SATELLITE IMAGERY TO MACHINE LEARNING DATASETS: AN AUTOMATED SYSTEM FOR SOIL WATER STRESS MONITORING IN AGRICULTURE

Alexandre Heideker[1], Eduardo Speranza[2], Ednaldo Ferreira[2], Dener Silva[1], Carlos Kamienski[2], Reinaldo Bianchi[1]

[1]Centro Universitário FEI, Brazil
[2]Brazilian Agricultural Research Corporation (Embrapa), Brazil
[3]Federal University of the ABC, Brazil

A paper from the Proceedings of the

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

13-15 July 2026

Porto Alegre, Brazil

Abstract. *Satellite remote sensing is a key data source for precision agriculture, enabling vegetation indices such as NDMI and EVI to monitor vegetation health and soil water stress over large areas. Its practical use, however, is often constrained by manual image selection, downloading, and preprocessing workflows, which are time-consuming, hard to reproduce, and ill-suited to building the large, multi-temporal datasets that machine learning requires. This study presents an automated system for satellite image acquisition and preprocessing that enables scalable dataset generation for soil water stress monitoring. The system retrieves historical and current Sentinel-2 multispectral imagery through the Copernicus API, guided by predefined spatial and temporal parameters. Areas of interest are defined as multiple polygons in GeoJSON files representing individual agricultural sectors, which are grouped within a simplified bounding area to reduce redundant queries and improve retrieval efficiency. By eliminating manual scene inspection, the system reduces human effort while ensuring consistency and reproducibility, enabling systematic expansion of monitored regions and supporting the creation of high-quality datasets for training and deploying machine learning models in irrigation management. Its modular design accommodates the integration of additional satellite missions, spectral indices, and preprocessing methods in the future.*

Keywords. *Sentinel-2; satellite remote sensing; vegetation indices; EVI; NDMI; machine learning datasets; cloud masking; soil water stress; coffee crop monitoring.*

Introduction

Satellite remote sensing has established itself as a fundamental data source for precision agriculture, enabling non-invasive monitoring of vegetation dynamics, soil conditions, and crop water status across large areas at regular temporal intervals [1]. The Sentinel-2 mission, operated by the European Space Agency (ESA) under the Copernicus program, provides multispectral imagery at spatial resolutions from 10 to 60 meters across 13 spectral bands, a 290 km swath width, and a revisit interval of approximately five days [2]. The Copernicus program's open, free

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

data policy makes Sentinel-2 one of the most widely used satellite platforms for agricultural monitoring worldwide [3].

Among the spectral indices derived from multispectral imagery, the Enhanced Vegetation Index (EVI) and the Normalized Difference Moisture Index (NDMI) are particularly relevant for crop monitoring. EVI was developed to optimize the vegetation signal in high-biomass regions by reducing soil and atmospheric influences [4], overcoming the saturation limitations of the traditional Normalized Difference Vegetation Index (NDVI). NDMI uses the near-infrared (NIR) and short-wave infrared (SWIR) bands to estimate vegetation water content and canopy moisture status [5], thereby serving as a direct proxy for crop water stress monitoring. Both indices have been widely applied in precision agriculture research on irrigation management [6,7], including studies conducted at the same experimental site as this work [8].

Despite the remarkable potential of public data sources, the construction of satellite image datasets suitable for machine learning applications in agriculture remains constrained by practical limitations in acquisition and preprocessing workflows. Conventional approaches rely on manual steps such as browsing image catalogs, visually selecting scenes with low cloud cover, downloading large files, and performing ad hoc preprocessing for each area of interest. These manual workflows are time-consuming, error-prone, and difficult to reproduce at scale [9]. They are fundamentally incompatible with the systematic construction of large, multi-temporal datasets required by modern machine learning pipelines.

The discontinuities in obtaining, processing, and interpreting data for data-driven agriculture are addressed within the scope of the SMART project (Smart Monitoring and Automation for Rural Technologies), an international initiative between Brazil and China focused on developing reliable IoT-continuum platforms for precision agriculture [10]. The SMART project employs IoT sensors, 5G/6G networks, edge computing, and machine learning to support data-driven decision-making for irrigation and fertilization management. One of the project's pilot sites is the Nossa Senhora da Aparecida da Água Branca (NSAAB) coffee farm in Brotas, São Paulo, Brazil, where the automated satellite monitoring system described in this paper has been deployed as a component of a smart irrigation decision-support pipeline.

In this context, this paper presents an automated system for satellite image acquisition and preprocessing designed to enable scalable dataset generation for soil water stress monitoring. The system retrieves Sentinel-2 imagery via the Copernicus Data Space Ecosystem API, performs spatial clipping to multiple areas of interest defined by GeoJSON files, harmonizes band resolutions by resampling and/or interpolation, reconstructs cloud-contaminated pixels using temporal compositing, and assesses EVI and NDMI spectral indices. The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 describes the proposed system; Section 4 presents the experimental results; and Section 5 draws conclusions.

Literature Review

Remote Sensing for Precision Agriculture

The application of satellite remote sensing to precision agriculture has been extensively explored over the past two decades. The availability of medium-resolution, free, and open-access missions such as Sentinel-2 has substantially reduced the barrier to large-scale crop monitoring. Reviews of Sentinel-2 features and applications in precision agriculture highlight its utility across a wide range of tasks, including crop type mapping, biophysical parameter estimation, and abiotic stress detection [1]. The integration of satellite time series with machine learning algorithms has enabled the automated classification of crop types and the detection of phenological events at scales ranging from farms to regions [9,11].

For coffee cultivation in Brazil, remote sensing has been applied to estimate leaf water potential, classify crop phenological stages, and support irrigation planning [12,13]. São Paulo state, the third largest producer among Brazilian states, presents a challenging environment for satellite remote sensing due to frequent cloud cover during the wet season (October–March), reinforcing the need for robust temporal reconstruction methods.

Vegetation Indices: EVI and NDMI

The EVI was introduced as a standard vegetation product for the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor and has since been widely adopted by other platforms, including Sentinel-2 [4]. It uses Sentinel-2 bands B8 (NIR), B4 (Red), and B2 (Blue), and was designed to be more responsive to canopy structure variations and less susceptible to saturation in high-biomass regions than NDVI. Equation (1) presents the EVI formulation, where NIR, R, and B denote near-infrared, red, and blue surface reflectances, respectively.

$$EVI = 2.5 \times \frac{NIR - R}{NIR + 6R - 7.5B + 1} \qquad (1)$$

The Normalized Difference Water Index (NDWI) was originally proposed [5] for remote sensing of vegetation liquid water from space. Using Sentinel-2 bands B8A (NIR) and B11 (SWIR), it is expressed as in Equation (2), where NIR and SWIR denote near-infrared and short-wave infrared reflectances, respectively. Values approaching +1 indicate high vegetation moisture; values near or below zero may signal water stress or senescence.

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \qquad (2)$$

The study by Ferreira et al. [8] demonstrated the differential sensitivity of NDMI and EVI to soil moisture dynamics, as measured by in situ porous-matrix sensors. The study found that NDMI exhibited a peak cross-correlation with soil matric potential at a two-day lag (Pearson coefficient, $r = 0.71$), indicating a rapid canopy water response immediately after dry soil conditions. EVI showed lower temporal sensitivity ($r = 0.114$ at five days), consistent with its metric properties, and nonspecific to water. Critically, the manual workflow used to obtain the satellite imagery in that study — relying on the RAVI QGIS plugin [14] for image downloading and index extraction — required per-session manual intervention and would not scale to continuous multi-site monitoring. The automated system proposed here directly addresses that limitation, enabling the systematic, reproducible acquisition of the same Sentinel-2 data streams required by such analyses across multiple farms and time periods without human intervention.

Cloud Contamination and Temporal Reconstruction

Cloud cover is one of the primary limitations of optical satellite imagery in tropical and subtropical regions. Multiple approaches have been developed to address this challenge, ranging from simple temporal compositing methods to advanced machine-learning-based reconstruction techniques that leverage multi-source optical and SAR data [15,16]. While complex deep learning-based gap-filling methods offer superior accuracy, simple temporal averaging within a configurable window provides a robust and computationally lightweight solution for operational monitoring pipelines, particularly for perennial crops with relatively smooth spectral trajectories over weekly to monthly timescales.

Automation of Satellite Image Pipelines

Efforts to automate satellite image acquisition and preprocessing have been facilitated by cloud computing platforms such as Google Earth Engine, as well as Python libraries including Rasterio, GDAL, and Sentinelsat. The Copernicus Data Space Ecosystem (CDSE), launched in 2023 as the successor to the Copernicus Open Access Hub, provides programmatic access to the full Sentinel data archive through standardized REST APIs including OData, STAC, openEO, and Sentinel Hub [2]. Despite these advances, end-to-end automated pipelines that integrate API-based acquisition, multi-polygon GeoJSON clipping, resolution harmonization, and cloud reconstruction into a single reproducible system specifically designed for machine learning dataset generation remain scarce in the published literature.

Proposed System

The proposed system automates the full pipeline from raw Sentinel-2 image retrieval to the generation of analysis-ready spectral indices. Figure 1 provides an overview of the four main processing stages.

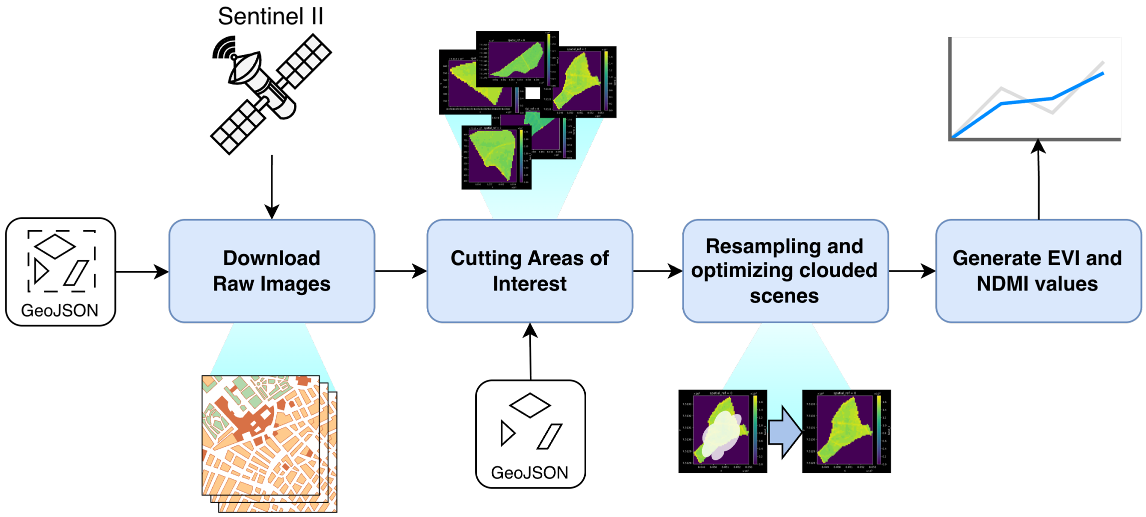

Fig. 1. Overview of the automated satellite image pipeline: from raw Sentinel-2 acquisition to EVI and NDMI dataset generation.

Automatic Image Acquisition via Copernicus API

The acquisition module performs automated queries to the Sentinel-2 catalog provided by the Copernicus Data Space Ecosystem (CDSE) API [2], using user-configurable spatial and temporal parameters. Input parameters include temporal window (start and end dates), product processing level (this work uses Level-2A surface reflectance products), and maximum cloud coverage percentage per scene. The system authenticates via OAuth2 tokens and queries the OData catalog endpoint to identify available scenes, then downloads the relevant spectral bands automatically.

To minimize the number of API calls and reduce data transfer volume, the areas of interest defined by the input GeoJSON polygons are enclosed within a unified bounding box. This simplified spatial extent is used as the search criterion in the API query, resulting in the download of scenes covering the entire monitored region in a single request batch, without the need for separate per-polygon queries.

Cutting Area of Interest

The monitored areas are defined through GeoJSON files containing one or more polygon features, each representing an individual agricultural sector, field block, or farm sub-unit. This approach allows seamless configuration for different farms or multi-site monitoring campaigns by simply updating the input GeoJSON file.

After downloading the raw Sentinel-2 tiles, the system performs precise spatial clipping for each polygon individually. Each clipped raster corresponds exactly to the boundary defined in the GeoJSON file, ensuring spectral isolation from adjacent areas (e.g., roads, native vegetation, buildings). All clipped images are co-registered and stored as georeferenced GeoTIFF files.

Preprocessing: Band Resolution Harmonization

Sentinel-2 bands are provided at three spatial resolutions: 10 m (B2, B3, B4, B8), 20 m (B5, B6, B7, B8A, B11, B12), and 60 m (B1, B9, B10). The computation of EVI requires bands B2, B4, and B8, while NDMI requires B8A and B11, combining bands from different native resolutions. The preprocessing module performs bilinear resampling of all 20 m bands to 10 m resolution, ensuring that all spectral indices are computed at a uniform spatial resolution.

Cloud Pixel Reconstruction

The cloud reconstruction module addresses the gap-filling problem for cloud-contaminated pixels using a temporal averaging strategy. Cloud and cloud-shadow pixels are identified using the Scene Classification Layer (SCL) provided with each Sentinel-2 Level-2A product. For each pixel flagged as invalid, the system retrieves corresponding pixel values from previous cloud-free acquisitions within a configurable temporal window (e.g., 30, 45, or 60 days prior). The mean of available valid observations within that window is then used to substitute the invalid pixel, producing a temporally reconstructed image (Figure 2) [15].

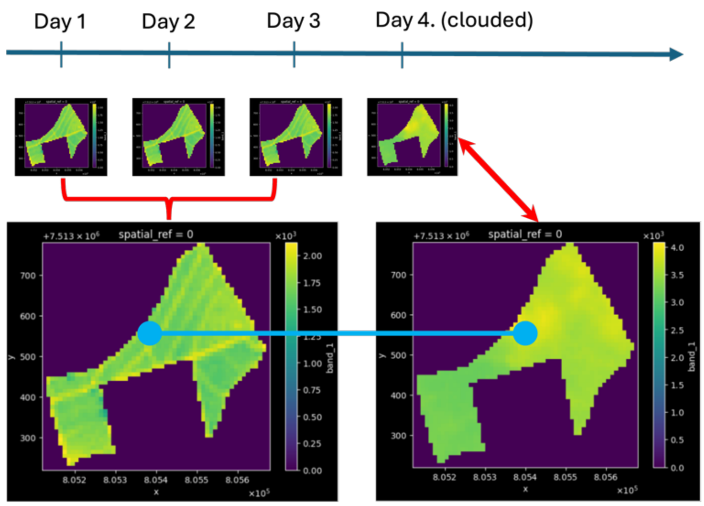

Fig. 2. Cloud pixel reconstruction mechanism: invalid pixels in the current scene are substituted by the temporal mean of valid observations within a user-defined lookback window.

This approach is grounded in the assumption that perennial crops such as coffee exhibit relatively gradual spectral trajectories over weekly to monthly timescales, making temporal averaging a reasonable and computationally efficient approximation. The configurable window parameter allows users to balance temporal fidelity against data availability. Reconstructed index values are flagged in the output time series, enabling downstream models to distinguish original from imputed observations. The proposed pipeline preserves local temporal consistency and prevents the generation of unrealistic values in the time series. Nevertheless, detailed temporal dynamics cannot be accurately reconstructed during prolonged periods of cloud contamination. The complete source code of our proposed solution is publicly available on the SAT-ML website [17].

Experimental Results: NSAAB Coffee Farm, Brotas/SP, Brazil

Study Area and Experimental Setup

The system was evaluated at the NSAAB coffee farm, a pilot farm of the SMART project [10], and satellite-derived spectral indices constitute a key remote-sensing data layer in the project's smart irrigation decision-support pipeline. Notably, prior work at this same farm site has already demonstrated the agronomic relevance of EVI and NDMI for irrigation monitoring using manually acquired Sentinel-2 data [7]; the automated system presented here is designed to replace and scale that manual acquisition process.

The Brotas region is located in the Western Plateau of São Paulo state (approximately 22.3°S, 48.1°W) at elevations between 600 and 750 m. Climate is classified as Cwa (humid subtropical with dry winters) under the Köppen system, with hot and rainy summers (October–March) and dry winters (April–September). This seasonal pattern results in high cloud frequency during the austral summer, making the region a realistic test case for the cloud reconstruction mechanism. The NSAAB farm contains multiple coffee blocks (Coffea arabica), each defined as an individual polygon in the input GeoJSON file.

The Sentinel-2 constellation crosses the Brotas region every 5 days, yielding approximately 54 potential acquisitions between August 2025 and May 2026. For each acquisition, the system automatically performed the full pipeline: API query, download, band clipping to each farm polygon, resolution harmonization, cloud masking, temporal reconstruction, and EVI/NDMI computation. Also, a 15-day temporal window for cloud coverage optimization was used, representing about 3 samples due to the satellite's overflight of the NSAAB region every 5 days.

EVI Time Series

Figure 3 presents the EVI time series for the NSAAB farm over the monitored period. Each acquisition is plotted in a continuous red line, and the blue bars indicate the cloud cover percentage of the scene. The dashed line represents the optimized EVI computed after cloud pixel reconstruction [4,7].

The EVI series reflects the expected phenological trajectory of Arabica coffee in São Paulo. During the dry season (August–September 2025), EVI values are moderate, consistent with reduced photosynthetic activity and the absence of new vegetative flush. With the onset of the rainy season (October–November 2025), EVI shows a progressive increase, indicating canopy expansion and increased vegetative vigor. The period from December 2025 to February 2026 contains the acquisitions with the highest cloud-cover percentages, highlighting the importance of the reconstruction mechanism. The optimized EVI series (dashed line) presents a smoother, phenologically consistent trajectory, effectively filling the gaps introduced by cloud-contaminated scenes.

NDMI Time Series

Figure 4 presents the equivalent analysis for the NDMI index. The NDMI is more directly sensitive to vegetation water content than EVI, providing a more immediate indicator of the coffee crop's hydration status [5,6,8].

The NDMI series shows lower values during the dry season (August–September 2025), indicating reduced foliar water content typical of pre-irrigation periods. A gradual recovery is evident from October 2025 onward, tracking rainfall onset and progressive canopy rehydration. The joint analysis of EVI and NDMI time series provides complementary information: while EVI captures canopy development and photosynthetic activity, NDMI directly reflects vegetation water status, enabling the identification of periods of potential water stress that are not always evident from EVI alone [8]. Both series generated by the proposed system are directly applicable to training and validating machine learning models for smart irrigation decision support within the SMART project framework.

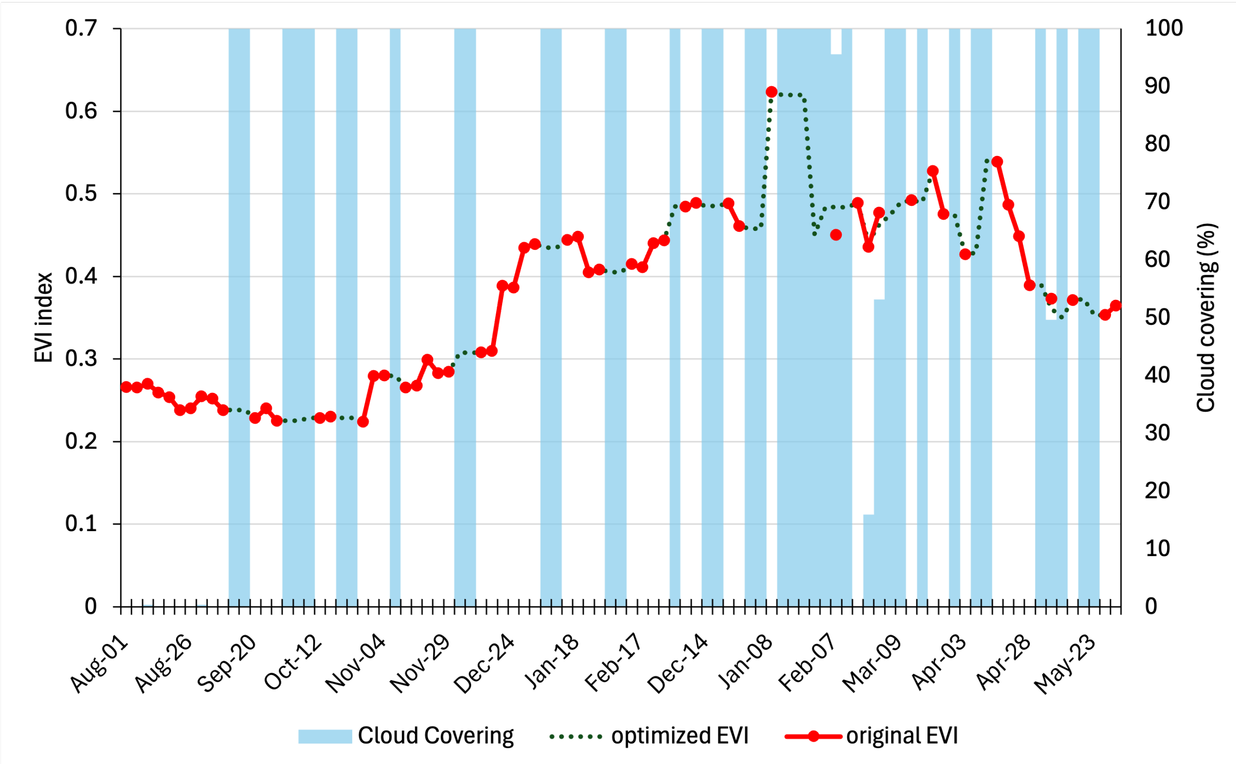

Fig. 3. EVI time series for the NSAAB coffee farm (August 2025 – May 2026). Solid markers indicate original values, and blue bars indicate the cloud cover. The dashed line shows optimized values after temporal cloud reconstruction.

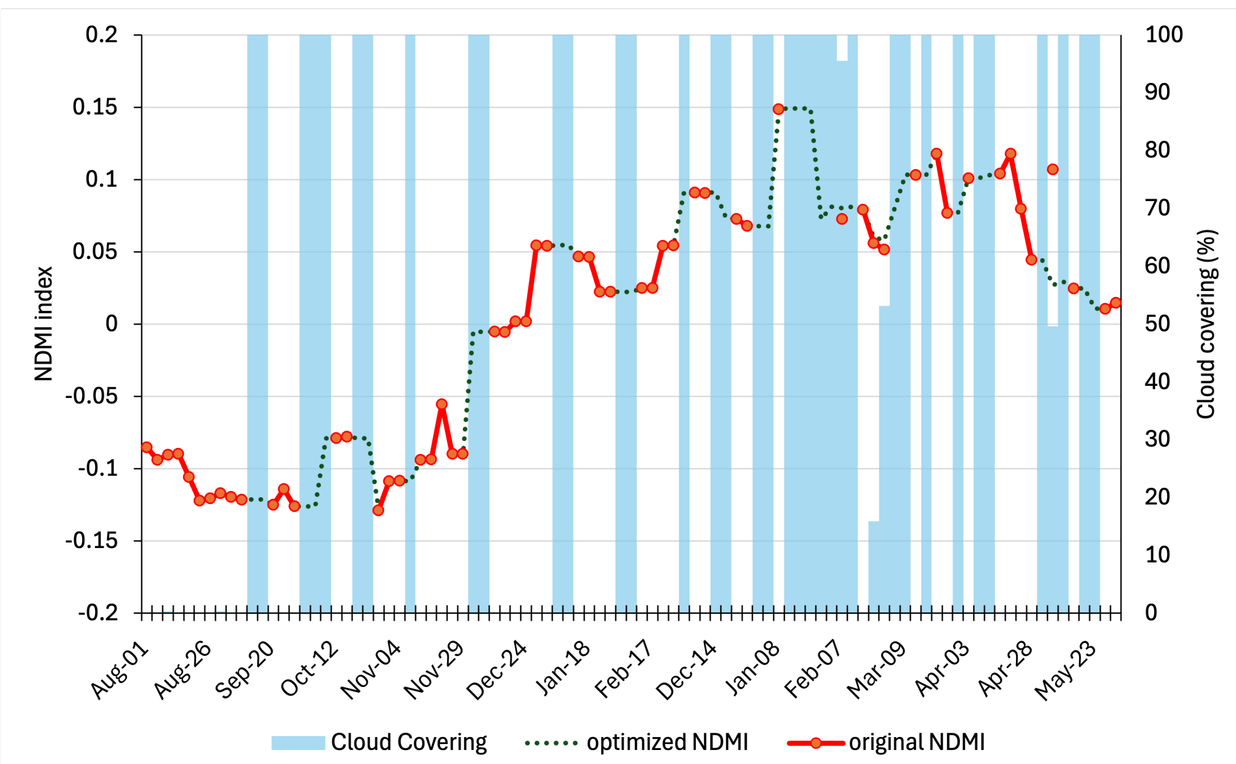

Fig. 4. NDMI time series for the NSAAB coffee farm (August 2025 – April 2026). Solid markers indicate original values, and blue bars indicate the scene cloud cover. The dashed line shows optimized values after temporal cloud reconstruction.

Discussion

The results demonstrate the viability and practical utility of the proposed automated pipeline for continuous satellite-based crop monitoring. The automation of the acquisition, clipping, preprocessing, and index generation stages eliminates the manual bottlenecks that have historically limited the construction of large-scale satellite datasets for precision agriculture. The GeoJSON-based area definition and bounding-box optimization strategy allow the system to efficiently scale from single-field to multi-farm monitoring without changes to the processing architecture.

A key motivation for this work is the direct continuity with prior research conducted at the same NSAAB farm site [8], which demonstrated the agronomic relevance of EVI and NDMI for irrigation timing decisions using Sentinel-2 data manually extracted via the RAVI QGIS plugin. In that study, the authors acknowledge scalability as a limitation: the plugin-based workflow requires manual session-by-session interaction and cannot support continuous, multi-site, or long-term monitoring without substantial human effort. The system proposed here directly resolves this bottleneck, providing the automated data acquisition infrastructure needed to deploy the vegetation-index-based irrigation approach at operational scale.

The temporal cloud reconstruction mechanism proved effective in maintaining the continuity of EVI and NDMI time series during the high-cloud-cover summer months in the Brotas region. The temporal mean within a configurable lookback window is a lightweight, interpretable gap-filling approach well-suited to operational monitoring pipelines [15,16], particularly for perennial crops with gradual seasonal spectral dynamics. The configurable window parameter allows users to adapt the trade-off between temporal accuracy and data availability to their specific context. In this experimental setup, a 15-day temporal window represents three samples - larger temporal windows can yield values unrelated to the actual climate profile and crop status, as lower values can reduce the samples to an impractical solution.

It is important to acknowledge the limitations of the mean-based reconstruction approach. In periods of rapid phenological change — such as flowering, fruit set, or harvest — or following abrupt stress events (frost, severe drought), temporal averaging may introduce bias in the reconstructed values. The development of more sophisticated gap-filling methods, such as temporal weighted interpolation, harmonic fitting, or SAR-optical fusion-based reconstruction, represents a natural extension of the current system.

The modular architecture of the system facilitates integration with additional satellite missions, new spectral indices, and advanced preprocessing algorithms. Within the SMART project framework, the satellite data pipeline described here is designed to complement in situ IoT sensor networks, enabling multi-source data fusion to improve irrigation management decisions.

Conclusion

This paper presented an automated system for satellite image acquisition and preprocessing designed to support large-scale dataset generation for soil water stress monitoring in precision agriculture. The system integrates Sentinel-2 imagery retrieval via the Copernicus Data Space Ecosystem AP, multi-polygon spatial clipping using GeoJSON files, band-resolution harmonization, temporal cloud-pixel reconstruction, and EVI and NDMI index generation into a fully automated, reproducible pipeline.

The system was evaluated at the NSAAB coffee farm in Brotas, SP, Brazil, a pilot site of the SMART project for intelligent irrigation, monitored from August 2025 to April 2026. The results demonstrated the effectiveness of the automated pipeline and the temporal cloud reconstruction mechanism in generating continuous, analysis-ready EVI and NDMI time series. The system provides the automated acquisition infrastructure required to scale the vegetation-index-based irrigation monitoring approach previously validated at this site to continuous, multi-site operational deployment.

Future work includes extensions to incorporate additional satellite missions, advanced gap-filling methods, and integration with IoT sensor data streams within the SMART project platform.

Acknowledgments

This work was funded by the São Paulo Research Foundation (FAPESP) under grants #2025/11944-7 and #2024/15527-9 (SMART project)".

References

1. Segarra, J., Buchaillot, M. L., Araus, J. L., & Kefauver, S. C. (2020). Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. Agronomy, 10(5), 641. https://doi.org/10.3390/agronomy10050641
2. D.Kovács, D., Musial, J., Bojanowski, J. et al. Copernicus Data Space Ecosystem establishes public cloud processing for earth observation data. Sci Data 13, 537 (2026). https://doi.org/10.1038/s41597-026-06765-8
3. Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y., & Gong, P. (2022). An Overview of the Applications of Earth Observation Satellite Data: Impacts and Future Trends. Remote Sensing, 14(8), 1863. https://doi.org/10.3390/rs14081863
4. A Huete, K Didan, T Miura, E.P Rodriguez, X Gao, L.G Ferreira, (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sensing of Environment, Volume 83, Issues 1–2, Pages 195-213, ISSN 0034-4257, https://doi.org/10.1016/S0034-4257(02)00096-2.
5. Bo-cai Gao, (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sensing of Environment, Volume 58, Issue 3, Pages 257-266, ISSN 0034-4257, https://doi.org/10.1016/S0034-4257(96)00067-3.
6. Er-Raki, S., & Chehbouni, A. (2023). Remote Sensing in Irrigated Crop Water Stress Assessment. Remote Sensing, 15(4), 911. https://doi.org/10.3390/rs15040911
7. Ahmad, U., Alvino, A., & Marino, S. (2021). A Review of Crop Water Stress Assessment Using Remote Sensing. Remote Sensing, 13(20), 4155. https://doi.org/10.3390/rs13204155
8. Ferreira, E., Speranza, E., Vaz, C., Neto, A. T., Bassoi, L., Rabello, L., ... & Kamienski, C. (2025). Estimating Vegetation Response Lag to Soil Moisture in Irrigated Coffee Using Porous-Matrix Sensor and Satellite Data. In 2025 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor) (pp. 412-417). IEEE. https://doi.org/10.1109/MetroAgriFor66923.2025.11512614
9. Wang, Y., Zhang, Z., Feng, L., Ma, Y., & Du, Q. (2021). A new attention-based CNN approach for crop mapping using time series Sentinel-2 images. Computers and electronics in agriculture, 184, 106090. https://doi.org/10.1016/j.compag.2021.106090
10. SMART Project. (2024). Smart Monitoring and Automation for Rural Technologies: IoT-continuum platforms for precision agriculture in Brazil and China. Funded by FAPESP and NSFC.
11. Pelletier, C., Webb, G. I., & Petitjean, F. (2019). Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. Remote Sensing, 11(5), 523. https://doi.org/10.3390/rs11050523
12. Maciel, D. A., Silva, V. A., Alves, H. M. R., Volpato, M. M. L., Barbosa, J. P. R. A. D., Souza, V. C. O. D., ... & Oliveira dos Santos, J. (2020). Leaf water potential of coffee estimated by landsat-8 images. Plos one, 15(3), e0230013. https://doi.org/10.1371/journal.pone.0230013
13. Parreiras, T. C., Santos, C. d. O., Bolfe, É. L., Sano, E. E., Leandro, V. B. S., Bayma, G., Silva, L. A. P. d., Furuya, D. E. G., Romani, L. A. S., & Morton, D. (2025). Dense Time Series of Harmonized Landsat Sentinel-2 and Ensemble Machine Learning to Map Coffee Production Stages. Remote Sensing, 17(18), 3168. https://doi.org/10.3390/rs17183168
14. Vasconcelos, J. C. S., Arantes, C. S., Speranza, E. A., Antunes, J. F. G., Barbosa, L. A. F., & Cançado, G. M. D. A. (2025). Predicting sugarcane yield through temporal analysis of satellite imagery during the growth phase. Agronomy, 15(4), 793.
15. Tsardanidis, I., Koukos, A., Sitokonstantinou, V., Drivas, T., & Kontoes, C. (2025). Cloud gap-filling with deep learning for improved grassland monitoring. Computers and Electronics in Agriculture, 230, 109732. https://doi.org/10.1016/j.compag.2024.109732
16. Maleki, R., Wu, F., Qu, G., Oubara, A., & Yang, G. (2025). Enhancing satellite image compositing with temporal proximity weighting for deep learning–based cropland segmentation. International Journal of Applied Earth Observation and Geoinformation, 143, 104804. https://doi.org/10.1016/j.jag.2025.104804
17. SAT-ML. (2026). Satellite Imagery to Machine Learning Datasets — source code repository. https://github.com/heideker/SAT-ML