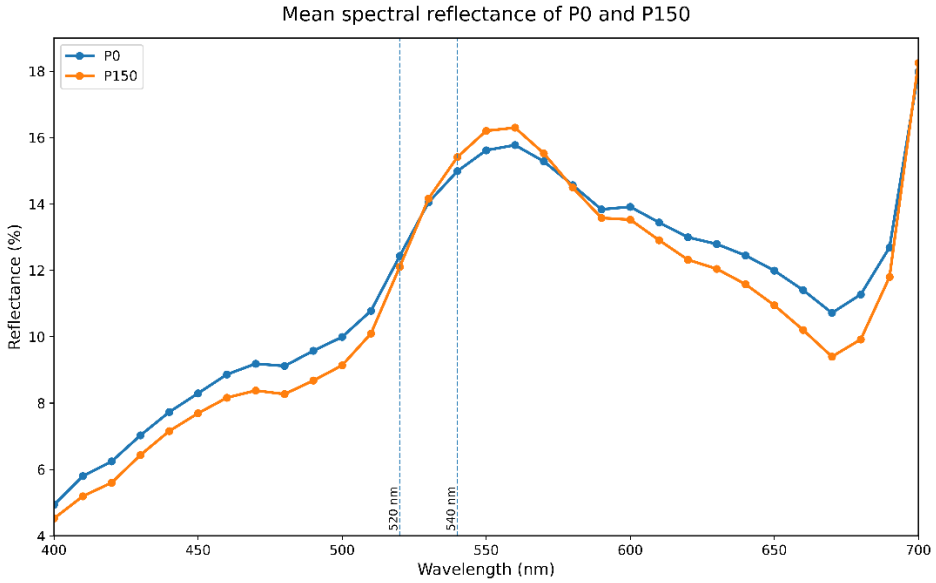


Figure 2. Mean spectral reflectance of P0 and P150 across the visible range.
Source: Authors' own work.

2.2 Construction of the complete spectral candidate space

The search space was constructed using the 31 available spectral bands between 400 and 700 nm. Rather than restricting the analysis a priori to a subset of wavelengths, all possible combinations of two distinct bands were evaluated. The 31 bands yielded 465 possible spectral pairs through combinations without repetition.

For each wavelength pair a and b, four spectral transformations were generated:

$$D(a,b) = R_a - R_b \quad (2)$$

$$R(a,b) = \frac{R_a}{R_b} \quad (3)$$

$$ND(a,b) = \frac{(R_a - R_b)}{(R_a + R_b} \quad (4)$$

$$S(a,b) = \frac{R_a - R_b}{a - b} \quad (5)$$

where R_a and R_b denote reflectance at wavelengths a and b, respectively.

Combining the 465 spectral pairs with the four transformations produced 1,860 spectral candidates. This procedure allowed the relationship between wavelengths and the effect of the mathematical transformation to be explored without defining a specific wavelength pair or final formulation a priori. The complete candidate set formed the basis for subsequent statistical screening, temporal evaluation, repeated-measures modeling, and selection of the proposed spectral formulation.

2.3 Statistical screening of candidates

Each of the 1,860 spectral candidates was initially evaluated using a one-way analysis of variance (ANOVA) to quantify the global difference between P0 and P150. For each candidate, the F statistic, p value, number of observations, mean, standard deviation, and coefficient of variation (CV) were obtained. Candidates were ranked from highest to lowest F statistic.

ANOVA was used as a global screening procedure to identify spectral relationships with greater differentiation between sampling positions. Because the data had a longitudinal structure and repeated measurements from the same plant, ANOVA results were treated as screening evidence rather than as the sole criterion for final index selection.

In addition, the 31 individual bands were evaluated by ANOVA to characterize the distribution of discriminatory behavior across the visible range. This analysis was used as a reference for interpreting the spectral location of relationships identified during candidate generation.

2.4 Temporal evaluation and multiplicity control

Temporal consistency was evaluated independently for each of the 13 sampling campaigns. For each candidate, paired comparisons between P0 and P150 were established within plant and campaign, thereby preserving correspondence between the two positions of the same individual.

Within each campaign, a paired-samples t test was applied to the 1,860 candidates. Because of the large number of simultaneous comparisons, p values were adjusted using the Benjamini–Hochberg procedure, yielding within-campaign false discovery rate (FDR)-adjusted q values.

In addition to statistical significance, the standardized effect size of the paired differences (d_z) was estimated for each candidate and campaign. The direction of the P150−P0 contrast was also recorded.

Temporal consistency was summarized by (i) the number of campaigns showing significance after FDR adjustment, (ii) the frequency of the contrast direction, and (iii) the distribution of the absolute d_z values across campaigns. Direction was used descriptively rather than as an isolated selection criterion because its interpretation depends on the mathematical formulation.

2.5 Evaluation using GEE models of paired differences

To explicitly incorporate dependence arising from repeated measurements on the same plant, the paired difference was calculated as:

$$\Delta_{ij} = P150_{ij} - P0_{ij} \quad (6)$$

where i denotes the plant and j the campaign.

For each candidate, a generalized estimating equation (GEE) model with a Gaussian distribution and identity link was fitted. Campaign was included as a factor and plant was used as the clustering unit. An exchangeable working correlation structure and consistent sandwich standard errors were used (Liang & Zeger, 1986).

For each candidate, the adjusted mean difference, consistent standard error, Wald statistic, z statistic, and p value were obtained. P values were adjusted using the Benjamini–Hochberg procedure to facilitate comparison among candidates.

The GEE model was used to evaluate inferential evidence for P0–P150 differences while accounting for repeated measurements. Because candidates were evaluated and selected using the same experimental database, this stage was considered internal evidence for the selection procedure rather than independent validation.

2.6 Identification of spectral families and redundancy assessment

Candidate comparison was not limited to individual statistical rankings. Because different wavelength pairs and mathematical transformations can generate highly related responses, candidate similarity was assessed from the paired P150−P0 differences.

Candidates showing consistent temporal evidence and relevant effect magnitudes were grouped according to the correlation among their paired-difference vectors. Hierarchical clustering was used to group candidates with highly correlated paired-difference vectors, using an absolute correlation threshold of \(|r| \geq 0.95\) to define the main spectral families. This procedure identified sets of candidates with similar spectral response patterns and distinguished central relationships, relationships connecting central and red regions, and relationships localized mainly in the red region.

An exchangeable working correlation structure was used because repeated observations from the same plant were expected to share within-plant correlation without assuming a specific correlation pattern across campaigns.

Representative candidates from the main families were subsequently selected to determine whether simultaneous inclusion of different spectral relationships provided additional information for differentiating P0 and P150. This evaluation used multivariate GEE models, with sampling position as the response, plant as the clustering unit, and campaign as an adjustment factor.

The purpose of this stage was to determine whether different families provided sufficient complementary evidence to justify retaining more than one spectral relationship in the final formulation.

2.7 Selection of the relationship and proposed spectral formulation

Final selection was based on a hierarchical assessment of evidence rather than an arbitrary composite score. The following were considered jointly: (i) global screening behavior; (ii) significance after FDR across campaigns; (iii) magnitude and temporal stability of effect size; (iv) GEE evidence; (v) redundancy among candidates and spectral families; and (vi) mathematical properties of the formulations.

The comparison was performed at two levels: first, the spectral relationship between wavelengths and, subsequently, the mathematical transformation applied to that relationship. This approach prevented differences in scale among formulations from being interpreted as evidence of different spectral signals.

For the 520–540 nm relationship, normalized difference and ratio formulations were considered:

$$ND_{520-540} = (R_{520} - R_{540}) / (R_{520} + R_{540}) \quad (7)$$

$$R_{520-540} = R_{520} / R_{540} \quad (8)$$

Both formulations represent the same spectral relationship and were not considered independent indices. Their mathematical equivalence is expressed as:

$$ND_{520-540} = (R_{520-540} - 1) / (R_{520-540} + 1) \quad (9)$$

Accordingly, $ND_{520-540}$ was established as the primary proposed formulation, whereas $R_{520-540}$ was retained as an equivalent supporting formulation. This choice was based on the normalization and bounded scale of the ND formulation and does not attribute absolute statistical superiority to it over the ratio.

2.8 Plant-grouped cross-validation for stability assessment

Stability of the leading candidates was evaluated using five-fold cross-validation, with plant as the grouping unit to prevent observations from the same individual from appearing simultaneously in training and test sets.

Within each fold, candidates were ranked using only the training set according to the absolute paired effect size. Candidates selected in training were subsequently evaluated in the held-out plant set using the effect size of the P150−P0 differences.

Stability was summarized by the frequency with which each candidate appeared among the top 15 positions across the five folds, together with the mean, standard deviation, and mean absolute effect size obtained in the held-out sets.

This procedure was used as a stability assessment of the selection and not as external validation because all folds originated from the same experimental database.

2.9 Final evaluation of $ND_{520-540}$

Once the proposed formulation had been established, $ND_{520-540}$ was characterized using descriptive statistics and measures of separation between P0 and P150.

The mean, standard deviation, and coefficient of variation of the index were calculated. For paired observations, the mean P150−P0 difference and standardized effect size d_z were estimated.

Separation between the P0 and P150 distributions was additionally examined using the area under the curve (AUC), the Mann–Whitney test, and the Kolmogorov–Smirnov (KS) distance. These metrics were interpreted as complementary measures of the discriminative behavior of the index within the experimental database.

Results from this stage were considered internal evidence of index performance under the experimental conditions studied. They were not interpreted as external validation or as sufficient evidence for establishing a universal classification threshold.

3. Results

3.1 Evaluation of the candidate space

The evaluation comprised 1,860 spectral candidates derived from 465 wavelength pairs and four mathematical transformations. The global analysis showed that the highest F statistics were concentrated mainly in relationships involving the 550–680 nm region. The leading candidates were $R_{560-670}$ (F = 707.82), $R_{550-670}$ (F = 694.09), and $R_{560-660}$ (F = 692.40). This result showed that global ANOVA screening identified several spectral relationships with discriminatory capacity and that the initial ranking did not establish a single relationship as the final formulation. The statistical discrimination of the candidate spectral relationships across the visible range is presented in Figure 3.

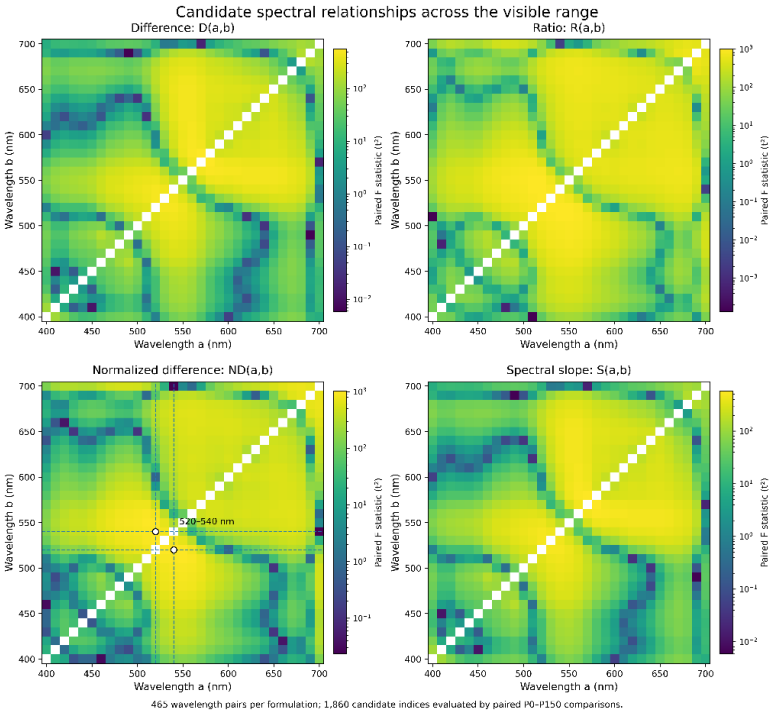


Figure 3. Statistical discrimination of candidate spectral relationships based on paired P0–P150 comparisons. Source: Authors' own work.

Selection was therefore complemented by evaluation of consistency across campaigns and by models accounting for repeated measurements. This allowed effect magnitude and stability of the P0–P150 differences to be considered jointly.

***Table 1**. Leading candidates according to the ANOVA F statistic.*

Rank	Index	F	p	CV (%)
1	$R_{560-670}$	707.82	1.03×10^{-126}	17.77
2	$R_{550-670}$	694.09	1.02×10^{-124}	18.24
3	$R_{560-660}$	692.40	1.80×10^{-124}	15.94
4	$R_{560-680}$	685.02	2.21×10^{-123}	17.54
5	$R_{570-670}$	677.88	2.47×10^{-122}	16.39
6	$R_{570-660}$	676.86	3.48×10^{-122}	14.33
7	$R_{550-680}$	676.01	4.61×10^{-122}	17.94
8	$R_{550-660}$	668.50	5.94×10^{-121}	16.56
9	$R_{560-650}$	660.57	8.59×10^{-120}	14.23
10	$R_{570-650}$	652.67	1.27×10^{-118}	12.47

3.2 Temporal consistency of candidates

Campaign-specific evaluation showed that 274 of the 1,860 candidates were significant after Benjamini–Hochberg correction in all 13 campaigns. This result indicated that multiple spectral relationships produced repeated P0–P150 differences, so selection could not be based solely on the global F statistic.

Among candidates significant in all 13 campaigns, the largest median absolute effect sizes were concentrated mainly in central spectral relationships. $ND_{520-540}$ showed the highest median |d_z|, 1.159, followed by $R_{520-540}$, 1.143. Both candidates were significant in all 13 campaigns and showed the same P150 < P0 direction in 12 of the 13 campaigns.

$ND_{520-540}$ had a mean |d_z| of 1.091 and a standard deviation of 0.401 across campaigns. The corresponding values for $R_{520-540}$ were 1.085 and 0.402. The paired comparison of absolute effect sizes did not show a statistically significant difference between the two formulations (W = 23; p = 0.127).

These results showed that the 520–540 nm relationship combined relatively large effect magnitudes with consistent behavior across campaigns. The temporal behavior of $ND_{520-540}$ across the 13 campaigns is shown in Figure 4.

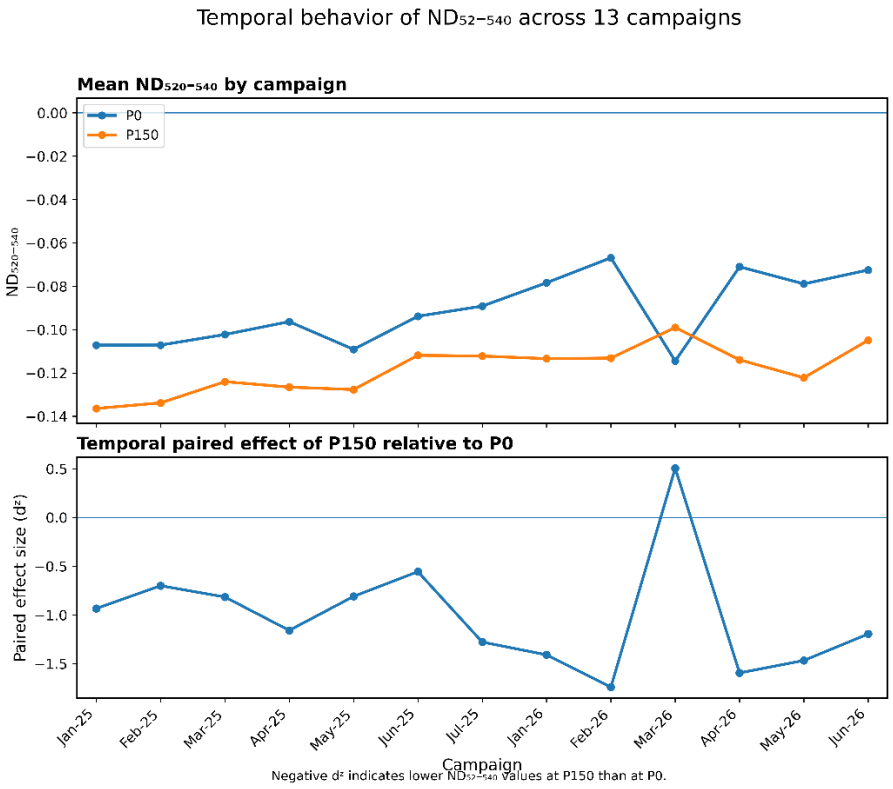


Figure 4. Temporal behavior of $ND_{520-540}$ across 13 campaigns.
Source: Authors' own work.

3.3 GEE evaluation

The GEE analysis of paired differences incorporated the repeated-measures structure through plant-level clustering and campaign adjustment. Candidates with the strongest inferential evidence were again concentrated around the central spectral region.

The largest z statistic corresponded to $ND_{510-540}$ ($z = -12.505$; $p = 7.04 \times 10^{-36}$), followed by $ND_{510-530}$ ($z = -12.502$; $p = 7.28 \times 10^{-36}$) and $ND_{500-540}$ ($z = -12.361$; $p = 4.27 \times 10^{-35}$). $ND_{520-530}$ also showed consistent statistical evidence ($z = -12.301$; $p = 9.00 \times 10^{-35}$).

For $ND_{520-540}$, z = −12.026, with an adjusted mean P150−P0 difference of −0.03080 ($p = 2.59 \times 10^{-33}$). Its equivalent $R_{520-540}$ formulation yielded z = −11.855 and an adjusted mean difference of −0.05008 ($p = 2.02 \times 10^{-32}$).

The GEE ranking did not completely coincide with the ANOVA ranking. This difference showed that candidate position depends on the statistical criterion used and that no single criterion provides a sufficient basis for final formulation selection.

***Table 2**. Representative candidates according to the GEE model.*

Candidate	z	p	Adjusted Δ P150–P0	Mean |d_z|
$ND_{510-540}$	−12.505	7.04×10^{-36}	−0.05848	1.050
$ND_{510-530}$	−12.502	7.28×10^{-36}	−0.04927	1.017
$ND_{500-540}$	−12.361	4.27×10^{-35}	−0.07779	0.997
$ND_{500-550}$	−12.324	6.72×10^{-35}	−0.07878	1.015
$ND_{520-530}$	−12.301	9.00×10^{-35}	−0.02078	1.076
$ND_{510-550}$	−12.274	1.25×10^{-34}	−0.05992	1.063
$ND_{520-540}$	−12.026	2.59×10^{-33}	−0.03080	1.091

3.4 Identification of spectral families

Candidate comparison showed that several relationships exhibited similar response patterns. To characterize this structure, correlations among the paired P150−P0 differences were evaluated for candidates with consistent temporal evidence.

The analysis distinguished a central family, a family connecting central and red regions, and a family located primarily in the red region. $ND_{520-540}$, $R_{560-670}$, and $ND_{680-700}$ were considered representative candidates, respectively.

The paired differences of these candidates showed high-magnitude correlations. The correlation between $ND_{520-540}$ and $ND_{680-700}$ was r = 0.827, whereas the correlation between $ND_{520-540}$ and $R_{560-670}$ was r = −0.824. These results indicate that part of the variation associated with sampling position is shared across different spectral relationships.

In the multivariate model, adding $R_{560-670}$ to $ND_{520-540}$ did not provide evidence of a significant additional contribution (p = 0.571). Adding $ND_{680-700}$ provided limited additional evidence (p = 0.0504), while the model including all three candidates also did not demonstrate an unequivocal additional contribution from the complementary relationships. Consequently, the results favored a single representation of the spectral signal.

3.5 Selection of $ND_{520-540}$ as the proposed formulation

The integrated evaluation identified the 520–540 nm relationship as a consistent representation within the central candidate family. $ND_{520-540}$ showed the highest median |d_z| among candidates significant in all 13 campaigns, while $R_{520-540}$ showed similar temporal behavior.

Plant-grouped cross-validation supported the stability of both formulations. $ND_{520-540}$ appeared among the top 15 candidates in all five folds and had a mean held-out absolute effect size of 0.860. $R_{520-540}$ also appeared among the top 15 in all five folds, with a mean held-out absolute effect size of 0.858.

Comparison with nearby candidates showed that $ND_{520-540}$ was not statistically superior to all neighboring relationships. The evidence was therefore interpreted as support for a consistent spectral family around the 520–540 nm region, within which $ND_{520-540}$ provides a normalized and parsimonious representation.

The final proposed formulation was:

$$ND_{520-540} = (R_{520} - R_{540}) / (R_{520} + R_{540}) \quad (10)$$

Its equivalent formulation was:

$$R_{520-540} = R_{520} / R_{540} \quad (11)$$

The ratio was retained as complementary evidence of the same spectral relationship rather than as a second independent index.

3.6 Global behavior of the proposed index

For $ND_{520-540}$, 1,411 paired plant–campaign comparisons were analyzed. The global index mean was −0.104834, with a standard deviation of 0.030421 and a coefficient of variation of 29.018%.

The mean P150−P0 difference was −0.027127, with a standard deviation of 0.031898 and a global standardized effect size d_z = −0.850. The AUC was 0.761. As complementary measures of distributional separation, the Mann–Whitney and Kolmogorov–Smirnov statistics indicated differences between the P0 and P150 distributions ($p = 1.58 \times 10^{-127}$), whereas the Kolmogorov–Smirnov test yielded D = 0.385 and $p = 1.58 \times 10^{-93}$. A total of 1,411 valid paired plant–campaign comparisons were available for $ND_{520-540}$. The global discrimination and cross-validation stability of $ND_{520-540}$ are shown in Figure 5.

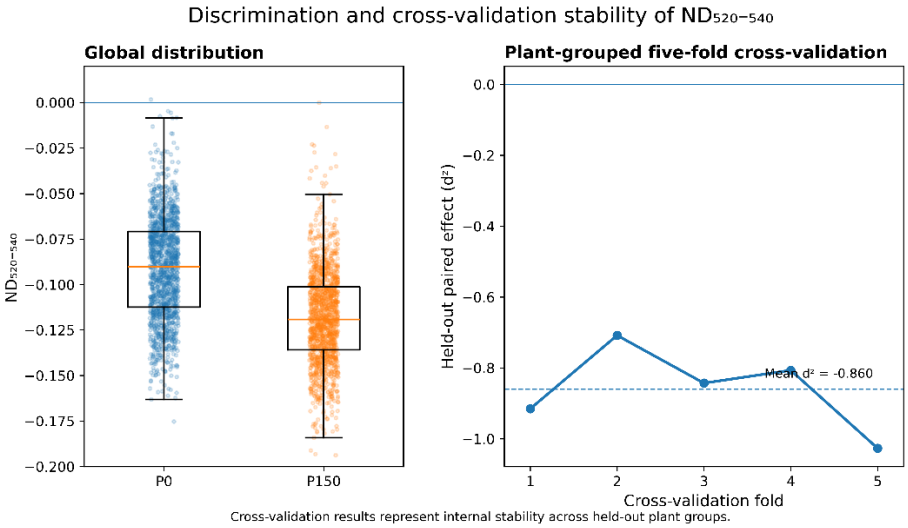


Figure 5. Discrimination and cross-validation stability of $ND_{520-540}$.
Source: Authors' own work

Overall, these results demonstrate differentiation between P0 and P150 within the analyzed database, although the two distributions remained overlapping.

***Table 3**. Global statistics for $ND_{520-540}$.*

Metric	$ND_{520-540}$
Global mean	−0.104834
SD	0.030421
CV	29.018%
Paired comparisons	1,411
Mean Δ (P150−P0)	−0.027127
d_z	−0.850
AUC	0.761
Mann–Whitney p	1.580×10^{-127}
KS D	0.385
KS p	1.583×10^{-93}

3.7 Temporal behavior of $ND_{520-540}$

$ND_{520-540}$ remained statistically significant after FDR correction in all 13 campaigns. The P150 < P0 direction was observed in 12 of the 13 campaigns.

The magnitude of the contrast varied across campaigns, and one campaign showed a reversal in direction. Therefore, the observed temporal consistency represents repeated differentiation between positions rather than a temporally constant response.

The observed variability prevents establishment of a universal campaign-independent threshold with the available evidence.

4. Discussion

Evaluation of the candidate space showed that differentiation between P0 and P150 was not restricted to a single spectral combination. Global ANOVA, temporal evaluation, and GEE models identified several relationships with discriminatory capacity, with a concentration of relevant candidates in central regions of the visible range. This pattern indicates that the signal associated with sampling position can be represented by different combinations of bands and mathematical transformations.

Selection of $ND_{520-540}$ considered effect magnitude, consistency across campaigns, repeated-measures structure, and stability observed through plant-grouped cross-validation. This approach avoided relying on a single statistic as the decision criterion. The results did not show absolute statistical superiority of $ND_{520-540}$ over all nearby candidates; therefore, its selection should be interpreted as the choice of a consistent and parsimonious representation within a family of related spectral relationships.

The concurrent behavior of $ND_{520-540}$ and $R_{520-540}$ among the most consistent candidates suggests that the discriminative information is primarily contained in the relationship between reflectance at 520 and 540 nm rather than in a specific mathematical transformation. The two formulations are mathematically equivalent and show an almost perfect correspondence; therefore, they should not be interpreted as two independent indices.

The normalized $ND_{520-540}$ formulation was adopted as the primary representation because it expresses the relative contrast between the two bands on a normalized and bounded scale. This is a mathematical representation choice rather than a claim of statistical superiority over $R_{520-540}$.

Results for nearby candidates, such as $ND_{510-540}$, $ND_{520-530}$, and $ND_{520-550}$, show that the signal should not be interpreted as exclusive to a single wavelength pair. In this sense, 520–540 nm can be considered a representative relationship within a central spectral interval showing discriminatory behavior under the conditions studied.

Correlation analysis of paired differences identified candidate families with related response patterns. In addition to the central family represented by $ND_{520-540}$, relationships connecting central and red regions and relationships located mainly in the red region were observed.

Multivariate evaluation did not provide sufficient evidence to justify retaining a second relationship as an independent index. In particular, $R_{560-670}$ did not show a significant additional contribution after $ND_{520-540}$ was included in the model ($p = 0.571$). The contribution of $ND_{680-700}$ was limited and did not unequivocally establish the need for a second variable. The final formulation was therefore retained as a single index, avoiding additional complexity without sufficient evidence of complementary information.

This result also indicates that different candidates should not be interpreted as independent signals. Part of their discriminative capacity may derive from shared variation patterns across the visible spectrum.

$ND_{520-540}$ was statistically significant in all 13 campaigns, and the P150 < P0 direction was observed in 12 of them. This repeated contrast provides evidence that the spectral relationship can differentiate the sampling positions across different acquisition campaigns.

However, effect magnitude varied across campaigns and one campaign showed a reversal in direction. The results therefore do not support interpreting the index as a temporally invariant variable. The observed variation should be considered part of the index behavior under different acquisition conditions and within the experimental context.

Accordingly, the index can be used as a discriminative variable for the P0–P150 contrast under the conditions studied, but definition of a fixed threshold or campaign-independent classification rule requires additional evaluation.

The scope of the results is delimited by the experimental design and the database used for index development and evaluation. All selection and evaluation stages were derived from the same database; therefore, GEE analyses, temporal comparisons, separation metrics, and cross-validation constitute internal evidence of the procedure rather than external validation.

Plant-grouped cross-validation provides information about selection stability across different partitions of individuals but does not replace evaluation using an independent dataset. External validation will be required to determine generalization to new plants, campaigns, or experimental conditions.

No independent biochemical or physiological measurements were available to associate the 520–540 nm relationship specifically with a pigment, water content, tissue structure, or another physiological mechanism. Accordingly, interpretation is limited to the observed spectral association between P0 and P150.

Finally, the results do not support establishment of a universal classification threshold. Variation across campaigns indicates that any decision criterion should be evaluated subsequently using independent data under clearly defined acquisition conditions.

The principal methodological contribution of this study is a procedure for developing a specific spectral variable from information restricted to the visible range. Evaluation of relationships among bands and mathematical transformations allowed identification of a formulation that summarizes the observed difference between sampling positions without relying on near-infrared bands.

Selection of $ND_{520-540}$ as the primary formulation provides a normalized variable that can be evaluated in future independent studies. The result is not intended to replace indices developed for other objectives or spectral regions, but rather to establish a visible-range relationship specific to the experimental contrast considered.

The evidence also indicates that spectral-relationship analysis should account for dependence among observations and temporal variability. The combination of paired comparisons, multiplicity control, repeated-measures models, and stability assessment provides a framework for future research aimed at developing application-specific spectral variables.

5. Conclusions

Analysis of spectral signatures in the 400–700 nm visible range identified spectral relationships capable of differentiating experimental positions P0 and P150 in *Stenocereus queretaroensis*. Evaluation of 1,860 candidates showed that discriminatory capacity was distributed across different wavelength combinations and mathematical transformations; therefore, formulation selection cannot be supported by a single statistical criterion.

Joint evaluation of effect magnitude, consistency across campaigns, repeated-measures models, and plant-level stability identified the 520–540 nm region as a representative relationship within a spectral family showing consistent behavior. Within this relationship, $ND_{520-540}$ and $R_{520-540}$ showed comparable results. Because of its normalized and bounded formulation, $ND_{520-540}$ was established as the primary proposed formulation, while $R_{520-540}$ was treated as an equivalent representation of the same spectral relationship.

$ND_{520-540}$ showed a mean P150−P0 difference of −0.027127 and a global standardized effect size d_z = −0.850 in the paired comparisons. The AUC of 0.761 and the Mann–Whitney and Kolmogorov–Smirnov tests indicated separation between the P0 and P150 distributions within the studied database.

Index significance remained after FDR correction in all 13 campaigns, while the P150 < P0 direction was observed in 12 campaigns. Variation in effect magnitude and the direction reversal observed in one campaign indicate consistent differentiation but not temporal invariance.

Plant-grouped cross-validation showed stability of $ND_{520-540}$ selection across the analyzed partitions. Nevertheless, this assessment and the other metrics used constitute internal evidence derived from the same experimental database and do not replace validation with independent data.

Overall, $ND_{520-540}$ constitutes a proposed visible-range spectral formulation for the P0–P150 contrast under the experimental conditions studied. Its application to new plants, campaigns, or acquisition conditions should be evaluated with independent data before establishing classification criteria or generally applicable thresholds.

Acknowledgments

Rosa Janette Pérez Chimal has received a fellowship No. CVU 290119 from Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI, formerly CONAHCYT) to conduct a research stay at Universidad de Guanajuato.

References

Acevedo-Correa, C., Goez, M., Torres-Madronero, M. C., & Rondon, T. (2024). Low-cost clamp for the measurement of vegetation spectral signatures. HardwareX, 19, e00557. https://doi.org/10.1016/j.ohx.2024.e00557

Cao, Q., Yang, G., Wang, F., Chen, L., Xu, B., Zhao, C., Duan, D., Jiang, P., Xu, Z.-T., & Yang, H. (2022). Discrimination of tea plant variety using in-situ multispectral imaging system and multi-feature analysis. Computers and Electronics in Agriculture, 202, 107360. https://doi.org/10.1016/j.compag.2022.107360

Carter, G. A. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. International Journal of Remote Sensing, 15(3), 697–703. https://doi.org/10.1080/01431169408954109

García-Cruz, L., Valle-Guadarrama, S., Guerra-Ramírez, D., Martínez-Damián, M. T., & Zuleta-Prada, H. (2022). Cultivation, quality attributes, postharvest behavior, bioactive compounds, and uses of Stenocereus: A review. Scientia Horticulturae, 304, 111336. https://doi.org/10.1016/j.scienta.2022.111336

Hernández, J., & Montaner, D. (2009). Patrones de respuesta espectral. In C. Mena (Ed.), Tecnologías geoespaciales: Experiencias aplicadas al estudio y gestión del territorio. Facultad de Ciencias Forestales Universidad de Chile. https://www.gep.uchile.cl/Publicaciones/Hern%C3%A1ndez%20%26%20Montaner%202009%20SAF.pdf

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Huete, A. R. (2012). Vegetation indices, remote sensing and forest monitoring. Geography Compass, 6(9), 513–532. https://doi.org/10.1111/j.1749-8198.2012.00507.x

Jacquemoud, S., & Ustin, S. L. (2001). Leaf optical properties: A state of the art. Proceedings of the 8th International Symposium on Physical Measurements & Signatures in Remote Sensing, 223–332.

Klančnik, K., & Gaberščik, A. (2016). Leaf spectral signatures differ in plant species colonizing habitats along a hydrological gradient. Journal of Plant Ecology, 9(4), 442–450. https://doi.org/10.1093/jpe/rtv068

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. Methods in Enzymology, 148, 350–382. https://doi.org/10.1016/0076-6879(87)48036-1.

Nobel, P. S., & Pimienta-Barrios, E. (1995). Monthly stem elongation for Stenocereus queretaroensis: Relationships to environmental conditions, net CO_2 uptake and seasonal variations in sugar content. Environmental and Experimental Botany, 35(1), 17–24. https://doi.org/10.1016/0098-8472(94)00037-6.

Pimienta-Barrios, E., & Nobel, P. S. (1994). Pitaya (Stenocereus spp., Cactaceae): Un cultivo antiguo y moderno de frutas en México. Economic Botany, 48, 76–83. https://doi.org/10.1007/BF02901385

Pimienta-Barrios, E., & Nobel, P. S. (1998). Vegetative, reproductive, and physiological adaptations to aridity of pitayo (Stenocereus queretaroensis, Cactaceae). Economic Botany, 54, 401–411. https://doi.org/10.1007/BF02862071

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index (MSAVI). Remote Sensing of Environment, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. Proceedings of the Third Earth Resources Technology Satellite Symposium, NASA SP-351, 309–317. https://ntrs.nasa.gov/citations/19740022614

Roy, P. S. (1989). Spectral reflectance characteristics of vegetation and their use in estimating productive potential. Proceedings: Plant Sciences, 99, 59–81. https://doi.org/10.1007/BF03053419

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. Journal of Sensors, 2017, 1353691. https://doi.org/10.1155/2017/1353691

Zeng, Y., Hao, D., Huete, A. R., Dechant, B., Berry, J. A., Chen, J. M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., & Xiao, J. (2022). Optical vegetation indices for monitoring terrestrial ecosystems globally. Nature Reviews Earth & Environment, 3(7), 477–493. https://doi.org/10.1038/s43017-022-00298-5.