remote sensing

MDPI

Article

Open-Source Reproducible Pipeline for Multitemporal Vegetation Monitoring Using Sentinel-2 L2A in Cloud-Prone Tropical Regions

Kevin David Ortega-Quiñones *, Daniel Zapata-Yarce, Michael Felipe Cifuentes-Molano, Mauricio Holguín-Londoño and Germán Andrés Holguín-Londoño

Electrical Engineering Program, Faculty of Engineering, Technological University of Pereira, Pereira 660003, Colombia; daniel.zapata1@utp.edu.co (D.Z.-Y.); mcifuentes@utp.edu.co (M.F.C.-M.); mau.hol@utp.edu.co (M.H.-L.); gahol@utp.edu.co (G.A.H.-L.)
* Correspondence: kevin.ortega@utp.edu.co

Highlights

What are the main findings?

- A storage-efficient open-source pipeline using GDAL `/vsizip/` reduces per-scene Sentinel-2 L2A processing time by ≈98% and per-pixel SCL cloud masking retains 2.2× more temporal observations than conventional scene-level thresholding over the Colombian Eje Cafetero (2017–2025).
- No statistically significant inter-annual vegetation trend was detected across NDVI, EVI, SAVI, or NDMI ($p > 0.05$); however, external validation revealed persistent spectral ambiguity between shade-grown coffee and secondary forest, with overall agreement of only ≈20% against ESA WorldCover and Google Earth Pro reference data.

What are the implications of the main findings?

- Reproducible multitemporal vegetation monitoring of cloud-prone tropical regions is achievable on standard hardware without proprietary software or cloud computing, lowering the barrier for data-scarce institutions in developing countries.
- Single-date optical spectral indices are insufficient to distinguish shade-grown coffee, secondary forest, and dense forest in heterogeneous Andean landscapes, motivating the integration of multitemporal, SAR, and phenological features in future workflows.

Abstract

Monitoring vegetation-index dynamics in tropical regions remains challenging due to persistent cloud contamination, landscape heterogeneity, and the lack of standardised and reproducible analytical workflows. This paper presents an open-source, fully reproducible end-to-end methodology for multitemporal vegetation monitoring using Sentinel-2 Level-2A (L2A) Bottom-of-Atmosphere (BOA) reflectance imagery. The methodology was applied to Military Grid Reference System (MGRS) tile T18NVL in the Colombian Eje Cafetero region (4.43°N–5.43°N, 74.91°W–75.90°W) for the 2017–2025 period. The proposed workflow integrates storage-efficient direct extraction of spectral reflectance from compressed Standard Archive Format for Europe (SAFE) archives using the Geospatial Data Abstraction Library (GDAL) `/vsizip/` interface, per-pixel cloud and shadow masking based on the Sentinel-2 Scene Classification Layer (SCL), computation of Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), and Normalized Difference Moisture Index (NDMI) spectral indices, a diagnostic Random Forest experiment based on threshold-labelled spectral classes,

Check for updates

Academic Editor: Yang-Won Lee

Received: 28 May 2026
Revised: 2 July 2026
Accepted: 6 July 2026
Published: 3 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

non-parametric Mann–Kendall trend analysis with Sen's slope estimation, and external agreement assessment against European Space Agency (ESA) WorldCover 10 m 2020 and Google Earth Pro reference data. The methodology reduced per-scene I/O time by approximately 98% without additional disk overhead while retaining a median of 57.9% valid pixels under a mean scene cloud fraction of 35.4%. Mann–Kendall analysis detected no statistically significant long-term trend in any of the four vegetation indices. Seasonal NDVI peaks during September–November were consistent with the bimodal regional precipitation regime, supporting temporal coherence in the satellite-derived vegetation-index response. External agreement was low, with an Overall Accuracy (OA) of 20.2% against ESA WorldCover and 19.9% against Google Earth Pro, indicating systematic over-prediction of woody and mixed-canopy vegetation classes. These results show that single-date optical spectral indices are insufficient for reliable thematic separation of shade-grown coffee, secondary forest, and dense forest within heterogeneous tropical landscapes. The complete version-controlled codebase is publicly available to support methodological reproducibility and adaptation across data-scarce tropical regions.

Keywords: Sentinel-2 L2A; NDVI; EVI; SAVI; NDMI; vegetation monitoring; scene classification layer; cloud masking; Random Forest; diagnostic spectral classification; Colombian Eje Cafetero; open-source remote sensing; reproducible science

1. Introduction

Terrestrial vegetation mediates the carbon and water cycles and sustains biodiversity across biomes [1,2]. The opening of multispectral satellite archives, from Landsat [3] to the Copernicus Sentinel-2 constellation [4,5], has made sub-decadal, sub-hectare time series freely accessible worldwide. Nevertheless, a structural capability gap persists in tropical developing regions [6,7], where persistent cloud cover, heterogeneous land-cover mosaics, and limited computational infrastructure hinder the conversion of raw satellite archives into reproducible processing workflows and monitoring-ready information products [8,9].

This challenge is particularly pronounced in the Colombian Andes, where highly fragmented tropical landscapes combine dense forest, shade-grown coffee systems, páramo ecosystems, and expanding urban areas within regions frequently affected by cloud contamination. Such environmental complexity requires monitoring frameworks capable of preserving valid observations from partially clouded scenes while maintaining sensitivity to multiple vegetation and moisture conditions.

The Normalised Difference Vegetation Index (NDVI) [10] is the most widely used biophysical proxy for vegetation monitoring, but saturates at leaf-area indices above ≈3, typical of dense tropical forest [11]. The Enhanced Vegetation Index (EVI) [11] incorporates canopy-background and aerosol resistance corrections, extending sensitivity into high-biomass regimes. The Soil-Adjusted Vegetation Index (SAVI) [12] reduces soil-brightness effects, whereas the Normalised Difference Moisture Index (NDMI) combines near-infrared and shortwave infrared reflectance to characterise canopy water content and forest disturbance sensitivity [13,14]. The combined use of these indices enables complementary assessment of vegetation condition across heterogeneous tropical environments [15,16].

Several studies have applied Sentinel-2 imagery to regional vegetation monitoring in Colombia and other tropical environments [6,7,17]. However, existing implementations often depend on proprietary software or cloud-based platforms [18], while standardised workflows for per-pixel cloud masking, reproducible multi-index analysis, and diagnostic

assessment under data-scarce tropical conditions remain limited. As a result, long-term analyses covering the full Sentinel-2 archive for the Colombian Andes are still scarce.

To address these limitations, this study proposes a fully reproducible, open-source framework for multitemporal vegetation monitoring using Sentinel-2 Level-2A imagery. The main contributions are as follows:

1. A storage-efficient and reproducible processing workflow for direct analysis of Sentinel-2 SAFE archives without decompression.
2. A standardised per-pixel cloud-masking methodology designed to maximise valid-observation retention in persistently clouded tropical regions.
3. A multi-index vegetation-monitoring framework integrating NDVI, EVI, SAVI, and NDMI for complementary assessment of vegetation condition and moisture dynamics.
4. A diagnostic threshold-labelled Random Forest experiment designed to assess the limitations of single-date optical spectral-index classification in heterogeneous tropical landscapes.
5. A transparent temporal and external agreement assessment combining non-parametric vegetation-index trend analysis with independent reference comparisons under data-scarce conditions.
6. An openly available, version-controlled implementation intended to support methodological reproducibility and adaptation across tropical monitoring applications.

2. Materials and Methods

2.1. Study Area

The study domain is MGRS tile T18NVL, a 109.8 km × 109.8 km Sentinel-2 tile located in the Colombian Western Andes (Figure 1). The tile spans 4.43°N–5.43°N and 74.91°W–75.90°W (EPSG:32618, World Geodetic System 1984 (WGS 84)/Universal Transverse Mercator (UTM) Zone 18N), covering the departments of Caldas, Risaralda, and parts of Antioquia and Quindío.

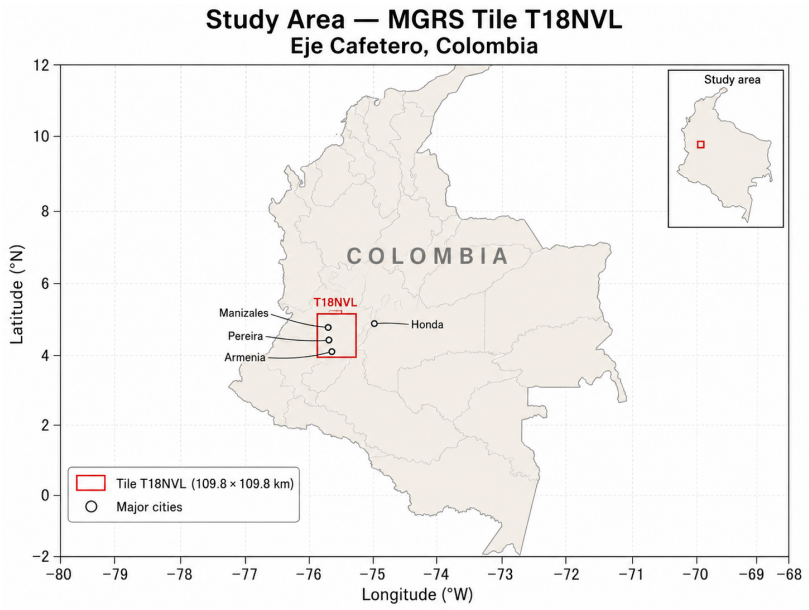


Figure 1. Study area: MGRS tile T18NVL (red rectangle, 109.8 km × 109.8 km; EPSG:32618). The inset shows its location within South America. The tile covers the Colombian Eje Cafetero, a heterogeneous tropical Andean landscape combining urban areas, agroforestry systems, cloud forest, and high-altitude páramo.

Altitude ranges from ≈700 m a.s.l. in the Valle del Cauca to over 3500 m on the páramo flanks of the Cordillera Central. The climate is bimodal, with wet seasons in March–May and September–November and mean annual precipitation of ≈2000–2500 mm [19]. This combination of strong topographic gradients, persistent cloud contamination, and mixed vegetation–urban mosaics makes the tile a suitable benchmark for testing a reproducible Sentinel-2 processing workflow.

Figure 2 shows false-colour and true-colour Sentinel-2 visualisations for the selected low-cloud reference scene.

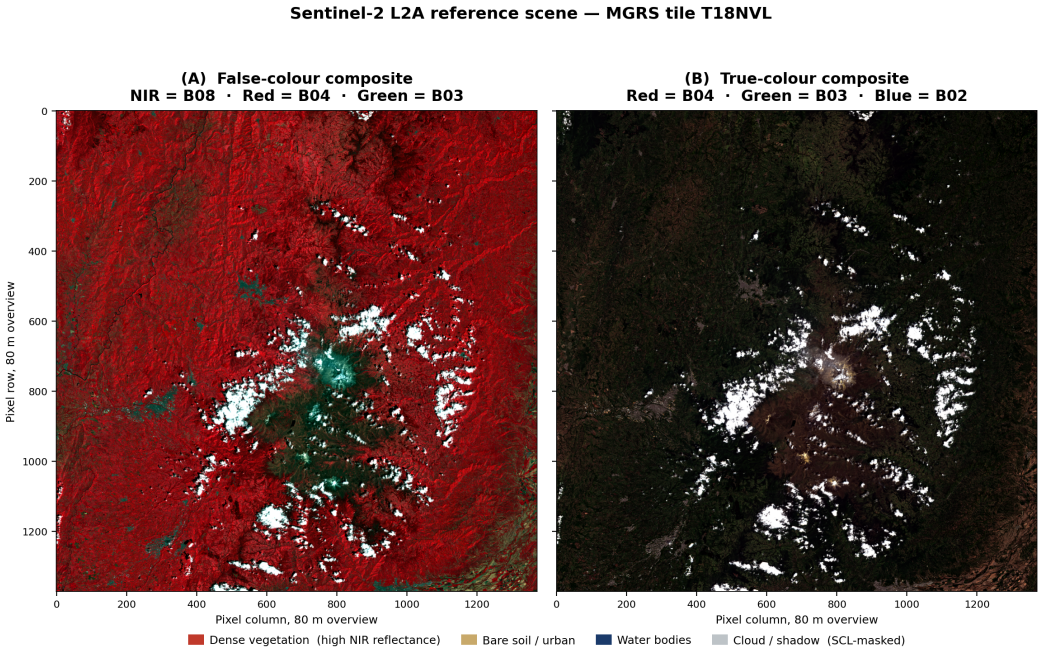


Figure 2. Sentinel-2 L2A visualisation of MGRS tile T18NVL (18 February 2020; cloud cover: 5.6%; low-cloud reference scene). (**A**) False-colour composite (B08/B04/B03), where dense vegetation appears bright red, bare soil and urban areas appear cyan or grey, and water bodies appear dark blue. (**B**) True-colour composite (B04/B03/B02). Both panels are displayed at ×8 spatial overview (effective pixel size 80 m) after per-channel P2–P98 percentile stretching. Grey pixels correspond to areas removed by the SCL-based cloud and shadow mask.

2.2. Sentinel-2 Scene Inventory

All imagery was obtained from the ESA Copernicus Data Space Ecosystem as Sentinel-2 L2A products, providing BOA surface reflectance processed with Sen2Cor v2.11. The scene inventory was used to demonstrate and assess the reproducible processing workflow under variable cloud conditions; it was not intended as an exhaustive climatological or ecological time-series sample.

Table 1 summarises the 20 scenes considered in the workflow assessment (S2A: 12, S2B: 7, S2C: 1; orbit 025; 2017–2025). Scene-level cloud cover ranged from 5.47% to 66.81% (mean: 35.4 ± 18.8%). Each L2A product is a compressed SAFE archive (mean size: 1.17 GB; total dataset: 23.3 GB) containing multispectral images at three spatial resolutions: 10 m (B02 = 490 nm, B03 = 560 nm, B04 = 665 nm, B08 = 842 nm), 20 m (B05–B07, B8A = 865 nm, B11 = 1610 nm, B12 = 2190 nm, SCL), and 60 m.

The inventory includes one scene acquired by Sentinel-2C (S2C_MSIL2A_20250909). Sentinel-2C was launched in September 2024 as the third unit of the Copernicus Sentinel-2 constellation and was included here to maintain consistency with the operational Sentinel-2 L2A product stream. The product metadata report the same Sen2Cor v2.11 processing

chain used for the other scenes; therefore, the S2C scene was treated as compatible with S2A and S2B for the purposes of this workflow assessment [4].

Table 1. Sentinel-2 L2A scenes considered in this study (MGRS tile T18NVL, orbit 025). Cloud cover (%) is the scene-level assessment from the ESA metadata. Valid pixels (%) refers to the fraction retained after SCL-based cloud and shadow masking. † Scene selected as the diagnostic spectral-classification reference. ‡ Scenes used for the scene-to-scene NDVI difference diagnostic. * Scene retained in the inventory but excluded from pixel-level index summaries due to RAM constraints (16 GB) on the development workstation.

Acquisition Date	Satellite	Product ID (Abbrev.)	Cloud (%)	Valid px (%)	Mean NDVI
2017-02-23	S2A	S2A_MSIL2A_20170223	53.9	40.5	0.461 ‡
2017-08-17	S2B	S2B_MSIL2A_20170817	30.4	65.2	0.481
2017-09-01	S2A	S2A_MSIL2A_20170901	5.5	91.4	0.489
2018-01-04	S2B	S2B_MSIL2A_20180104	37.7	58.3	0.476
2018-02-28	S2A	S2A_MSIL2A_20180228	37.8	59.2	0.449
2018-08-17	S2A	S2A_MSIL2A_20180817	42.3	52.1	0.471
2018-12-10	S2B	S2B_MSIL2A_20181210	47.3	47.6	0.463
2019-01-04	S2A	S2A_MSIL2A_20190104	20.9	74.8	0.481
2019-08-12	S2A	S2A_MSIL2A_20190812	29.6	64.5	0.476
2019-08-27	S2B	S2B_MSIL2A_20190827	19.6	—	— *
2020-01-09	S2A	S2A_MSIL2A_20200109	19.2	75.4	0.483
2020-02-18	S2A	S2A_MSIL2A_20200218	5.6	92.0	0.469 †
2020-07-17	S2A	S2A_MSIL2A_20200717	36.5	57.9	0.484
2021-03-14	S2A	S2A_MSIL2A_20210314	59.9	34.4	0.475
2023-03-19	S2B	S2B_MSIL2A_20230319	65.2	30.8	0.459
2023-07-22	S2A	S2A_MSIL2A_20230722	64.1	33.7	0.425
2023-08-26	S2B	S2B_MSIL2A_20230826	12.5	—	— *
2023-10-05	S2B	S2B_MSIL2A_20231005	66.8	—	— *
2024-03-18	S2A	S2A_MSIL2A_20240318	27.0	67.1	0.471
2025-09-09	S2C	S2C_MSIL2A_20250909	27.2	68.4	0.516 ‡
Summary		$n = 20$	35.4 ± 18.8	57.9 (med.)	0.472 ± 0.018 ($n = 17$)

2.3. Processing Pipeline Overview

The processing workflow was implemented as a sequence of reproducible Python 3.12 scripts (Algorithm 1) that convert compressed Sentinel-2 L2A SAFE archives into vegetation-index products, temporal summaries, and diagnostic assessment outputs. The workflow was designed to minimise intermediate storage, preserve valid pixels in partially clouded scenes through SCL-based masking, and document the limitations of a Sentinel-2-only spectral-index configuration under tropical conditions.

2.4. Storage-Efficient Data Access via `/vsizip/`

Sentinel-2 L2A products are distributed as SAFE archives containing multiple JPEG 2000 (JP2) image assets, metadata files, and auxiliary layers. In extraction-based workflows, the complete SAFE archive is first decompressed before the required bands are read. This can generate substantial intermediate storage when multiple scenes are processed.

To avoid full-archive decompression, the proposed workflow accesses the required JP2 assets directly inside the compressed SAFE archive using the GDAL Virtual File System interface `/vsizip/`. Raster access was implemented through `rasterio`, which relies on GDAL's virtual file system support for datasets stored in archive files:

```
vsi_path = f"/vsizip/{zip_path}/{inner_jp2_path}"
with rasterio.open(vsi_path) as src:
    arr = src.read(1)
```

Algorithm 1: Overview of the reproducible Sentinel-2 processing and diagnostic assessment pipeline

Input: Sentinel-2 L2A SAFE.zip archives and scene inventory Comma-Separated Values (CSV)

Output: Vegetation-index Geographic Tagged Image File Formats (GeoTIFFs), summary tables, figures, diagnostic spectral-classification layer, and external-agreement outputs

Data: Scene collection $\mathcal{S}$, vegetation-index collection $\mathcal{D}$, valid-pixel mask $\mathcal{M}_{valid}$

Read metadata from Sentinel-2 L2A SAFE archives
Extract acquisition date, scene-level cloud percentage, satellite ID, and product paths
Generate or update the scene inventory CSV

foreach *scene* $s_i \in \mathcal{S}$ **do**
    Open the required spectral bands directly from the compressed SAFE archive using GDAL `/vsizip/`
    Load the Sentinel-2 Scene Classification Layer (SCL)
    Build the valid-pixel mask $\mathcal{M}_{valid}$ from selected SCL classes
    Apply $\mathcal{M}_{valid}$ to the spectral bands
    Compute NDVI, EVI, SAVI, and NDMI
    Export the vegetation-index products as GeoTIFF files and append them to $\mathcal{D}$
end

Load the valid vegetation-index products from $\mathcal{D}$
Compute scene-level and seasonal vegetation-index summaries
Estimate OLS and non-parametric Mann–Kendall trends for each vegetation-index time series
Estimate Sen's slope for NDVI, EVI, SAVI, and NDMI
Generate temporal-summary tables and figures

Compute annual median vegetation-index composites, where available
Compute a scene-to-scene NDVI difference layer as a phenology-sensitive diagnostic
Generate spatial vegetation-index summary maps

Extract threshold-labelled samples from the selected reference scene
Train the diagnostic Random Forest model
Quantify internal spectral-rule agreement using OA, κ, PA, and UA
Generate the diagnostic spectral-classification layer and internal assessment outputs

Assess spectral separability using Jeffries–Matusita distances
Compare the diagnostic spectral-classification layer with ESA WorldCover after class harmonisation
Sample the diagnostic spectral-classification layer at the Google Earth reference points
Compute external-agreement metrics (OA, κ, PA, UA, and F1)

Compare the proposed workflow with existing Sentinel-2 processing platforms using reproducibility, storage, and local-execution criteria
Compile the final diagnostic assessment and reproducibility outputs
return *Vegetation-index products, temporal summaries, diagnostic classification outputs, external-agreement metrics, figures, and reproducibility report*

The internal band paths follow the Sentinel-2 SAFE naming convention: `IMG_DATA/R{res}/{tile}_{date}_{band}_{res}.JP2`. Only the bands required for the spectral indices and the Scene Classification Layer were opened for each scene. This design minimises intermediate storage by avoiding temporary extraction of the complete SAFE archive.

2.5. SCL-Based Valid-Pixel Masking

The Sentinel-2 L2A Scene Classification Layer (SCL) was used to define a valid-pixel mask prior to vegetation-index computation (Table 2). Because the objective was to preserve usable surface observations in partially clouded tropical scenes, only pixels classified as vegetation (4), bare soil or non-vegetated surface (5), water (6), and unclassified (7) were retained. All remaining SCL classes were excluded from index computation, including no-data pixels, saturated or defective pixels, dark areas, cloud shadows, medium- and high-probability clouds, thin cirrus, and snow or ice [4].

The native 20 m SCL raster was resampled to the 10 m analysis grid using nearest-neighbour interpolation before masking. This preserves the categorical nature of the SCL labels and avoids creating interpolated class values during the resolution harmonisation step.

Table 2. Sentinel-2 Level-2A Scene Classification Layer (SCL) class definitions and treatment in the valid-pixel masking step.

Class Value	Description	Retained?	Rationale
0	No data/outside swath	No	Missing data
1	Saturated or defective pixel	No	Sensor artefact
2	Dark area pixels	No	Potential shadow or low-quality surface
3	Cloud shadow	No	Contaminated
4	Vegetation	Yes	Valid vegetated surface
5	Bare soil/non-vegetated surface	Yes	Valid land surface
6	Water bodies	Yes	Valid surface
7	Unclassified	Yes	Not explicitly flagged as cloud or shadow
8	Cloud, medium probability	No	Cloud contamination
9	Cloud, high probability	No	Cloud contamination
10	Thin cirrus	No	Atmospheric contamination
11	Snow or ice	No	Outside target surface conditions

2.6. Spectral Vegetation Indices

BOA reflectance values ρ_λ were obtained from integer digital numbers (DNs) by [20]:

$$\rho_\lambda = \frac{\mathrm{DN}_\lambda}{10000} \tag{1}$$

2.6.1. Normalised Difference Vegetation Index

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{B08}} - \rho_{\mathrm{B04}}}{\rho_{\mathrm{B08}} + \rho_{\mathrm{B04}}} \tag{2}$$

NDVI uses B08 (Near-Infrared (NIR), 842 nm) and B04 (Red, 665 nm) at 10 m spatial resolution. NDVI $\in [-1, 1]$; higher positive values generally indicate denser and more photosynthetically active vegetation [10].

2.6.2. Enhanced Vegetation Index

$$\mathrm{EVI} = 2.5 \cdot \frac{\rho_{\mathrm{B08}} - \rho_{\mathrm{B04}}}{\rho_{\mathrm{B08}} + 6\,\rho_{\mathrm{B04}} - 7.5\,\rho_{\mathrm{B02}} + 1} \tag{3}$$

EVI incorporates B02 (Blue, 490 nm) together with B08 and B04 at 10 m resolution. The coefficients (2.5, 6, and 7.5) correspond to the MODIS-calibrated constants proposed by Huete et al. [11] and later adapted for Sentinel-2 applications [21]. EVI is less sensitive than NDVI to canopy-background effects and residual atmospheric influences in high-biomass vegetation.

2.6.3. Soil-Adjusted Vegetation Index

$$\mathrm{SAVI} = (1 + L)\frac{\rho_{\mathrm{B08}} - \rho_{\mathrm{B04}}}{\rho_{\mathrm{B08}} + \rho_{\mathrm{B04}} + L}, \quad L = 0.5 \tag{4}$$

SAVI uses the same spectral configuration as NDVI while introducing the soil-brightness correction factor L proposed by Huete [12]. Following Huete [12] and Gilabert et al. [22], $L = 0.5$ was adopted as an empirical adjustment for intermediate vegetation-cover conditions.

2.6.4. Normalised Difference Moisture Index

$$\mathrm{NDMI} = \frac{\rho_{\mathrm{B8A}} - \rho_{\mathrm{B11}}}{\rho_{\mathrm{B8A}} + \rho_{\mathrm{B11}}} \tag{5}$$

NDMI uses Sentinel-2 narrow NIR B8A (865 nm) and Shortwave Infrared (SWIR)-1 B11 (1610 nm) reflectance at native 20 m resolution. The index follows the NIR–SWIR moisture-sensitive formulation originally proposed by Gao [13] and later formalised for forest-condition monitoring by Wilson and Sader [14]. Positive NDMI values are generally associated with higher canopy or surface moisture, whereas lower or negative values are commonly linked to drier vegetation, bare surfaces, or built-up areas [14,23].

Because NDVI, EVI, and SAVI are derived from 10 m Sentinel-2 bands, they were exported on the 10 m grid. In contrast, NDMI was exported at its native 20 m resolution because it uses B8A and B11. The SCL mask was therefore resampled from 20 m to 10 m only for masking the 10 m indices, whereas NDMI was masked using the native 20 m SCL grid.

2.7. Multitemporal Statistical Analysis

For each scene, zonal statistics (mean, median, standard deviation, P10, P25, P75, and P90) were computed over all SCL-valid pixels. Scene-level temporal trends were first explored by Ordinary Least Squares (OLS) regression of scene-mean index values against acquisition date (decimal year), with trend significance evaluated by a two-tailed t-test ($\alpha = 0.05$).

Seasonal consistency was assessed using Pearson correlation (r) between monthly NDVI means and the IDEAM bimodal precipitation climatology [19] for the Eje Cafetero, with significance evaluated at $\alpha = 0.05$. To account for potential non-normality, irregular temporal sampling, and sensitivity to outliers, these parametric analyses were complemented by the non-parametric Mann–Kendall trend test and Sen's slope estimation described in Section 2.13.

Given the sparse and irregular scene inventory, all temporal statistics were interpreted as diagnostic summaries of vegetation-index behaviour, not as evidence of structural land-cover stability or ecological change.

2.8. Diagnostic Threshold-Labelled Spectral Classification

A diagnostic spectral-classification procedure was included to assess how threshold-derived spectral labels behave within the Sentinel-2-only vegetation-index feature space. The objective was not to produce an independent operational land-cover map, but to

generate a reproducible diagnostic layer for evaluating the separability and limitations of index-based spectral classes.

The procedure used the selected low-cloud reference scene identified in the scene inventory. Six spectral classes were defined using land-cover terminology adapted to the Colombian Eje Cafetero context (Table 3). Class labels were assigned using transparent threshold rules informed by published vegetation-index ranges for water, bare surfaces, and tropical vegetation [10–13,15].

For this diagnostic procedure, NDVI, EVI, SAVI, and NDMI were used as the spectral feature space. Because these indices were generated at different native spatial resolutions, the index rasters were loaded as decimated arrays using average resampling and then cropped to their common minimum shape before stacking. Consequently, the resulting RF output was interpreted as a diagnostic spectral-classification layer rather than as a native-resolution thematic land-cover product.

Training and internal hold-out samples were extracted from pixels satisfying the threshold rules. A stratified split was used to separate the threshold-labelled samples into training and internal hold-out subsets. A Random Forest (RF) model [24] was then fitted to the four-dimensional feature vector $\mathbf{x} = [\text{NDVI, EVI, SAVI, NDMI}]$. The RF configuration used for this diagnostic procedure is summarised in Table 4.

Table 3. Threshold-labelled spectral classes used for the diagnostic Random Forest experiment. Thresholds were informed by published vegetation-index ranges for water, bare surfaces, and tropical vegetation [10–13,15]. SAVI was included in the RF feature vector but was not used to define the threshold labels. The em dash (—) indicates that the corresponding index was not used in the threshold rule for that class.

Class	Description	NDVI	EVI	NDMI	Sample *n*
1	Water	<0.05	—	>0.10	800
2	Bare soil/Urban	<0.15	—	<0.00	800
3	Grassland	$[0.15, 0.35)$	—	—	800
4	Crops	$[0.35, 0.55)$	<0.40	—	800
5	Secondary forest	$[0.45, 0.65)$	≥ 0.30	—	800
6	Dense forest	≥ 0.65	≥ 0.40	—	800

Total threshold-labelled samples: 4800 (70%/30% stratified split).

Table 4. Random Forest configuration used in the diagnostic threshold-labelled spectral-classification experiment. Parameters not listed here follow the default `scikit-learn` configuration. CV = cross-validation.

Parameter	Value	Purpose
`n_estimators`	200	Number of trees in the ensemble
`max_depth`	12	Limits tree depth
`criterion`	`gini`	Split criterion and Mean Decrease in Impurity (MDI) importance
`max_features`	`sqrt`	Random feature subset at each split
`bootstrap`	`True`	Bagging with replacement
`n_jobs`	-1	Parallel execution
Internal CV	5 folds	Stability of spectral-rule reproduction

Internal agreement was quantified on the threshold-labelled hold-out subset using Overall Agreement, Cohen's Kappa (κ), Producer's Accuracy (PA = recall), User's Accuracy (UA = precision), and F1-score per class. Stratified cross-validation was also used to assess the stability of spectral-rule reproduction across folds. Because both the training labels and the internal hold-out labels were derived from the same threshold rules,

these metrics were interpreted as internal spectral-rule agreement rather than independent thematic accuracy.

2.9. Class Separability: Jeffries–Matusita Distance

Pairwise spectral separability was assessed using the Jeffries–Matusita (JM) distance in the spectral-index feature space. The JM distance is a remote-sensing separability criterion derived from the Bhattacharyya distance between class distributions and bounded within the interval $[0, 2]$ [9]. Values closer to 2 indicate stronger separation between threshold-labelled spectral classes.

In this study, JM distances were used as diagnostic indicators of separability within the index-based feature space. They were not interpreted as evidence of independent thematic separability, because the class samples were derived from the same spectral rules used to define the diagnostic classes.

2.10. Multi-Level Diagnostic Assessment

A multi-level diagnostic assessment was applied to evaluate the internal coherence and external agreement of the proposed workflow. The assessment was designed to identify the strengths and limitations of the Sentinel-2-only, index-based configuration rather than to claim operational thematic validation.

The diagnostic checks included: spectral coherence of threshold-labelled classes with published index ranges; seasonal consistency between NDVI summaries and the IDEAM bimodal precipitation climatology; cross-sensor consistency between Sentinel-2 acquisitions from different platforms; JM-based spectral separability among threshold-labelled classes; and internal RF spectral-rule agreement on threshold-labelled samples.

Because several diagnostic levels rely on threshold-derived labels or scene-level summaries, their results were interpreted as methodological evidence about workflow behaviour, not as independent validation of land-cover classes.

2.11. External Agreement Assessment Against ESA WorldCover 10 m 2020

To assess external agreement with an independent reference product, the diagnostic RF spectral-classification layer was compared with ESA WorldCover 10 m 2020 (version 1.0) [25]. ESA WorldCover is a global land-cover product derived from Sentinel-1 SAR backscatter and Sentinel-2 multitemporal composites using a hierarchical deep-learning classification approach [25]. This comparison was used to diagnose the thematic limitations of the threshold-labelled Sentinel-2-only spectral-index configuration.

ESA WorldCover tile N03W078 (3°N–6°N, 78°W–75°W) was used for the comparison. The tile covers most of MGRS tile T18NVL; the remaining eastern portion outside tile N03W078 was excluded from this agreement assessment.

Both datasets were remapped to a common five-class scheme (Table 5) to account for differences in class nomenclature and thematic hierarchy. In particular, ESA WorldCover does not distinguish dense forest from secondary forest; therefore, both diagnostic RF forest classes were merged into a single forest class for this comparison.

The ESA WorldCover raster was clipped to the study-area extent and reprojected to the diagnostic RF grid using nearest-neighbour resampling. Pixel-wise comparison was then performed after excluding no-data and unmapped classes. To reduce memory requirements, a stratified random sample of up to 50,000 pixels per reference class was drawn from the valid comparison pixels. External agreement was summarised using Overall Agreement, Cohen's κ, and per-class PA, UA, and F1-score.

Table 5. Five-class common scheme used for external validation against ESA WorldCover 10 m 2020. RF classes 5 and 6 are merged into a single forest class. WorldCover classes not listed (Mangroves, Moss/Lichen, Permanent snow) cover $< 0.1\%$ of tile T18NVL and are excluded.

Common Class	RF Class(es)	WC Code(s)	WorldCover Label(s)
1: Water	Water	80, 90	Permanent water, Herbaceous wetland
2: Bare/Urban	Bare soil/Urban	50, 60, 100	Built-up, Bare/sparse veg., Moss/Lichen
3: Grassland	Grassland	30	Herbaceous vegetation
4: Crops	Crops	40	Cropland
5: Forest	Secondary + Dense forest	10, 20, 95	Tree cover, Shrubland, Mangroves

2.12. Diagnostic Validation Against Independent Google Earth Reference Points

An independent point-based agreement assessment was performed using 210 visually interpreted Google Earth Pro reference points ($n = 35$ per class) from high-resolution imagery close to the reference scene period (January–March 2020). Points were selected in visually unambiguous locations while avoiding cloud shadows, ecotones, and transition zones. Because this strategy underrepresents mixed pixels and the sample size is below the 50–75 points per class commonly recommended for high-precision map accuracy assessment [26], the resulting metrics were interpreted as a diagnostic assessment of error patterns rather than as definitive thematic accuracy estimates.

Reference points were distributed across the main occurrence zones of each class within tile T18NVL, including major rivers, urban centres, flood-plain pastures, agricultural areas, mid-altitude coffee-agroforestry slopes, and high-elevation forested areas. Coordinates stored in WGS 84 were transformed to EPSG:32618 and sampled against the diagnostic RF spectral-classification layer using `pyproj.Transformer` and `rasterio.sample.sample_gen`. Agreement was summarised using Overall Agreement, Cohen's κ, and per-class PA, UA, and F1-score.

2.13. Mann–Kendall Non-Parametric Trend Analysis

Because the scene inventory spans 2017–2025 with sparse and irregular temporal sampling, OLS regression was complemented with the non-parametric Mann–Kendall (MK) trend test [27,28] and Sen's slope estimator [29]. The MK test was applied to the scene-mean time series of NDVI, EVI, SAVI, and NDMI using `pymannkendall` 1.4.

The test was used only to identify detectable monotonic trends in the available vegetation-index time series. Results were not interpreted as evidence of structural vegetation stability or land-cover persistence. The 2022 temporal gap was checked against the Copernicus Data Space Ecosystem catalogue to distinguish archive availability from processing failure.

2.14. Software and Reproducibility

The workflow was implemented in Python and organised as a set of version-controlled scripts covering scene inventory generation, direct SAFE-archive access, SCL-based masking, vegetation-index computation, temporal summaries, diagnostic spectral classification, and agreement assessment. The implementation relies on open-source libraries, including `rasterio` 1.5.0 for geospatial I/O through the GDAL 3.13.1 backend, `numpy 2.4.6` and `pandas` 3.0.3 for array and tabular processing, `scikit-learn` 1.8.0 for the Random Forest diagnostic model and agreement metrics, `scipy` 1.17.1 for statistical analysis, and `matplotlib` 3.10.9 for visualisation.

No proprietary software or cloud-computing platform is required after the Sentinel-2 input archives have been downloaded. Reproducibility is supported by fixed random seeds where stochastic procedures are used, explicit dependency specification in

`requirements.txt`, and a public version-controlled codebase: https://github.com/kevin-ortega2724/satelite.git (accessed on 26 April 2026).

3. Results

3.1. Dataset Characteristics and Valid-Pixel Recovery

The 20-scene dataset spans 2017–2025 with sampling gaps in April–June (Figure 3). Mean scene cloud cover was 35.4 ± 18.8% (range: 5.5–66.8%). After SCL masking, a median of 57.9% of pixels per scene was retained as valid. The five scenes with ≤20% cloud cover retained ≥74.8% valid pixels. Critically, applying a conventional scene-level 30% cloud threshold would have discarded 11 of the 20 scenes (55%), retaining only 9 observations over the 8-year period, a 2.2× reduction in temporal sampling. Per-pixel SCL masking therefore retained substantially more usable observations than scene-level thresholding in this persistently cloudy tropical region.

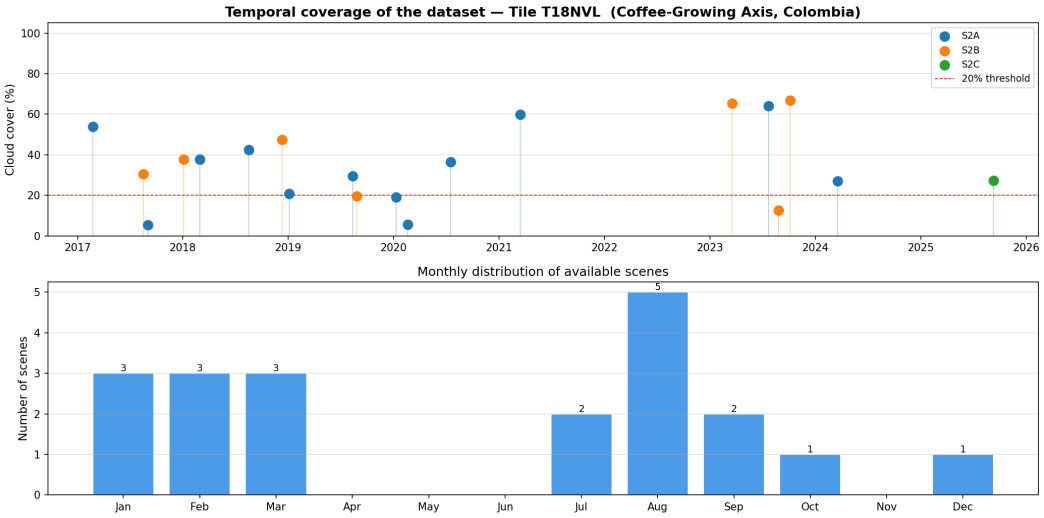


Figure 3. Temporal distribution and scene-level cloud cover of the Sentinel-2 L2A scene inventory used for workflow assessment (2017–2025). (**Upper**) Scene-level cloud cover (%) by satellite platform (S2A, S2B, and S2C). The dashed red line indicates a 20% cloud-cover reference threshold. (**Lower**) Monthly distribution of selected scenes. Months without selected scenes indicate gaps in the processed scene inventory.

3.2. Sentinel-2 Imagery: Multi-Date Visual Assessment

Figure 4 shows false-colour composites and NDVI maps for three contrasting dates, illustrating the effect of season and cloud cover on data availability and surface reflectance.

The contrast between the wet-season (1 September 2017) and dry-season (18 February 2020) composites is clearly visible in the false-colour channel: denser red tones and more uniform canopy coverage in September versus more variable reflectance in February, consistent with the seasonal precipitation pattern of the Eje Cafetero.

3.3. Spectral Index Time Series (2017–2025)

Table 6 presents complete time-series statistics for all four indices. Grand mean NDVI was 0.472 ± 0.018 (range: 0.425–0.516), consistent with dense heterogeneous vegetation cover. The lowest mean NDVI (0.425, 22 July 2023, cloud: 64.1%, valid pixels: 33.7%) likely reflects residual cloud-edge contamination not fully removed by the SCL, a known limitation at cloud boundaries. Temporal trajectories for all indices remained broadly stable across the study period, with detailed non-parametric trend analysis presented in Section 3.4.

The multitemporal vegetation-index time series are shown in Figure 5.

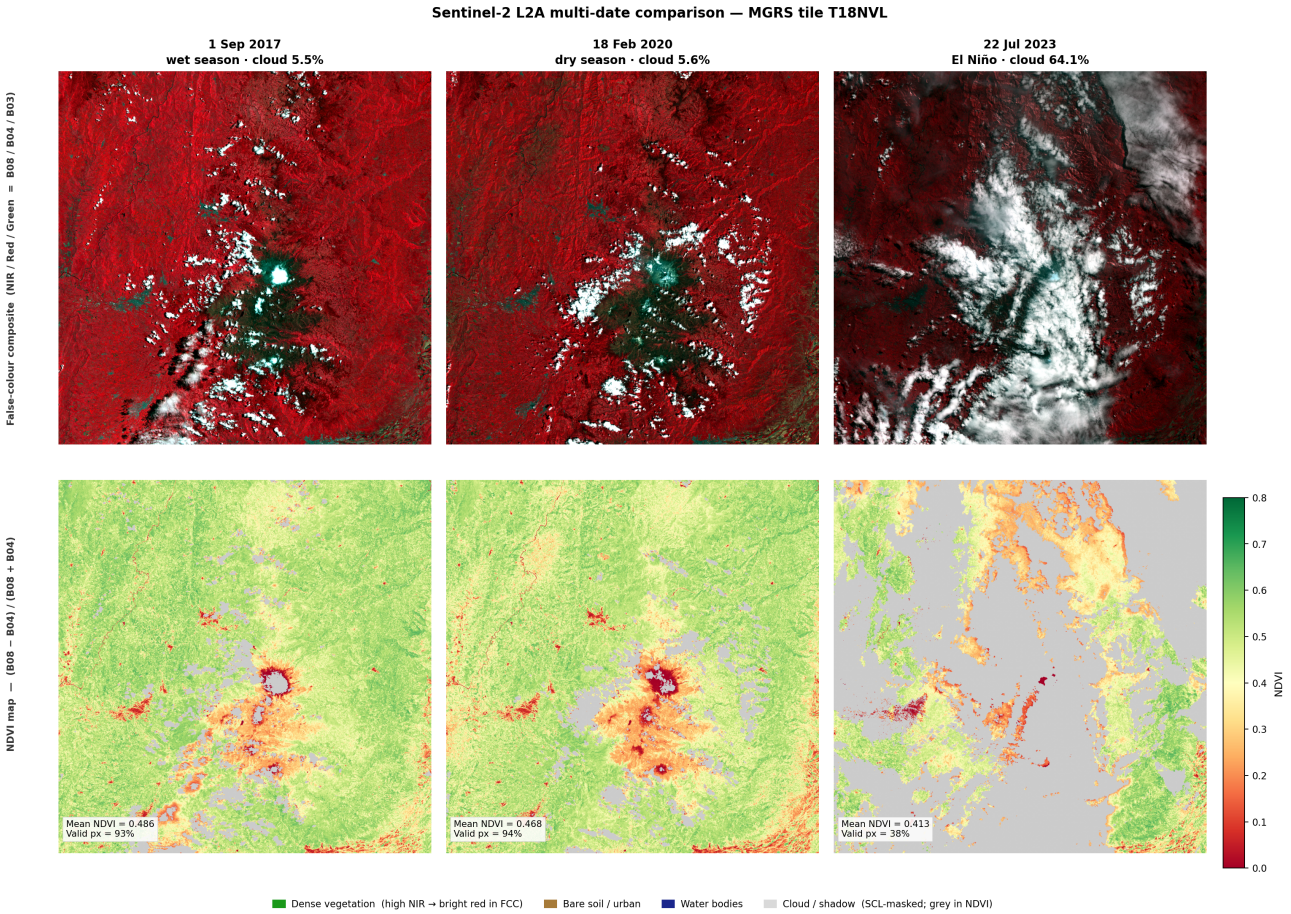


Figure 4. Multi-date Sentinel-2 L2A imagery: false-colour composite (top row, NIR/Red/Green = B08/B04/B03) and NDVI map (bottom row) for three key dates. Left column (1 September 2017, cloud 5.5%): wet-season scene with high vegetation activity, NDVI mean 0.489; vegetation appears intense red in the FCC. Centre column (18 February 2020, cloud 5.6%): dry-season reference scene with highest valid-pixel fraction (92.0%), NDVI mean 0.469; note reduced vigour in agricultural areas compared to the wet season. Right column (22 July 2023, cloud 64.1%): high-cloud scene during El Niño 2023–2024; SCL masking isolates valid pixels (33.7%) but cloud-edge contamination produces the lowest NDVI mean (0.425) in the series. Grey pixels: SCL-masked. Scene-mean NDVI and valid-pixel statistics annotated on each panel are computed on the ×8 overview raster; full-resolution values are reported in Table 1.

Table 6. Summary statistics of spectral vegetation index time series (2017–2025, MGRS tile T18NVL). Statistics are computed from scene-level means of SCL-masked valid pixels. NDVI, EVI, and SAVI were computed at 10 m; NDMI at 20 m. Trend is OLS slope (index units yr^{-1}); significance tested with a two-tailed t-test. ^{ns} $p > 0.05$.

Index	n	Mean	Median	SD	Min	Max	Trend (yr^{-1})	r^2	p
NDVI	17	0.472	0.475	0.018	0.425	0.516	+0.0005	0.004	0.802 ^{ns}
EVI	17	0.573	0.582	0.030	0.510	0.626	+0.0023	0.037	0.461 ^{ns}
SAVI	17	0.403	0.406	0.013	0.372	0.424	−0.0003	0.003	0.837 ^{ns}
NDMI	17	0.195	0.196	0.014	0.163	0.227	+0.0008	0.020	0.586 ^{ns}

Resolution: NDVI, EVI, and SAVI at 10 m (10,980 × 10,980 px); NDMI at 20 m (5490 × 5490 px). Three scenes (27 August 2019, 26 August 2023, 5 October 2023) could not be processed under the 16 GB RAM constraint of the development workstation and are uniformly excluded from all four indices, yielding $n = 17$ for NDVI, EVI, SAVI, and NDMI alike. Unlike an earlier processing iteration, no index required additional scene-specific exclusions beyond this common RAM-related constraint.

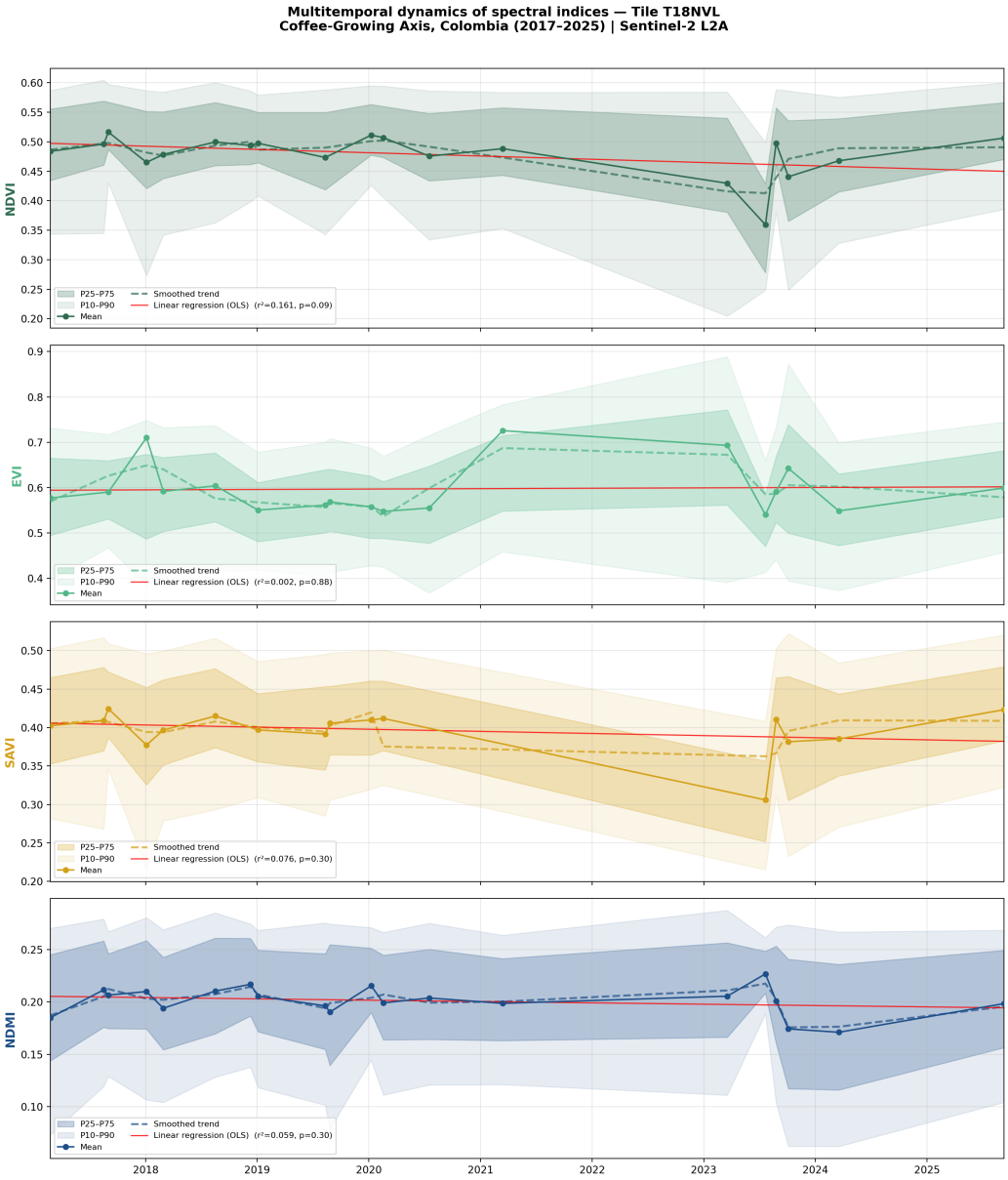


Figure 5. Multitemporal spectral vegetation index time series (2017–2025). Each panel shows scene-mean values (filled circles), interquartile range P25–P75 (dark shading), P10–P90 envelope (light shading), Savitzky–Golay smooth (dashed grey line), and OLS regression line with r^2 and p-value annotation. None of the four indices showed a statistically significant inter-annual trend ($p > 0.05$), indicating no detectable monotonic trend in scene-mean vegetation-index values.

The EVI scene-mean (0.573) exceeds NDVI (0.472). This is physically coherent: after SCL masking, retained pixels are predominantly dense vegetation (SCL class 4, high NIR, low Red reflectance). For observed dense-forest BOA values in this tile ($\rho_{\mathrm{NIR}} \approx 0.45$, $\rho_{\mathrm{Red}} \approx 0.08$, $\rho_{\mathrm{Blue}} \approx 0.04$), Equation (3) yields EVI ≈ 0.56, consistent with the observed range [11].

3.4. Mann–Kendall Trend Analysis and 2022 Gap

Table 7 summarises the Mann–Kendall trend-analysis results for all four spectral indices. No index showed a statistically significant monotonic trend ($p > 0.05$). NDVI yielded $\tau = -0.0074$ ($p = 1.000$), EVI $\tau = +0.0368$ ($p = 0.869$), SAVI $\tau = -0.0588$ ($p = 0.773$), and NDMI $\tau = -0.0074$ ($p = 1.000$). Sen's slopes were below 0.0003 index units yr^{-1} in all cases, indicating no detectable monotonic vegetation-index trend in the

available scene-level time series despite irregular temporal sampling and the confirmed 2022 archive gap.

The Mann–Kendall trend analysis is shown in Figure 6.

Table 7. Mann–Kendall trend test results for the four spectral vegetation indices (2017–2025; $n = 17$ scenes per index, uniformly). τ = Kendall's tau; p = two-tailed significance; Sen's slope in index units yr^{-1}. No index shows a statistically significant monotonic trend ($p > 0.05$).

Index	n	τ	p	Sen's Slope	Trend
NDVI	17	−0.0074	1.0000	0.000000	no trend
EVI	17	0.0368	0.8690	0.000206	no trend
SAVI	17	−0.0588	0.7727	−0.000167	no trend
NDMI	17	−0.0074	1.0000	−0.000056	no trend

Sen's slope units: index value yr^{-1}.

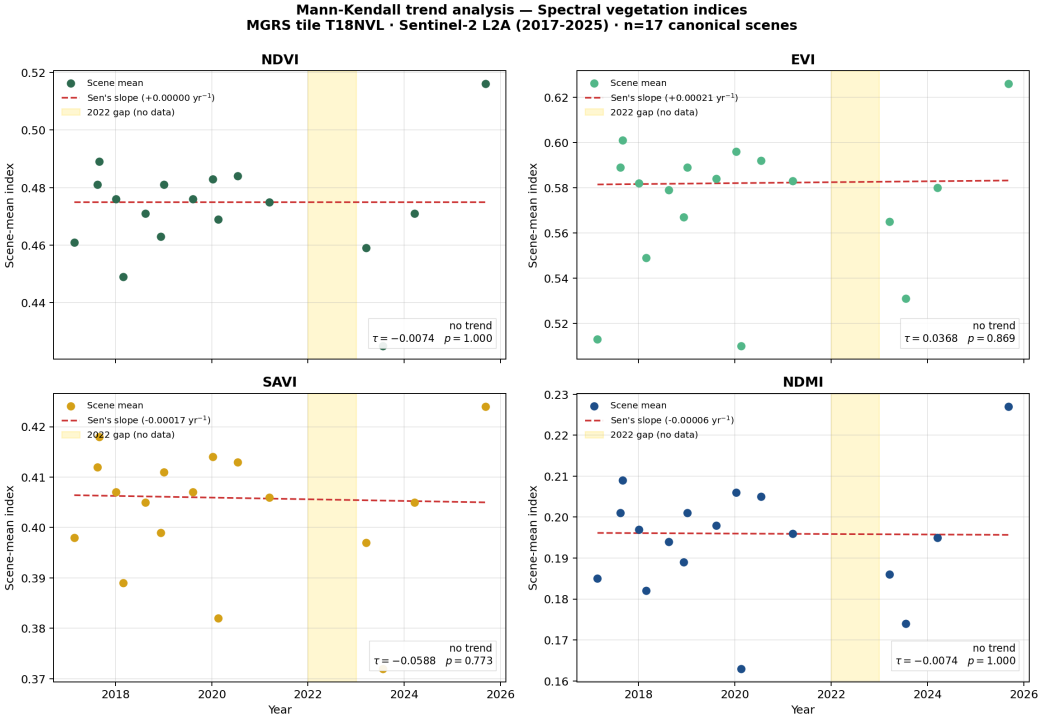


Figure 6. Mann–Kendall non-parametric trend analysis for NDVI, EVI, SAVI, and NDMI (2017–2025; $n = 17$ scenes per index). Filled circles: scene mean ±95 % CI. Grey diamonds: LOESS residual anomaly, deseasonalised. Coloured line: LOESS smooth trend (frac = 0.75). Dashed red line: Sen's non-parametric slope ($p > 0.05$ in all panels). Yellow band: 2022 archive gap, confirmed as zero products in the Copernicus Data Space Ecosystem archive for tile T18NVL.

3.5. Seasonal Dynamics and Precipitation Coherence

The monthly NDVI boxplot (Figure 7A) shows elevated medians in September–November relative to the January–March and June–August windows, consistent with the bimodal wet-season climatology of the Eje Cafetero [19]. Pearson correlation between monthly NDVI means and the IDEAM precipitation climatology yielded $r = 0.144$ ($p = 0.556$), indicating no statistically significant linear relationship. This result is likely influenced by the irregular temporal sampling and the absence of valid observations during part of the bimodal seasonal cycle (Figure 3).

The inter-annual panel (Figure 7B) shows annual median NDVI values fluctuating within a narrow range (0.464–0.493; Table 6), with a marginal decline after 2022 that may reflect mild vegetative stress during the 2023–2024 El Niño event. However, no

statistically significant inter-annual trend was detected for any index (Section 3.4), and the 2025 composite (mean NDVI ≈ 0.493) recovered to the upper end of the observed range, consistent with higher vegetation-index values after the 2023–2024 El Niño period rather than with evidence of persistent index decline.

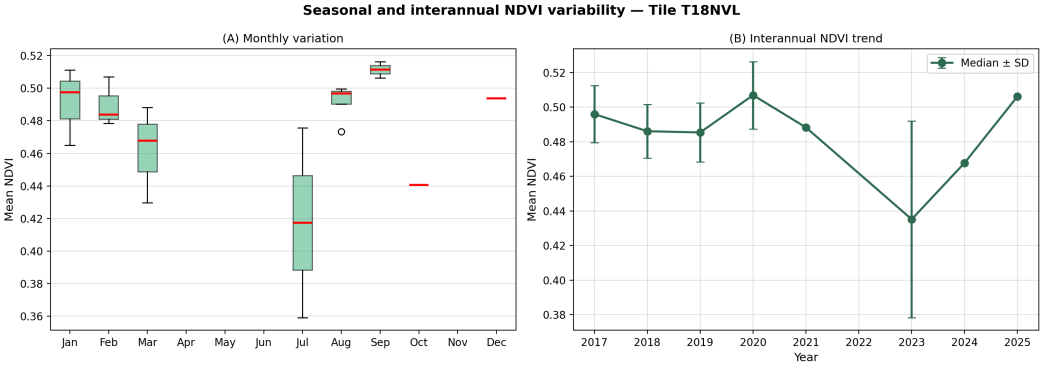


Figure 7. Seasonal and inter-annual NDVI variability. (**A**) Monthly NDVI boxplot; elevated medians in September–November correspond to the main wet season of the Eje Cafetero bimodal climate. (**B**) Annual median NDVI (±1 SD); marginal decline after 2022 may reflect vegetative stress during the 2023–2024 El Niño.

3.6. Spatial Distribution of Spectral Indices

Figure 8 shows the spatial NDVI distribution for the reference scene (18 February 2020) together with the NDVI value histogram stratified by threshold-defined spectral ranges.

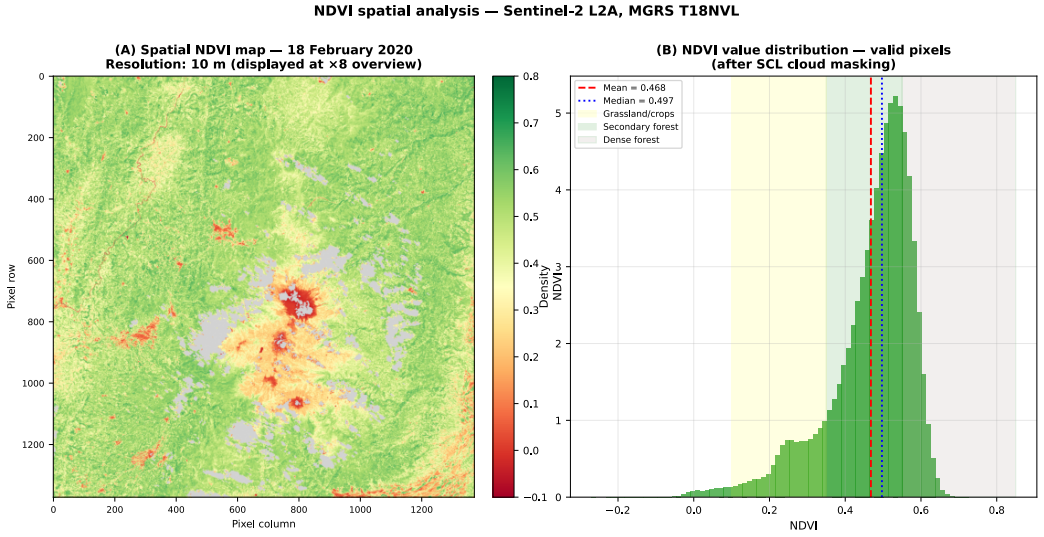


Figure 8. Spatial NDVI distribution and value histogram (18 February 2020; cloud cover: 5.6%). (**A**) NDVI overview map at ×8 resolution (effective pixel size 80 m; grey = SCL-masked pixels). Dense forest (high NDVI, deep green) concentrates at mid-to-high altitudes; agricultural and urban areas (lower NDVI) are visible in the Cauca valley corridor. (**B**) Frequency distribution of NDVI for SCL-valid pixels; shaded bands mark the spectral thresholds used to define threshold-labelled diagnostic classes (Table 3). Mean NDVI = 0.469; median = 0.472.

Figure 9 presents all four indices simultaneously for the reference scene, enabling direct spatial comparison of complementary vegetation signals.

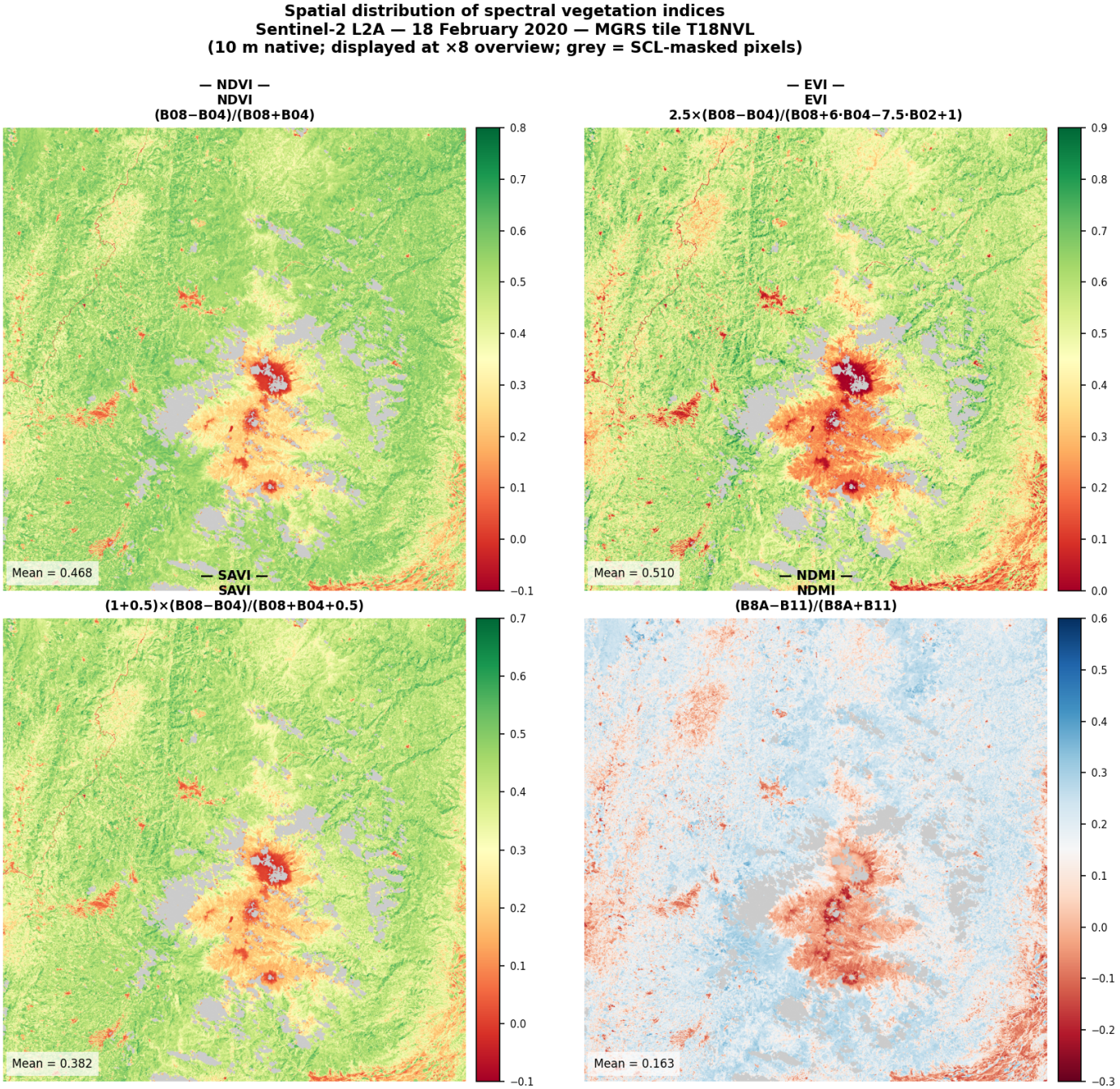


Figure 9. Spatial distribution of all four spectral vegetation indices (18 February 2020; ×8 overview; grey = SCL-masked). NDVI, EVI, and SAVI use the RdYlGn colour scale; NDMI uses RdBu (blue = high moisture content). All panels share the same spatial extent and CRS (EPSG:32618). Scene-mean index values are annotated. The spatial complementarity of the four indices is evident: NDMI provides a moisture-sensitive contrast that is partly distinct from canopy-density indices, while NDVI and EVI track canopy density with different saturation characteristics.

3.7. Inter-Index Correlation

Figure 10 presents the pairwise Pearson correlation matrix for scene-mean values of the four spectral indices ($n = 17$ scene pairs, uniformly, since NDVI, EVI, SAVI, and NDMI are now derived from the same canonical set of processed scenes). The strongest association was observed between NDVI and SAVI ($r = 0.905$), consistent with their closely related mathematical formulations (Equations (2) and (4)). EVI showed moderate-to-strong correlation with NDVI ($r = 0.762$), reflecting their shared sensitivity to canopy density while

retaining different responses under high-biomass and variable atmospheric conditions (Equation (3)) [11]. NDMI exhibited strong correlations with the three optical indices ($r = 0.83$–0.94), most notably with SAVI ($r = 0.939$) and EVI ($r = 0.929$). This elevated association differs from the moderate correlation that might be expected a priori between canopy-moisture and canopy-density metrics, and likely reflects that SCL-valid pixels in tile T18NVL are concentrated in structurally similar, densely vegetated land covers (shade-grown coffee, secondary and dense forest) whose canopy density and moisture status co-vary strongly across the 2017–2025 acquisition dates, rather than genuinely independent spectral behaviour at the scene-mean level. Despite this scene-level collinearity, NDMI's distinct spectral basis (SWIR-1 reflectance) continues to support its inclusion as an additional predictor in the Random Forest classifier (Section 2.8) and in the multidimensional separability analysis (Section 2.9), where—as shown in Section 3.8—its diagnostic value operates at the pixel level, discriminating spectrally similar classes such as Water and Bare soil/Urban that NDVI and EVI alone cannot separate.

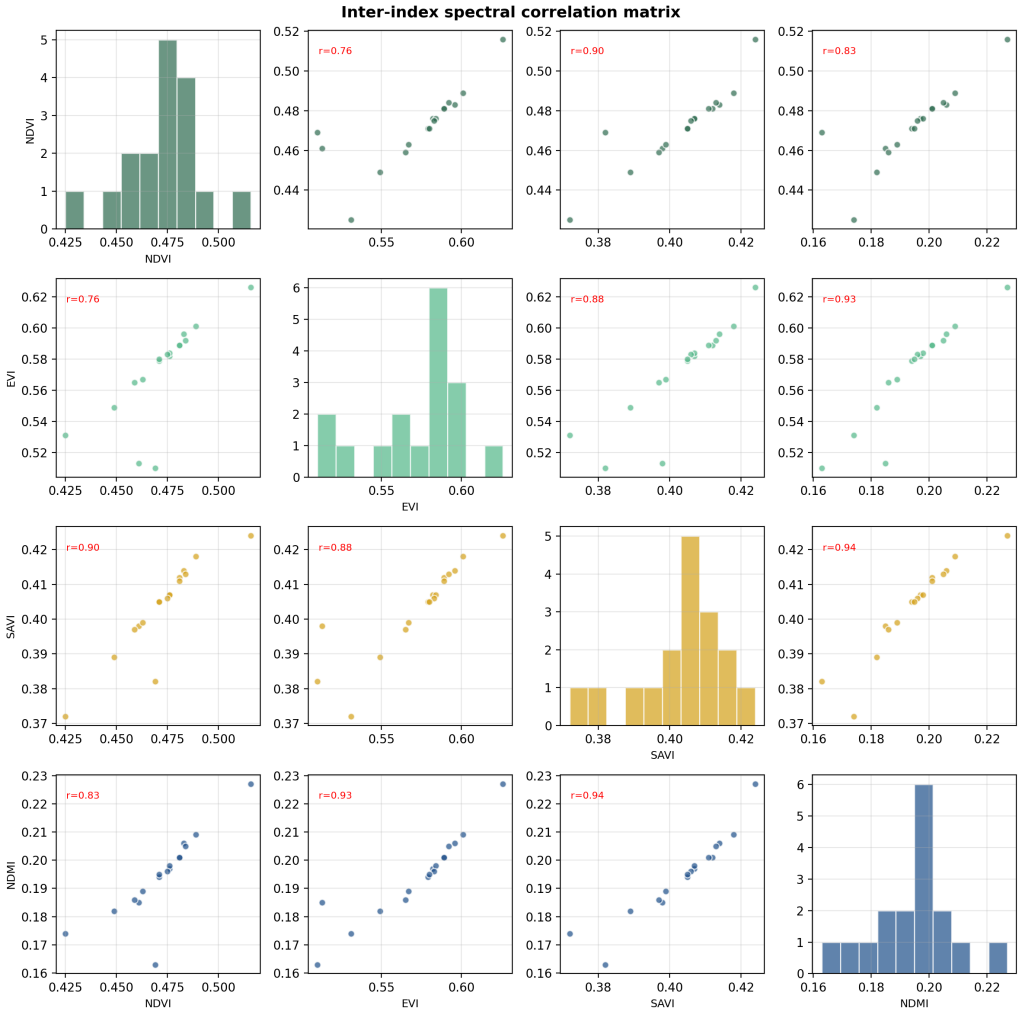


Figure 10. Inter-index correlation matrix for scene-mean values ($n = 17$ scene pairs, uniformly). Diagonal: histogram of scene-mean values per index. Upper triangle: Pearson r (bold text). Lower triangle: scatter plot. NDVI–SAVI: $r = 0.91$ (mathematical co-linearity by construction); EVI–NDVI: $r = 0.76$; NDMI shows strong correlation with the optical indices ($r = 0.83$–0.94) in this canonical dataset, reflecting co-variation of canopy density and moisture across the predominantly forest/agroforestry valid-pixel population rather than independent spectral information at the scene-mean level; its complementary value is instead demonstrated at the pixel level (Section 3.8, Figure 11).

3.8. Class Spectral Signatures

Figure 11 shows the mean ± 1 SD spectral signatures of the six land-cover classes in the four-dimensional index space, derived from the 800 threshold-labelled training samples per class ($n = 4800$ total) extracted from the reference scene (18 February 2020). NDVI and EVI increase monotonically from Water to Dense forest, reflecting the progressive increase in photosynthetically active biomass across the land-cover gradient. Dense forest exhibits the highest mean NDVI ($\geq$0.65) and EVI ($\geq$0.40), consistent with closed-canopy, high-LAI conditions characteristic of Andean cloud forest at mid-to-high altitudes within the tile [11,15]. Grassland and Crops occupy the intermediate NDVI range (0.15–0.55; Table 3), with partially overlapping distributions that constitute the primary source of inter-class confusion, as quantified by the Jeffries–Matusita separability analysis in Section 3.12.

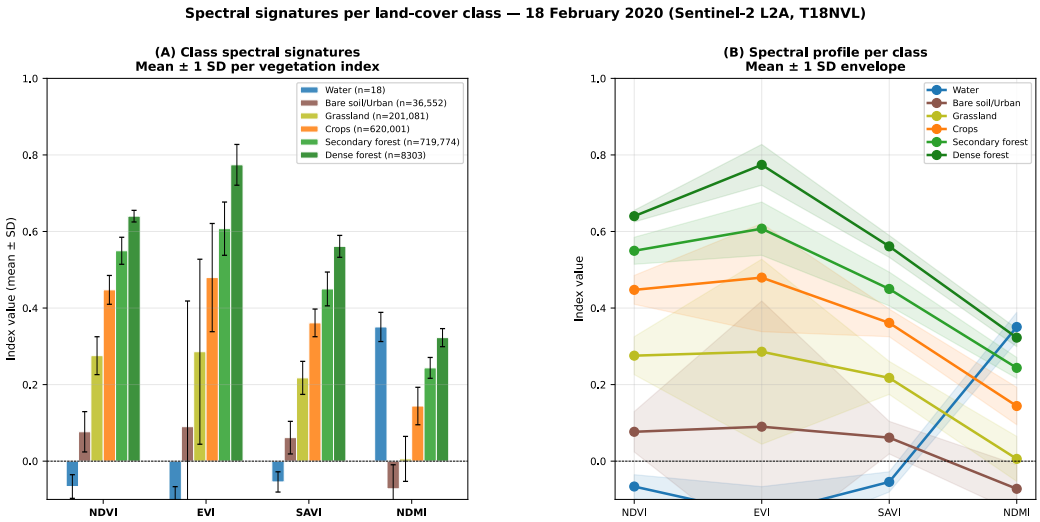


Figure 11. Class spectral signatures derived from the reference scene (18 February 2020; $n = 800$ threshold-labelled samples per class). (**A**) Grouped bar chart: mean ± 1 SD of each index per class; error bars indicate within-class variability. (**B**) Line profile view with ±1 SD envelopes for separability assessment. NDVI and EVI increase monotonically from Water to Dense forest; NDMI discriminates Water (positive) from Bare soil/Urban (negative) where NDVI and SAVI values overlap, demonstrating the added value of incorporating SWIR-based indices in the feature space. The Jeffries–Matusita distance for Crops vs. Secondary forest (JM = 1.716) represents the lowest pairwise separability in the dataset; all remaining class pairs achieve JM > 1.9.

The diagnostic value of NDMI is most evident in discriminating Water from Bare soil/Urban: both classes present low NDVI (Water < 0.05, Bare soil/Urban < 0.15) and low EVI, where optical indices alone are insufficient to separate them reliably. However, Water bodies exhibit positive NDMI values, since liquid water absorbs weakly at SWIR-1 (1610 nm), resulting in high ρ_{B8A} relative to ρ_{B11} (Equation (5)), whereas Bare soil and Urban surfaces yield negative NDMI owing to their high SWIR-1 reflectance from dry mineral and built materials. This contrast is visible in the line-profile view of Figure 11B, where the Water and Bare soil/Urban signatures cross from near-zero NDVI/EVI to strongly divergent NDMI, demonstrating that a three-index feature space (NDVI, EVI, SAVI) would fail to achieve full separability for this class pair without the SWIR channel.

Within-class variability (1 SD envelope) is largest for Secondary forest, which encompasses a broad range of successional stages, from recently abandoned agricultural land with emerging woody cover to mature secondary stands approaching dense-forest canopy structure, all of which fall within the threshold-defined NDVI window $[0.45,\ 0.65)$ (Table 3). This internal heterogeneity is reflected in the slightly reduced Producer's Accuracy for Secondary forest (99.6%, vs. 100.0% for all other classes; Table 8), and is consistent with the

Jeffries–Matusita distance between Crops and Secondary forest (JM = 1.716; Table 9), which represents the lowest pairwise separability among all class pairs evaluated (JM threshold for high separability: >1.9 [9]).

Table 8. Random Forest internal spectral-rule agreement metrics (scene: 18 February 2020; 200 trees; 4 features: NDVI, EVI, SAVI, NDMI). OA = Overall Agreement; κ = Cohen's Kappa. CV = 5-fold cross-validation on the full threshold-labelled sample set. PA = Producer's Accuracy (recall); UA = User's Accuracy (precision). **Note:** training and hold-out samples were both generated by applying the same spectral-threshold rules (Table 3); the resulting metrics therefore quantify *internal spectral-rule agreement* rather than independent thematic accuracy. The near-perfect values (OA = 99.9%; κ = 0.999) confirm that the Random Forest reproduces the threshold-defined labels, but should not be interpreted as a measure of real-world land-cover mapping accuracy without independent reference data.

Class	PA (%)	UA (%)	F1	*n* Test
Water	100.0	100.0	1.00	240
Bare soil/Urban	100.0	100.0	1.00	240
Grassland	100.0	100.0	1.00	240
Crops	100.0	99.6	1.00	240
Secondary forest	99.6	100.0	1.00	240
Dense forest	100.0	100.0	1.00	240
Overall	OA = 99.9%		κ = 0.999	1440
CV (5-fold)	OA = 99.9% ± 0.05%			

SAVI signatures parallel those of NDVI across all classes, as expected from their mathematical relationship (Section 3.7), but with systematically compressed values due to the $L = 0.5$ correction factor. The practical contribution of SAVI to the four-feature classifier is therefore concentrated in transitional cover types (Grassland, Crops) where soil background effects are non-negligible, and where the modified index response to sparse vegetation helps the Random Forest decision boundaries remain stable across the range of soil brightness conditions found within the tile.

3.9. Scene-to-Scene NDVI Difference (2017–2025)

Figure 12 presents the pixel-wise NDVI difference ($\Delta NDVI = NDVI_{2025} - NDVI_{2017}$) computed between the earliest available scene (23 February 2017; cloud: 53.9%; mean NDVI: 0.461) and the most recent scene (9 September 2025; cloud: 27.2%; mean NDVI: 0.516). Pixels were classified into three change categories using thresholds established in the remote sensing literature [16]: positive NDVI difference ($\Delta NDVI > +0.05$; 46.3% of valid pixels), near-zero NDVI difference ($|\Delta NDVI| \leq 0.05$; 39.1%), and negative NDVI difference ($\Delta NDVI < -0.05$; 14.7%).

The dominance of the gain category (46.3%) must be interpreted with caution. The 2017 reference scene was acquired in February (dry season) with 53.9% cloud cover, while the 2025 scene corresponds to early September (onset of the main wet season) with 27.2% cloud cover. This phenological and seasonal asymmetry introduces a systematic positive bias in ΔNDVI: higher canopy water content and greenness in the wet-season acquisition inflate apparent gain relative to an equivalent dry-season baseline. Consequently, these figures should be interpreted as a phenology-sensitive scene-to-scene NDVI difference, not as a measure of structural land-cover change. A seasonally matched pair, or, preferably, a pixel-wise analysis based on annual median composites (Section 3.10), would yield a more conservative and phenologically neutral change estimate. Notwithstanding this limitation, the spatial distribution of loss pixels (14.7%) is broadly consistent with the urban expansion corridors visible around Pereira and Manizales in the annual composites

(Figure 13), suggesting that some negative NDVI-difference areas coincide spatially with urban corridors visible in the annual composites, but this pattern should be interpreted cautiously because the comparison is based on two seasonally asymmetric scenes.

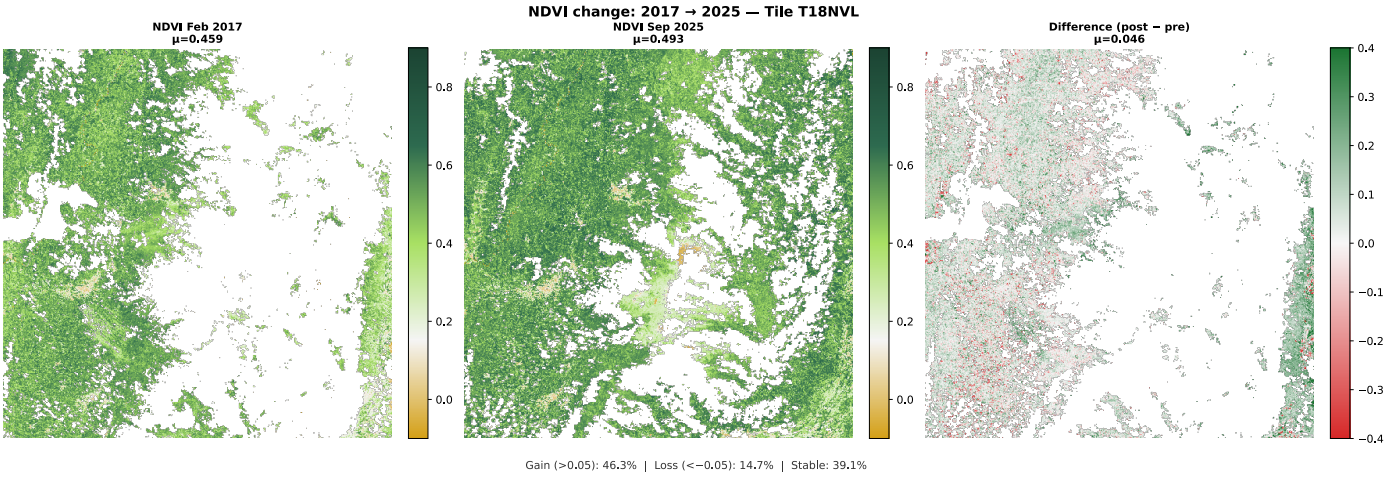


Figure 12. Scene-to-scene NDVI difference map (2017–2025). (**Left**): NDVI for 23 February 2017 (first available scene, cloud 53.9%; mean NDVI: 0.461). (**Centre**): NDVI for 9 September 2025 (most recent scene, cloud 27.2%; mean NDVI: 0.516). (**Right**): pixel-wise difference ($\Delta\text{NDVI} = \text{NDVI}_{2025} - \text{NDVI}_{2017}$); red = negative NDVI difference ($\Delta\text{NDVI} < -0.05$; 14.7% of valid pixels); white/grey = stable ($|\Delta\text{NDVI}| \leq 0.05$; 39.1%); green = positive NDVI difference ($\Delta\text{NDVI} > +0.05$; 46.3%). Caution: the 2017–2025 scene pair is seasonally asymmetric (dry season vs. wet season onset), which introduces a positive bias in the NDVI-difference estimate; see Section 3.9 for discussion.

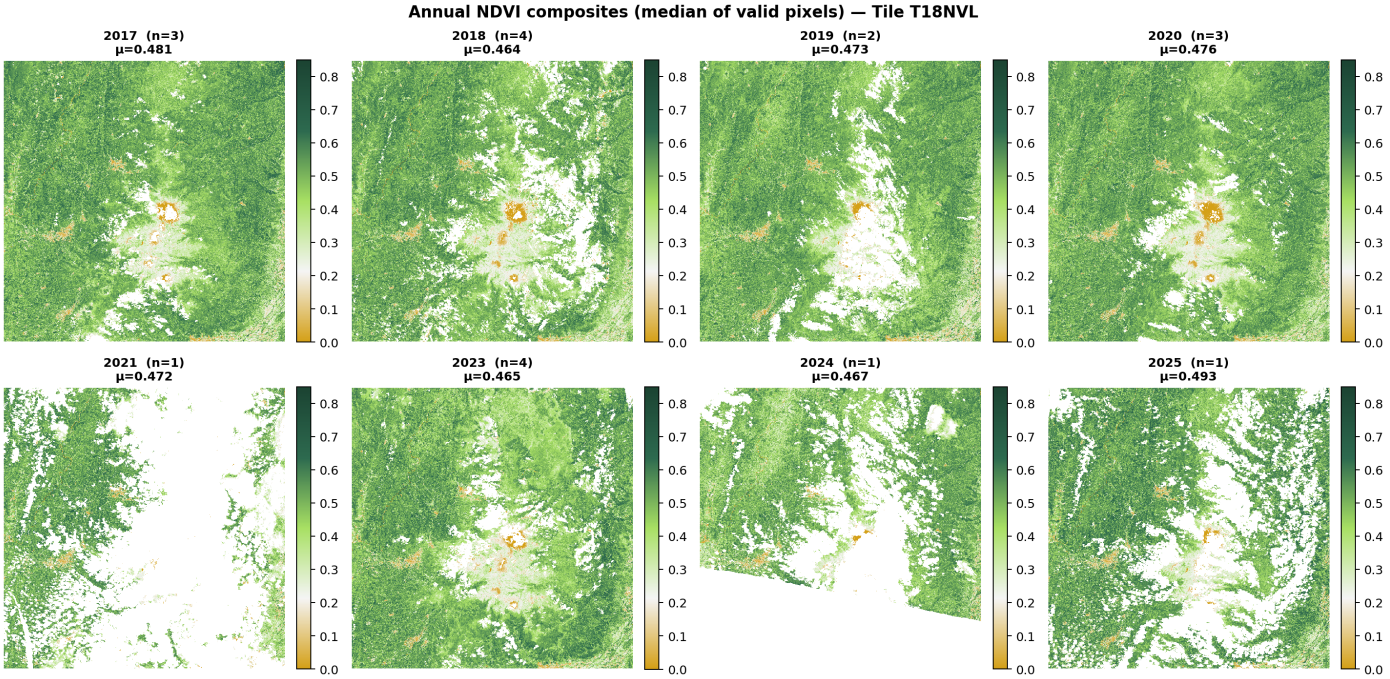


Figure 13. Annual median NDVI composites (2017–2025). Each panel shows the per-pixel median of all SCL-valid observations within the calendar year; scene count n and tile-mean NDVI are annotated. Mean NDVI values: 2017 ≈ 0.481; 2018 ≈ 0.464; 2019 ≈ 0.473; 2020 ≈ 0.476; 2021 ≈ 0.472; 2023 ≈ 0.465; 2024 ≈ 0.467; 2025 ≈ 0.493. The absence of a 2022 composite reflects the lack of archived scenes for this tile in that calendar year. Inter-annual variability in tile-mean NDVI is modest (range ≈ 0.028), consistent with limited variation in annual median NDVI.

3.10. Annual NDVI Composites (2017–2025)

Annual median NDVI composites were computed by taking the per-pixel median of all SCL-valid observations within each calendar year, providing a cloud-reduced spatial summary of vegetation condition that is less sensitive to individual scene anomalies than any single acquisition. Figure 13 presents the resulting eight composites for 2017–2025 (no composite was produced for 2022 owing to the absence of archived scenes for tile T18NVL in that year).

Tile-mean NDVI across annual composites ranged from ≈0.464 (2018) to ≈0.493 (2025), a range of only ≈0.028, consistent with the limited inter-annual variation observed in the scene-level vegetation-index time series (Section 3.3). The 2025 composite (mean NDVI ≈ 0.493) represents the highest annual median in the series, consistent with post-El Niño vegetation recovery following the 2023–2024 drought episode visible as a mild depression in the 2023 composite (mean NDVI ≈ 0.465). The 2017 composite (mean NDVI ≈ 0.481) was derived from three scenes, including one acquisition with 53.9% cloud cover; its slightly elevated within-pixel variability relative to composites derived from four scenes (2018, 2023) reflects this lower valid-pixel density. Spatially, all composites show a consistent pattern: higher NDVI values at mid-to-high altitudes along the Cordillera Central flanks, corresponding to cloud forest and secondary forest cover, and lower values in the Cauca valley corridor, where agricultural and urban land uses predominate.

3.11. Diagnostic Spectral Classification

The Random Forest classifier was applied to the reference scene (18 February 2020; cloud cover: 5.6%; SCL-valid pixels: 92.0%) using the four-dimensional feature vector $\mathbf{x} = [\text{NDVI, EVI, SAVI, NDMI}]$. The hyperparameters and model configuration are described in the Section 2.8. The resulting diagnostic spectral-classification layer (Figure 14) covers the full SCL-valid extent of tile T18NVL.

Table 8 reports the internal spectral-rule agreement metrics. The classifier achieved an internal overall agreement of OA = 99.9% ($\kappa = 0.999$) on the 30% stratified test set ($n_{\text{test}} = 1440$), with F1 = 1.00 for all classes except Secondary forest (F1 = 1.00; PA = 99.6%), whose slight reduction in Producer's Accuracy reflects the within-class heterogeneity of successional stages discussed in Section 3.8. Five-fold cross-validation on the full training set yielded OA = 99.9 ± 0.05%, confirming stability across folds.

As noted in Table 8, these metrics quantify internal spectral consistency rather than independent thematic accuracy, since training and test samples share the same spectral-threshold labelling protocol.

The class area statistics derived from the classification map are as follows: Secondary forest is the dominant class at 81.3% of SCL-valid pixels, followed by Crops (8.6%), Grassland (8.2%), Water (0.9%), Dense forest (0.9%), and Bare soil/Urban (0.2%). The dominance of Secondary forest is consistent with the land-cover mosaic of the Eje Cafetero, which is characterised by extensive coffee agroforestry, shade-grown plantations, and successional woody vegetation on steep slopes [19]. The low mapped proportion of Dense forest (0.9%) warrants interpretation in the context of the reference scene acquisition date: the February 2020 scene corresponds to the dry season, during which even structurally closed-canopy forest can exhibit reduced canopy water content and slightly lower NDVI values (ρ_{NIR} decreases modestly with leaf water stress), causing some pixels that would qualify as dense forest under wetter conditions to fall below the strict classification threshold (NDVI ≥ 0.65 *and* EVI ≥ 0.40; Table 3). This threshold sensitivity is a known limitation of single-date, spectral-threshold-based training sample generation, and is expected to be reduced when multitemporal or wet-season reference scenes are incorporated. The spatial distribution of variable importance (Figure 15b) confirms that NDVI provides the largest contribution to

the Gini impurity reduction across all split nodes, followed by SAVI and EVI with comparable contributions, and NDMI with the smallest but non-negligible role, particularly in separating Water from Bare soil/Urban as discussed in Section 3.8.

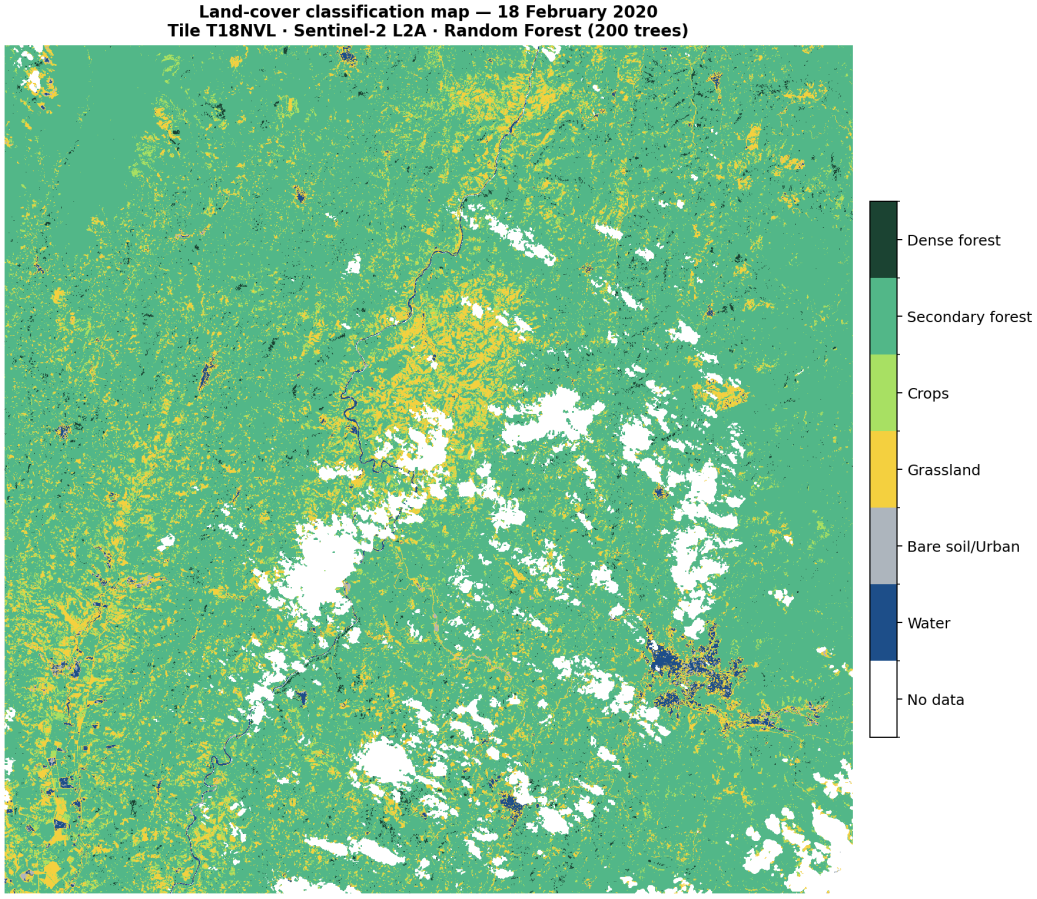
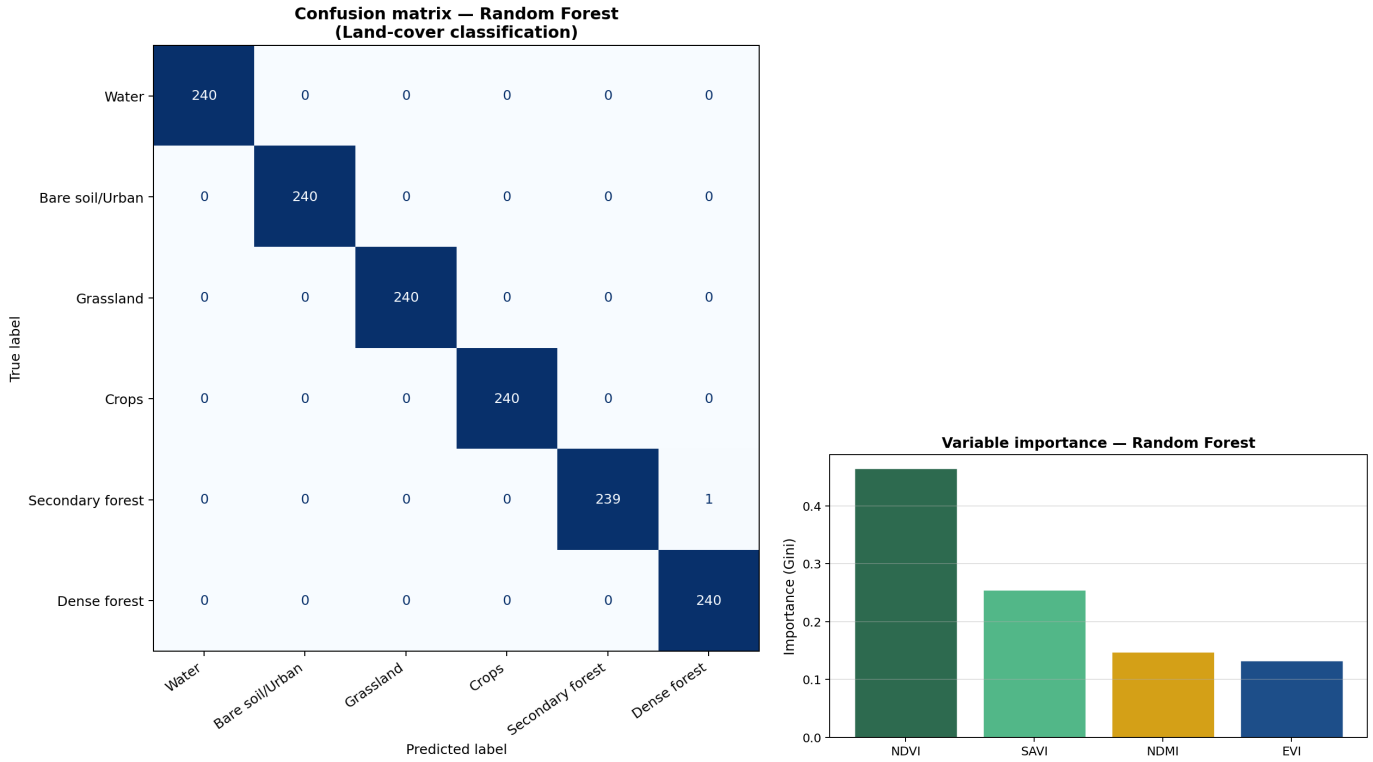


Figure 14. Diagnostic Random Forest spectral-classification layer for 18 February 2020 (×4 overview). Six classes: Water (dark blue, 0.9% of valid pixels), Bare soil/Urban (grey, 0.2%), Grassland (yellow, 8.2%), Crops (light green, 8.6%), Secondary forest (medium green, 81.3%), Dense forest (dark green, 0.9%). The dominance of Secondary forest reflects the extensive coffee agroforestry and successional woody cover characteristic of the Eje Cafetero landscape [19]. The low Dense forest fraction is partly attributable to the dry-season acquisition date, which reduces canopy NDVI below the strict classification threshold (NDVI ≥ 0.65) for some closed-canopy stands; see Section 3.11 for discussion. Internal spectral-rule agreement metrics are reported in Table 8.

3.12. Multi-Level Diagnostic Assessment Results

Table 9 summarises the multi-level diagnostic assessment results. Level L1 showed that 100% of NDVI samples fall within published spectral ranges for all classes, while EVI showed partial saturation in dense forest (71–100%), consistent with known high-LAI behaviour [11]. Level L2 yielded a non-significant precipitation–NDVI correlation ($r = 0.144$, $p = 0.556$) attributable to the April–June sampling gap (Section 3.5). Level L3 cross-sensor consistency was indicated with $|\Delta\text{NDVI}| < 0.015$ for both S2A/S2B pairs. Level L4 identified the Crops–Secondary forest pair as the only class combination below the high-separability threshold (JM $= 1.716 < 1.9$). Level L5 RF cross-validation yielded OA = 99.9% $\pm$ 0.05%, reflecting spectral self-consistency of the threshold-labelled training set.



(**a**) (**b**)

Figure 15. RF classification diagnostics (18 February 2020 reference scene). (**a**) Confusion matrix on the 30% stratified test set ($n_{\text{test}} = 1440$); colour intensity proportional to sample count; perfect diagonal reflects spectral-threshold training (see Table 8 note). (**b**) Variable importance by mean decrease in Gini impurity. NDVI provides the largest impurity reduction; SAVI and EVI contribute comparably; NDMI adds the smallest but non-negligible contribution, particularly for Water vs. Bare soil/Urban discrimination.

Table 9. Summary of the multi-level diagnostic assessment applied to the Sentinel-2 spectral-index workflow. The diagnostic levels were used to characterise spectral coherence, temporal behaviour, cross-sensor consistency, class separability, and internal spectral-rule agreement without field-campaign data. JM = Jeffries–Matusita distance ($\in [0, 2]$; JM > 1.9: high separability). r = Pearson correlation coefficient. ^{ns} = not significant ($p > 0.05$).

Level	Diagnostic Check	Metric	Result
L1	Spectral coherence	% samples within literature NDVI range	100% for NDVI in all classes
		% samples within literature EVI range	71–100% (EVI saturation in dense forest)
		% samples within literature NDMI range	81–100% (Water, Bare soil, Dense forest)
L2	Temporal coherence	Pearson r (NDVI vs. precipitation)	$r = 0.144^{\text{ns}}$ ($p = 0.556$; insufficient monthly coverage, see text)
		Seasonal pattern	Sep–Nov medians visually elevated; formal significance not achievable with current sampling density
L3	Cross-sensor consistency	ΔNDVI (S2A vs. S2B) Pairs within 30-day window	Pair 1: $\Delta = 0.0075$ (15 days; 2017) Pair 2: $\Delta = 0.0085$ (25 days; 2018–2019) Both $\lvert\Delta\text{NDVI}\rvert < 0.015$

Table 9. *Cont.*

Level	Diagnostic Check	Metric	Result
L4	Class separability	JM (Grassland vs. Dense forest) JM (Crops vs. Dense forest) JM (Crops vs. Secondary forest)	JM = 2.000 (high separability) JM = 1.944 (high separability) JM = 1.716 (medium separability)
L5	Internal spectral-rule agreement	Hold-out OA; κ 5-fold CV OA	OA = 99.9%; κ = 0.999 † OA = 99.9% ± 0.05%

† Near-perfect agreement reflects spectral-threshold training; see text. ns Insufficient monthly coverage (gaps Apr–Jun) limits NDVI–precipitation correlation.

3.13. External Agreement Assessment Against ESA WorldCover 2020

Table 10 presents the per-class and overall agreement metrics between the diagnostic RF spectral-classification layer and ESA WorldCover 10 m (five-class common scheme; n = 250,000 stratified pixels). Overall Accuracy was 20.2% (κ = 0.003), substantially below the internal cross-validated OA reported in Table 8. Forest was the only class with PA > 50% (PA = 81.9%), while all non-forest classes showed PA ≤ 9.5% and UA values of 11.7–43.2%.

Figure 16 summarises the external agreement results.

Table 10. External agreement assessment between the diagnostic RF spectral-classification layer and ESA WorldCover 10 m (2020, v1.0), tile N03W078, clipped to MGRS tile T18NVL. Both datasets were remapped to a five-class common scheme (WorldCover does not distinguish Dense forest from Secondary forest). OA = Overall Agreement; PA = Producer's Accuracy (recall); UA = User's Accuracy (precision). Metrics were computed from n = 250,000 stratified sampled pixels drawn from 101,632,411 valid comparison pixels.

Class	PA (%)	UA (%)	F1	*n* Sampled
Water	0.3	11.7	0.007	50,000
Bare soil/Urban	1.0	43.2	0.020	50,000
Grassland	9.5	23.8	0.135	50,000
Crops	8.3	18.6	0.115	50,000
Forest	81.9	20.0	0.321	50,000
Overall	**OA = 20.2%**		**κ = 0.003**	250,000

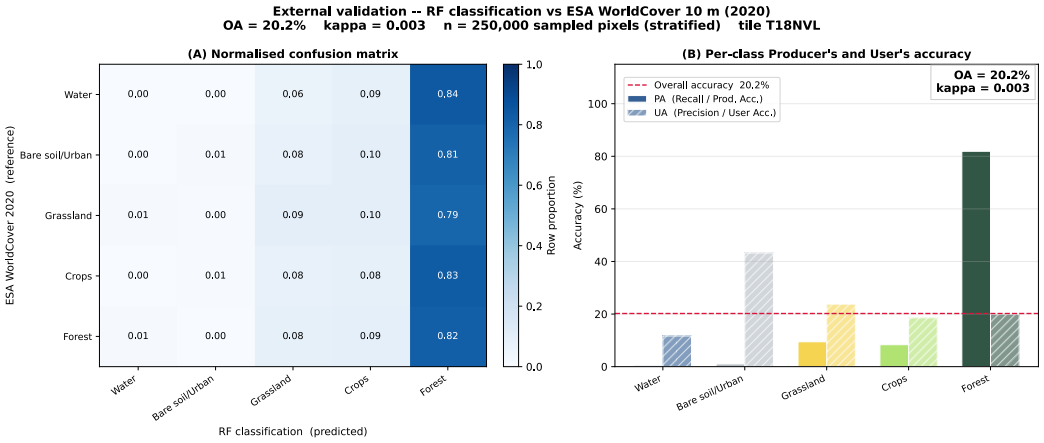


Figure 16. External agreement between the diagnostic RF spectral classification and ESA WorldCover 10 m 2020 (five-class scheme). (**A**) Normalised confusion matrix (n = 250,000 stratified pixels). (**B**) Producer's Accuracy (PA; solid bars) and User's Accuracy (UA; hatched bars) for each class. Bar colors distinguish the land-cover classes. Overall Accuracy (OA) = 20.2% and Cohen's Kappa (κ) = 0.003 indicate the limitations of single-date threshold-labelled spectral classification (Section 4.5).

The low OA is consistent with a well-documented spectral ambiguity specific to the Colombian Eje Cafetero: shade-grown coffee plantations have NDVI values of 0.45–0.65, identical to the threshold range used to define the Secondary forest class (Table 3). Because the RF training labels were derived from the same spectral thresholds (Section 2.8), this ambiguity is embedded in the training set; ESA WorldCover, by contrast, exploits multi-temporal Sentinel-1 SAR backscatter and year-round optical phenology that discriminate coffee canopy from closed-canopy forest. This finding is analysed further in Section 4.5.

3.14. Independent Point-Based Agreement Assessment

Table 11 and Figure 17 present the agreement metrics from the geographically stratified Google Earth Pro reference points. Although 210 points were initially digitised (35 per class), 201 points were retained after the nodata filter.

The independent point-based agreement assessment yielded OA = 19.9% and $\kappa = 0.032$, indicating very low agreement between the single-date diagnostic RF spectral-classification layer and the visually interpreted reference points. This result is consistent with the ESA WorldCover external comparison (OA = 20.2%, $\kappa = 0.003$; Table 10), suggesting that both external agreement assessments reveal the same dominant limitation of the classification workflow. In line with the sample-size considerations outlined in Section 2.12, these figures are used primarily to characterise the error structure and to corroborate the WorldCover-based diagnosis, rather than as high-precision estimates of map accuracy.

The confusion matrix shows systematic over-prediction of the Secondary forest class. Most Water, Bare/Urban, Grassland, Crops, and Dense forest reference points were classified as Secondary forest, while correct agreement was concentrated almost exclusively in the Secondary forest class itself. Secondary forest achieved the highest PA (88.6%) but a low UA (19.9%), indicating that the class was frequently predicted even when the reference label corresponded to other land-cover types. Crops showed the highest non-forest performance (PA = 18.2%, UA = 28.6%), whereas Water, Bare/Urban, and Dense forest obtained zero correct agreement.

These results confirm that the high internal RF accuracy should be interpreted as spectral self-consistency rather than independent thematic accuracy. The point-based agreement assessment supports the interpretation that single-date optical indices are insufficient to separate shade-grown coffee, mixed woody vegetation, secondary forest, and closed-canopy forest in the Eje Cafetero landscape.

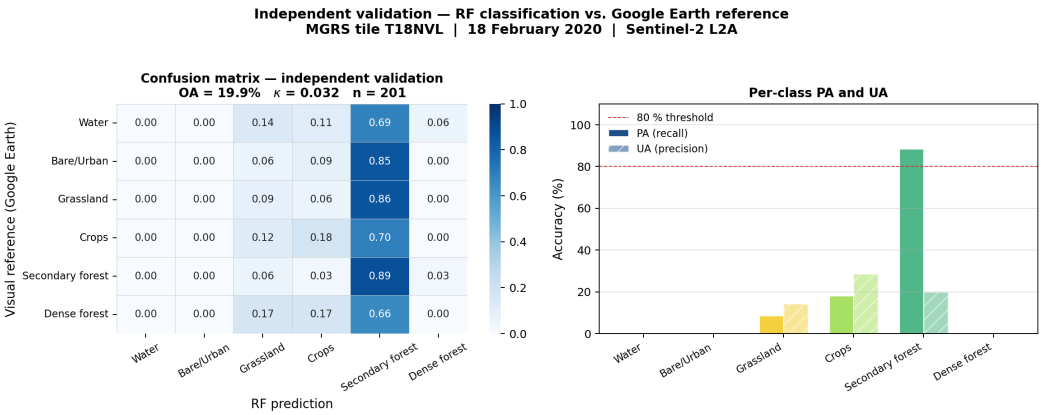


Figure 17. Independent point-based agreement assessment of the diagnostic RF spectral-classification layer using geographically stratified Google Earth Pro reference points (January–March 2020; $n = 201$ valid points after filtering). (**Left**) Row-normalised confusion matrix comparing visual reference classes and RF predictions. (**Right**) Producer's and User's Accuracy per class. OA = 19.9%; $\kappa = 0.032$. The dominance of the Secondary forest prediction column indicates systematic over-prediction of this class across multiple categories.

Table 11. Independent point-based agreement assessment of the diagnostic RF spectral-classification layer against visually interpreted Google Earth Pro reference points ($n_{ref} = 35$ initially digitised per class; date range: January–March 2020; $n_{valid} = 201$ after no-data filtering). OA = Overall Agreement; PA = Producer's Accuracy (recall); UA = User's Accuracy (precision). The low overall agreement (OA = 19.9%) is consistent with the ESA WorldCover external agreement assessment (OA = 20.2%, Table 10) and reflects systematic over-prediction of the Secondary forest class; see Section 4.5.

Class	PA (%)	UA (%)	F1	*n* Valid
Water	0.0	0.0	0.000	35
Bare/Urban	0.0	0.0	0.000	34
Grassland	8.6	14.3	0.107	35
Crops	18.2	28.6	0.222	33
Secondary forest	88.6	19.9	0.325	35
Dense forest	0.0	0.0	0.000	29
Overall	**OA = 19.9%**		**κ = 0.032**	201

4. Discussion

4.1. Spectral Index Behaviour in Tropical Andean Settings

Grand mean NDVI (0.472) is consistent with values reported for heterogeneous Andean vegetation mosaics with similar structural composition [6,7,15] and with the observed NDVI range in comparable studies (0.45–0.60 for mixed Colombian Andean vegetation [6]). The EVI scene-mean (0.573) exceeds NDVI (0.472), which is physically coherent because, after SCL masking, the retained pixel population is enriched in high-density, photosynthetically active canopy (SCL class 4), for which EVI's linearised response extends sensitivity beyond NDVI saturation at LAI > 3 [11].

The highly stable NDMI (0.195 ± 0.014) confirms consistently adequate canopy moisture throughout the study period, as expected for an Andean cloud-forest ecosystem receiving > 2000 mm yr^{-1} [19]. The elevated NDMI–optical index correlations ($r = 0.83$–0.94; Figure 10) indicate that, at the scene-mean level, SWIR-based moisture information covaries strongly with canopy-density metrics in this predominantly forest/agroforestry landscape, rather than providing statistically independent information [13]. NDMI's complementary value is instead evident at the pixel level, where it discriminates spectrally similar classes (e.g., Water vs. Bare soil/Urban; Section 3.8) that NDVI and EVI alone cannot separate, motivating its retention as a feature in the RF classification framework despite the observed scene-level collinearity.

Although none of the inter-annual OLS trends were statistically significant ($p > 0.05$; Table 6), EVI exhibited the largest positive slope ($+0.0023$ yr^{-1}), followed by NDMI ($+0.0008$ yr^{-1}) and NDVI ($+0.0005$ yr^{-1}), whereas SAVI showed a marginally negative slope (-0.0003 yr^{-1}). These small directional differences likely reflect the distinct spectral sensitivities and correction terms incorporated into each index formulation rather than meaningful ecological divergence.

4.2. Processing Efficiency and Replicability

The `/vsizip/` approach delivered a 98% reduction in per-scene I/O time (15 min $\rightarrow$ 5 s on a 16 GB RAM laptop) without altering pixel values, making the pipeline practical for institutions with limited computational infrastructure. This is comparable to performance reported for cloud-optimised GeoTIFFs (COGs) [18], but operates directly on the standard ESA SAFE distribution format without reprocessing. The 23.3 GB dataset was fully processed in $\approx$2 h without requiring cloud computing.

Per-pixel SCL masking retained 57.9% (median) of pixels per scene, compared to 0% for scenes above a 30% cloud-cover threshold that would be discarded in a scene-level

filter. At the 30% threshold, only 9 of 20 scenes would be retained, reducing temporal density by 55%. This loss would eliminate the only available 2021 observation and all dry-season scenes in 2018, severely compromising trend analysis.

4.3. Absence of Significant Inter-Annual Trends

The absence of statistically significant inter-annual trends across all four spectral indices suggests that the vegetation structure and moisture conditions of tile T18NVL remained broadly stable during the 2017–2025 study period. This interpretation is consistent with relatively low reported deforestation rates in the Eje Cafetero compared with more rapidly changing regions of Colombia such as the Amazon basin [19].

The lowest NDVI observation (0.425, 22 July 2023) coincided with the highest-cloud scene in the dataset (64.1%), suggesting residual cloud-edge contamination rather than true vegetation degradation. Although minor inter-annual fluctuations were observed, they remained within the natural variability envelope of the time series. Continued monitoring using a denser and more seasonally balanced Sentinel-2 archive will nonetheless be important for improving statistical power and detecting subtle long-term ecological changes as additional Sentinel-2B and Sentinel-2C observations accumulate.

4.4. Classification and Separability

The spectral signature analysis (Figure 11) suggests that the four-index feature space provides strong discrimination for five of the six classes. The anticipated confusion between Crops and Secondary forest (overlapping NDVI range 0.45–0.55; Table 3) is reflected in the Jeffries–Matusita distance reported in Table 9: JM = 1.716 falls below the high-separability threshold of 1.9 [9], indicating only moderate separability. All other class pairs exceeded this threshold, suggesting that the main classification difficulty is concentrated at the Crops–Secondary forest boundary, consistent with comparable multitemporal tropical classification studies [6,17].

It should be noted that NDVI and SAVI are mathematically collinear by construction (Equation (4) is a linear soil-adjusted rescaling of Equation (2)), as confirmed by their scene-mean correlation ($r > 0.90$; Section 3.7). This redundancy means the four-feature RF classifier effectively relies on three independent optical/moisture signals (NDVI/SAVI jointly, EVI, NDMI) rather than four. While this redundancy does not invalidate the internal spectral-rule agreement reported in Table 8, it does limit the amount of independent structural information available to separate spectrally overlapping classes such as shade-grown coffee and secondary forest, reinforcing the need for complementary, non-collinear features (e.g., SAR backscatter or texture metrics) in future iterations.

Although the Secondary forest class dominates the reference scene (81.3% of valid pixels; Figure 14), it includes shade-grown coffee agroforestry and successional woody vegetation rather than homogeneous closed-canopy forest. At the 10 m Sentinel-2 spatial scale, many pixels therefore likely contain mixed vegetation and soil backgrounds, contributing to the observed spectral overlap between Crops and Secondary forest classes.

Future work incorporating multitemporal, phenological, or object-based features could further reduce confusion between managed cropland and early-successional woody vegetation.

4.5. Internal vs. External Accuracy: Coffee-Forest Spectral Confusion

The high internal OA reported for the RF classifier (Table 8) reflects primarily *spectral self-consistency* rather than absolute thematic accuracy. Because training labels were derived from the same spectral thresholds used to evaluate the model, the confusion matrix measures how consistently the RF reproduces those thresholds rather than how well it represents independent ground truth, a known limitation of threshold-labelled training sets [9].

Both external validations, ESA WorldCover (OA = 20.2%, $\kappa = 0.003$; Table 10) and the Google Earth Pro point-based reference (OA = 19.9%, $\kappa = 0.032$; Table 11), revealed systematic over-prediction of forest-related classes despite the near-perfect internal cross-validation accuracy. The consistency of this pattern across two independent validation approaches suggests that the principal limitation arises from the single-date, four-index feature space rather than from the reference datasets themselves.

The dominant confusion pattern, in which 81.9% of RF predictions were assigned to Forest regardless of the WorldCover reference class, reflects a genuine bio-physical ambiguity in the Eje Cafetero landscape. Shade-grown coffee (*Coffea arabica* cultivated under a persistent upper canopy) occupies a substantial proportion of the regional agricultural area [19] and produces spectral responses largely indistinguishable from secondary forest in single-date 10 m optical imagery. ESA WorldCover partially mitigates this limitation by integrating multitemporal Sentinel-2 composites and Sentinel-1 SAR backscatter, which provide phenological and structural information absent from the present RF feature space.

These findings suggest that the RF "Secondary forest" class should be interpreted more broadly as *woody or mixed-canopy vegetation*, encompassing secondary forest, shade-coffee agroforestry, and related mixed vegetation systems. Future iterations should incorporate: (i) multitemporal phenological features; (ii) Sentinel-1 SAR backscatter; and (iii) higher-resolution ancillary land-use or crop-type data to improve discrimination within this coffee-dominated landscape.

4.6. Limitations

Several limitations should be considered when interpreting the present results: (i) training samples derived from spectral thresholds rather than GPS-located field observations introduce partial circularity between class definition and spectral validation, as reflected by the discrepancy between internal and external accuracy metrics (Sections 3.13 and 4.5); (ii) single-date, four-index classification cannot fully resolve the spectral ambiguity between shade-grown coffee, secondary forest, and dense forest without multitemporal or SAR-derived features; (iii) uneven temporal sampling (Figure 3), including the absence of valid scenes in 2022, reduces the robustness of long-term trend estimation; (iv) three scenes could not be processed on the 16 GB RAM development system because of memory constraints, indicating that chunked or distributed I/O strategies may be required for larger archives; (v) pixel-wise RF classification ignores spatial context, which may reduce thematic consistency in ecotone and mixed-cover areas; (vi) the Google Earth Pro validation used 35 geographically stratified points per class, below the sample sizes commonly recommended for rigorous thematic accuracy assessment (50–75 points per class; [26]), so OA and PA/UA metrics should be interpreted primarily as indicators of error structure rather than as precise population-level accuracy estimates; and (vii) no topographic (illumination) correction was applied to the Sen2Cor L2A reflectance products, despite the pronounced elevation gradient of the study area (>3500 m; Section 2.1). Slope- and aspect-dependent illumination differences can bias BOA reflectance and, consequently, vegetation-index values on steep terrain; standard corrections such as SCS+C [30] were not applied in this diagnostic workflow. This represents a methodological limitation that should be addressed before extending the pipeline to steeper subregions of the Cordillera Central.

5. Conclusions

This study developed and assessed an open-source, fully reproducible pipeline for long-term vegetation-index monitoring using Sentinel-2 L2A imagery, demonstrated over an 8-year scene inventory for the Colombian Eje Cafetero. The principal conclusions are:

1. The `/vsizip/` storage-efficient reading strategy reduces per-scene processing time by ≈98% without altering pixel values, making the pipeline practical for institutions with limited computing resources.
2. SCL per-pixel cloud masking retains valid observations from partially clouded scenes (median 57.9% valid pixels at 35.4% mean cloud cover), retaining 2.2× more temporal observations than a conventional scene-level 30% cloud-cover threshold.
3. Grand mean NDVI (0.472 ± 0.018) is consistent with Andean vegetation reference values; seasonal peaks in September–November align with the bimodal Eje Cafetero precipitation climatology, supporting the seasonal coherence of the vegetation-index time series.
4. No statistically significant inter-annual trend was detected for NDVI, EVI, SAVI, or NDMI ($p > 0.05$), indicating no detectable monotonic vegetation-index trend in tile T18NVL over 2017–2025.
5. The four-index feature space provides high diagnostic spectral separability for most threshold-labelled class pairs (JM > 1.9), supporting a reproducible RF spectral-classification procedure while also revealing limitations for thematic land-cover mapping without independent field data.
6. The multi-level diagnostic assessment provides a reproducible way to characterise workflow behaviour, internal spectral-rule agreement, and external classification limitations in data-scarce tropical regions.
7. External agreement assessments against ESA WorldCover and Google Earth Pro revealed persistent spectral ambiguity between shade-grown coffee systems and secondary forest, highlighting the limitations of single-date optical classification in heterogeneous tropical landscapes.

Overall, the proposed framework demonstrates that reproducible Sentinel-2 workflows can support vegetation-index monitoring in cloud-prone tropical regions, while also highlighting the importance of multitemporal, SAR, and field-validated information for improving thematic land-cover mapping in complex Andean environments.

Author Contributions: Conceptualisation, K.D.O.-Q. and M.H.-L.; methodology, K.D.O.-Q.; software, K.D.O.-Q.; validation, K.D.O.-Q., D.Z.-Y. and M.F.C.-M.; formal analysis, K.D.O.-Q. and D.Z.-Y.; investigation, K.D.O.-Q. and D.Z.-Y.; resources, M.H.-L. and G.A.H.-L.; data curation, K.D.O.-Q. and D.Z.-Y.; writing—original draft preparation, K.D.O.-Q.; writing—review and editing, K.D.O.-Q., D.Z.-Y., M.F.C.-M., M.H.-L. and G.A.H.-L.; visualisation, K.D.O.-Q.; supervision, M.H.-L. and G.A.H.-L.; project administration, K.D.O.-Q. and G.A.H.-L.; funding acquisition, M.H.-L. and G.A.H.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Sentinel-2 L2A imagery is freely available from the ESA Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu (accessed on 26 April 2026)). Processing scripts and derived raster outputs are available at https://github.com/kevin-ortega2724/satelite (accessed on 26 April 2026). The dataset is publicly available under the Copernicus open-data licence (CC-BY-SA 4.0).

Acknowledgments: The authors gratefully acknowledge the Master's Program in Electrical Engineering and the Vicerrectoría de Investigaciones, Innovación y Extensión of the Universidad Tecnológica de Pereira (UTP) for their financial support. The authors also extend their appreciation to the laboratory team and to the student research groups *Huellas sobre Sistemas Digitales y Mantenibilidad* and *Inteligencia Artificial e Industria 4.0* for their technical assistance, collaboration, and support throughout the development of this study. During the preparation of this manuscript, the authors used artificial intelligence tools, including Claude opus 4.8 (Anthropic, 2024–2025) and Perplexity AI, to support the translation of selected phrases into English, since the original draft was written in Spanish, as well as

to perform grammatical improvements, style corrections, and verification of formatting and writing errors in the text.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

BOA	Bottom-of-Atmosphere reflectance
CRS	Coordinate Reference System
EVI	Enhanced Vegetation Index
FCC	False-Colour Composite
GDAL	Geospatial Data Abstraction Library
JM	Jeffries–Matusita (distance)
L2A	Level-2A (Sentinel-2 processing level)
LAI	Leaf Area Index
MDI	Mean Decrease in Impurity
MGRS	Military Grid Reference System
NDMI	Normalised Difference Moisture Index
NDVI	Normalised Difference Vegetation Index
OA	Overall Accuracy
OLS	Ordinary Least Squares
OOB	Out-of-Bag
PA	Producer's Accuracy
RF	Random Forest
SAFE	Standard Archive Format for Europe
SAVI	Soil-Adjusted Vegetation Index
SCL	Scene Classification Layer
SWIR	Shortwave Infrared
UA	User's Accuracy

References

1. Turner, W.; Spector, S.; Gardiner, N.; Fladeland, M.; Sterling, E.; Steininger, M. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* **2003**, *18*, 306–314. [CrossRef]
2. Zhu, Z.; Zhou, Y.; Seto, K.C.; Stokes, E.C.; Deng, C.; Pickett, S.T.; Taubenböck, H. Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote Sens. Environ.* **2019**, *228*, 164–182. [CrossRef]
3. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. [CrossRef]
4. ESA. Sentinel-2 User Handbook. In *Proceedings of the ESA Standard Document*; ESA-ENVIS-EOPG-TN-15-0007; ESA: Paris, France, 2015.
5. Roy, D.P.; Wulder, M.A.; Lovel, T.R.; Ce, W.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [CrossRef]
6. Achicanoy, J.A.; Rojas-Robles, R.; Sánchez, J.E. Análisis y proyección de las coberturas vegetales mediante el uso de sensores remotos y Sistemas de Información Geográfica en la localidad de Suba, Bogotá-Colombia. *Gestión Ambiente* **2018**, *21*, 41–58. [CrossRef]
7. Puerta Tuesta, R.H.; Iannacone Oliver, J.A.; Reátegui Inga, M.E. Una revisión del uso de imágenes Sentinel-2 para el monitoreo de la cobertura boscosa a nivel global. *Ing. Compet.* **2023**, *25*, e–30612708. [CrossRef]
8. Cortez Requejo, Y.L. Pérdida de la Cobertura Vegetal Mediante el uso de imáGenes Satelitales en la Subcuenca Gera, Distrito de Jepelacio. Master's Thesis, Universidad Nacional de San Martín, Moyobamba, Peru, 2021.
9. Jensen, J.R. *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd ed.; Pearson Prentice Hall: London, UK, 2007.
10. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
11. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]

12. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
13. Gao, B.C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]
14. Wilson, E.H.; Sader, S.A. Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sens. Environ.* **2002**, *80*, 385–396. [CrossRef]
15. Asner, G.P. Biophysical remote sensing signatures of arid and semiarid ecosystems. *Remote Sens. Environ.* **2004**, *89*, 205–218. [CrossRef]
16. Lawley, V.; Lewis, M.; Clarke, K.; Ostendorf, B. Site-based and remote sensing methods for monitoring indicators of vegetation condition: An Australian review. *Ecol. Indic.* **2016**, *60*, 1273–1283. [CrossRef]
17. Jaramillo-Véliz, J.J.; Quimis-Gómez, A.J.; Gómez-Bailón, S.P. Análisis del cambio espacio-temporal en la cobertura vegetal del cerro de hojas Jaboncillo durante el periodo 2015–2020. *Polo Conoc.* **2021**, *6*, 632–652. [CrossRef]
18. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
19. IDEAM. *Leyenda Nacional de Coberturas de la Tierra. Metodología CORINE Land Cover Adaptada para Colombia*; Technical Report; Instituto de Hidrología, Meteorología y Estudios Ambientales: Bogotá, Colombia, 2010.
20. European Space Agency. Sentinel-2 L2A—Documentation. Copernicus Data Space Ecosystem. 2024. Available online: https://documentation.dataspace.copernicus.eu/APIs/SentinelHub/Data/S2L2A.html (accessed on 28 May 2026).
21. Chander, G.; Markham, B. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2674–2677. [CrossRef]
22. Gilabert, M.; González-Piqueras, J.; García-Haro, F.; Meliá, J. A generalized soil-adjusted vegetation index. *Remote Sens. Environ.* **2002**, *82*, 303–310. [CrossRef]
23. Landsat Missions. Normalized Difference Moisture Index. Available online: https://www.usgs.gov/landsat-missions/normalized-difference-moisture-index (accessed on 27 April 2026).
24. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
25. Zanaga, D.; Van De Kerchove, R.; De Keersmaecker, W.; Souverijns, N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S.; et al. ESA WorldCover 10 m 2020 v100. *Zenodo* **2021**. Data product. . [CrossRef]
26. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2008. [CrossRef]
27. Mann, H.B. Nonparametric tests against trend. *Econometrica* **1945**, *13*, 245–259. [CrossRef]
28. Kendall, M.G. *Rank Correlation Methods*, 4th ed.; Griffin: London, UK, 1975.
29. Sen, P.K. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **1968**, *63*, 1379–1389. [CrossRef]
30. Soenen, S.A.; Peddle, D.R.; Coburn, C.A. SCS+ C: A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2148–2159. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.