Generation of Spatially and Temporally Fine-Resolution Imagery Using STF Algorithms and CACAO Post-Processing

Jaejun Gou
Seoul National University
Dongwon Kang
Seoul National University
Hyeokjin Lee
Seoul National University
Seongju Jang
Seoul National University
Inhong Song
inhongs@snu.ac.kr
Seoul National University

Article

Keywords: Consistent Adjustment of the Climatology to Actual Observations (CACAO), Remote sensing, Satellite, Spatial Temporal Fusion (STF), UAV, NDVI

Posted Date: May 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9501632/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on June 30th, 2026. See the published version at https://doi.org/10.1038/s41598-026-60408-x.

Generation of Spatially and Temporally Fine-Resolution Imagery Using STF Algorithms and CACAO Post-Processing

Jaejun Gou[a], Dongwon Kang[b], Hyeokjin Lee[a], Seongju Jang[a], Inhong Song[c*]

[a]*Department of Rural Systems Engineering, Global Smart Farm Convergence Major, Seoul National University, Seoul 08826, South Korea*
[b]*Department of Rural Systems Engineering, Seoul National University, Seoul 08826, South Korea*
[c]*Department of Rural Systems Engineering, Global Smart Farm Convergence Major, Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, South Korea*
* *Corresponding author. Email: inhongs@snu.ac.kr*

Abstract

Spatio-temporal fusion (STF) has been widely used across various remote sensing applications, including environmental monitoring, land cover change detection, and water resource management by integrating multi-sensor data with different spatial and temporal resolutions. The objective of this study was to generate spatially and temporally fine-resolution imagery and to evaluate the performance of multiple STF algorithms and Consistent Adjustment of the Climatology to Actual Observations (CACAO) post-processing for parcel-level crop monitoring. The study was conducted in a soybean field located in Anseong, South Korea, using Planet SuperDove satellite imagery which has 3 m spatial resolution and near-daily temporal resolution, and Phantom 4 Multispectral Unmanned Aerial Vehicle (UAV) data which has 0.05 m spatial resolution, downscaled to target resolution 0.5 m, and 1 – 4 week irregular temporal resolution. Relative radiometric normalization was applied, followed by the implementation and comparison of four STF algorithms— Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM), Fitting, spatial Filtering and residual Compensation (FitFC), Flexible Spatiotemporal Data Fusion (FSDAF), and Variation-based Spatiotemporal Data Fusion (VSDF)—within a 4-fold cross-validation framework. CACAO post-processing was then employed to reconstruct temporally continuous NDVI trajectories, from which growth metrics such as Vegetation Growth Metrics (VGM)85 and VGMmax were derived. The validation results indicated that ESTARFM achieved the highest Normalized Difference Vegetation Index (NDVI) performance among the evaluated algorithms, with an Root Mean Square Error (RMSE) of 0.113 and the Universal Image Quality Index (UIQI) of 0.697, and CACAO further improved the results, with CA-ESTARFM providing the highest NDVI accuracy, with an RMSE of 0.108 and a UIQI of 0.740. NDVI histogram and spatial analyses demonstrated that CA-ESTARFM achieved the most consistent agreement with UAV observations while preserving fine-scale spatial heterogeneity. In addition, intra-field vegetation assessment using VGM85 and VGMmax confirmed that CA-ESTARFM enhanced the reliability of crop growth evaluation compared to simple linear interpolation of UAV observations. The proposed framework demonstrates strong potential for applications in comprehensive crop monitoring, precision agriculture management, and yield forecasting.

Keywords: Consistent Adjustment of the Climatology to Actual Observations (CACAO), Remote sensing, Satellite, Spatial Temporal Fusion (STF), UAV, NDVI

1. Introduction

Satellite remote sensing is widely used in precision agriculture to generate crop growth time series for diagnosing crop status and predicting yield. Shammi and Meng (2021) developed time-series crop growth metrics from MODerate resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) data and applied them to soybean yield modeling, suggesting phenology-based metrics that capture crop development and yield estimation. Radočaj and Jurišić (2025) evaluated multiple Sentinel-2 vegetation indices and phenological metrics against ground-based maize and soybean yield data, identifying index–phenology combinations for cropland suitability assessment. Recently, PlanetScope (hereafter referred to as Planet) imagery has been widely used since it offers near-daily observations at approximately 3 m spatial resolution. Somerscales and Chen (2026) utilized Planet time series to map crop presence and growth stages in smallholder farming systems, achieving overall accuracies exceeding 85% for crop presence classification and distinguishing key growth stages across fragmented and heterogeneous fields. However, low spatial resolution of satellite data has limitation for detailed analysis within field.

Unmanned Aerial Vehicle (UAV) remote sensing provides centimeter-level spatial resolution, enabling detailed observation of within-field crop variability. Lacerda et al. (2022) utilized UAV-based thermal imagery to estimate crop water status, reporting relationships between Crop Water Stress Index (CWSI) and predawn leaf water potential. Xiao et al. (2023) utilized UAV-based multispectral imagery and canopy height models to assess corn growth performance, reporting correlations between UAV-derived variables and agronomic parameters such as biomass and plant height. Despite these strengths in spatial detail, UAV observations are constrained by limited temporal frequency, as data acquisition depends on field operation and repeated flight missions during the growing season.

In overcoming the trade-off between temporal and spatial resolution of UAV and satellite, Spatio-Temporal Fusion (STF) is one of the most effective method to combine temporally frequent but spatially coarse imagery with spatially fine but temporally infrequent imagery. Early STF studies mainly focused on fusing MODIS and Landsat data, since MODIS provides global observations every 1-2 days and includes 250 m surface reflectance products, whereas Landsat provides 30 m observations with a 16-day repeat period (Feng et al., 2006; Li et al., 2023; Xu et al., 2022). Feng et al. (2006) proposed Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) to predict daily Landsat-like surface reflectance by blending Landsat and MODIS observations. In recent studies, STF applications have extended to UAV–satellite integration (Li et al., 2023; Zhang & Li, 2025).

STF approaches can generally be categorized into weighting-based, unmixing-based, learning-based, Bayesian-based, and hybrid methods (Bazrafkan et al., 2025; Li et al., 2023; Xu et al., 2022; Zhu et al., 2016). Weighting-based methods, such as STARFM, estimate fine-resolution reflectance by weighting spectrally similar neighboring pixels, assuming consistent reflectance changes across spatial scales (Feng et al., 2006). Enhanced STARFM (ESTARFM) introduces conversion coefficients and multi-temporal weighting schemes to better handle heterogeneous landscapes and mixed pixels (Zhu et al., 2010). Wang and Atkinson (2018) developed regression model Fitting, spatial Filtering and residual Compensation (FitFC) which improves temporal prediction accuracy by incorporating local linear regression, spatial filtering, and residual compensation mechanisms. These methods are computationally efficient and widely applied, but may show limitations in representing complex spatial heterogeneity and non-linear temporal changes (Xu et al., 2022; Zhu et al., 2016).

Other methodological frameworks have also been developed to address these challenges. Unmixing-based methods estimate fine-resolution reflectance by decomposing coarse pixels into endmember fractions based on spectral mixture theory (Bazrafkan et al., 2025; Zurita-Milla et al., 2008). Learning-based approaches utilize data-driven models, including convolutional neural networks and generative adversarial networks, to learn relationships between coarse and fine imagery (Ma et al., 2025; Xiao et al., 2023). Bayesian-based methods formulate the fusion process as a probabilistic estimation problem, enabling the integration of prior information and multi-source observations while accounting for uncertainty (Huang et al., 2013; Liao et al., 2016).

Hybrid methods have been proposed to integrate complementary strengths of these approaches. Zhu et al. (2016) developed the Flexible Spatiotemporal Data Fusion (FSDAF) method, which integrates spectral unmixing, thin plate spline interpolation, and neighborhood-based weighting. Xu et al. (2022) developed the Variation-based Spatiotemporal Data Fusion (VSDF) method, which integrates variation classification, guided filtering, residual distribution, and edge fusion. Hybrid methods improve performance by integrating complementary strengths of multiple approaches, enabling reliable image reconstruction under heterogeneous conditions and abrupt land cover changes (Swain et al., 2024).

Corresponding to the development of STF methods, several studies introduced STF to generate high spatial, temporal resolution images for agricultural and environmental monitoring. Meng et al. (2013) generated high spatial and temporal resolution NDVI maps using modified STARFM to MODIS and HJ-1 images, and applied them to winter wheat biomass estimation. Xiao et al. (2022) employed ESTARFM to integrate Sentinel-2 and MODIS imagery, to analyze the start, end, duration, and water demand of irrigation in paddy fields. Xu et al. (2024) evaluated five STF algorithms, STARFM, ESTARFM, FSDAF, RASTFM, and UBDF, using Landsat and MODIS imagery in monitoring inland water dynamics, including water body extraction, algae bloom detection, seasonal variations, and quantitative water quality analysis. However, these studies have primarily focused on regional-scale applications or algorithm comparisons, and the direct comparison of STF and temporal interpolation methods at the intra-field scale has not been fully explored.

Some studies introduced Consistent Adjustment of the Climatology to Actual Observations (CACAO) post processing, developed by Verger et al. (2013), which interpolates the temporal variation and makes a reflectance correction for smoothing phenological curves. Yin et al. (2017) employed the CACAO method to integrate MODIS and Landsat data to reconstruct temporally continuous NDVI and generate gap-free Leaf Area Index (LAI) maps. Yu et al. (2021) proposed a climate-integrated gap-filling method that reconstructs continuous Landsat NDVI time series by integrating climate variables and adjusting the predicted phenological trajectories using the CACAO-based post-processing. However, these studies did not employ STF techniques for satellite and UAV, within-field crop growth assessment. Li et al. (2023) proposed an STF framework that integrates UAV and Sentinel-2 imagery with CACAO-based phenological reconstruction to produce daily high-resolution data for winter wheat growth monitoring. However, this study did not directly compare the performance of various STF algorithms against CACAO, and did not extend to further analyses such as evaluating growth indicators within the field. Furthermore, while numerous advanced and hybrid STF algorithms have been developed, their actual effectiveness remains unverified in simplified field environments consisting of vegetation and bare soil, which are the two primary components driving crop phenology analysis.

The objective of this study is generating fine UAV-like images (target resolution 0.5 m), using coarse Planet images, evaluating and comparing the performance of two weighting-based and two hybrid STF methods, ESTARFM, fitFC, FSDAF, VSDF, and CACAO post processing in a simplified soybean field environment consisting vegetation and bare soil classes, aiming to accurately capture within-field crop growth variability.

2. Data and Methods

2.1 Study area

The study area is located in east side of Anseong city, Gyeong-gi province, South Korea (Fig. 1). The region, located at 37°N latitude, lies within a temperate monsoon climate zone, characterized by warm, humid summers and cold winters. The Region of Interest (ROI) is small-scale rectangular shaped field, which is approximately 100 m in width and 50 m in length. The ROI was originally used as paddy fields for rice cultivation, has recently been converted into upland fields for soybean cultivation. Since paddy fields have limited drainage ability due to its clay based soil environment, underground drainage pipe is installed for drainage enhancement. Soybean was sown on June 23rd and harvested on November 1st.

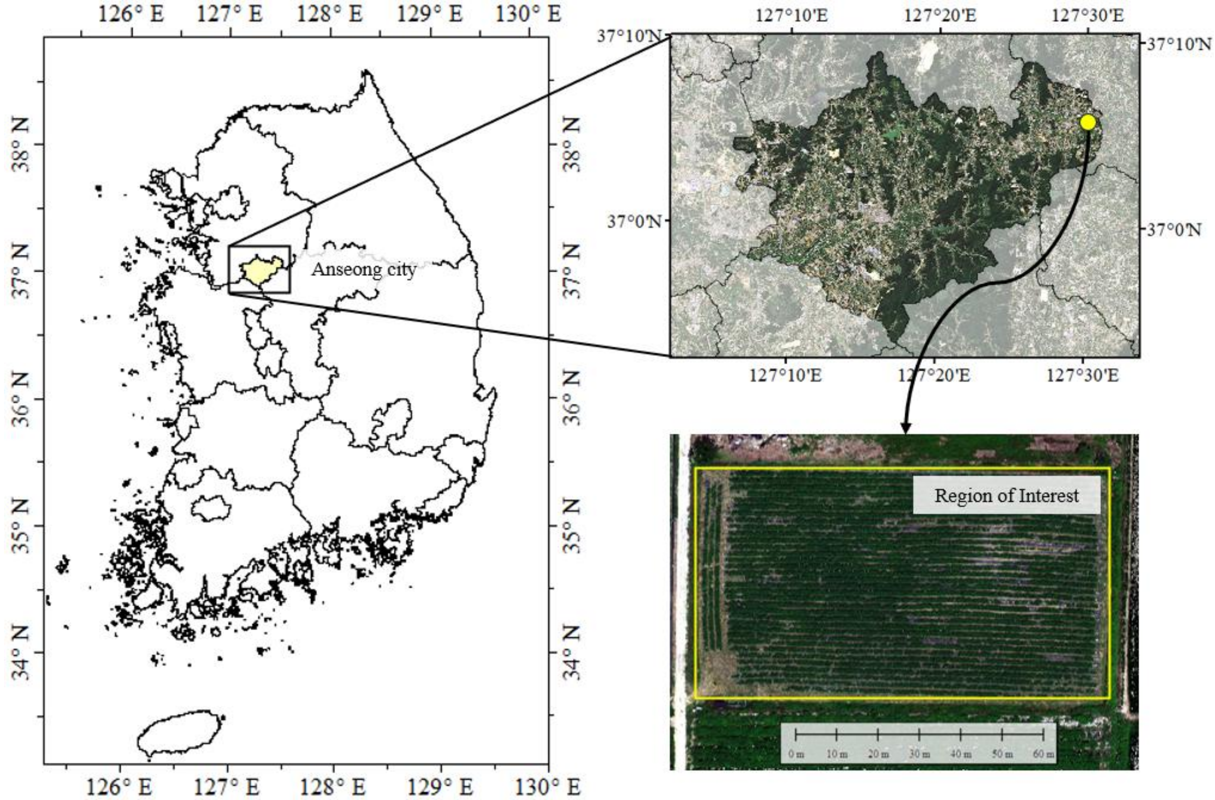


Fig. 1 Geological details of the study area, located in east side of Anseong city. The yellow box indicates the ROI (Region Of Interest),

2.2 Data Acquisition and Experiment Methods

Planet SuperDove (Planet Labs, USA), and Phantom 4 Multispectral (DJI, China) were used in this study to provide coarse and fine resolution imagery, respectively. Planet SuperDove is a constellation of high-resolution Earth observation satellites operated by Planet Labs, providing daily global imagery at approximately 3 m spatial resolution across eight spectral bands. Spectral details are represented at Table 1. Surface reflectance imagery, atmospherically corrected by Planet Labs, with UDM2, which is usable data mask for cloud, shadow, haze, and other quality flags, is available through Planet Explorer (https://www.planet.com/explorer/).

Fine resolution UAV-based imagery was acquired using a Phantom 4 Multispectral, which captures data in five spectral bands. Exact spectral details are represented at Table 1. UAV data acquisition was conducted eight times from transplanting (June 30) to just before harvest (October 23), and the exact acquisition dates are shown in Fig. 2. The flights were conducted at a ground sampling distance of 0.05 m, with a flight speed of 5 m s^{-1}, and image overlaps of 75% in the forward direction and 60% in the side direction. A total of 5 Ground Control Points (GCPs), each measuring 1 m × 1 m and distributed around the boundary of the field, were deployed for accurate georeferencing. The geographic coordinates of the GCPs were surveyed using an RTK-enabled GPS system (GeoXR, Trimble, USA). Image processing and orthomosaic generation were performed using Pix4Dmapper (version 4.8.4). While the spatial resolution of generated UAV-based imagery is 0.05 m, we downscaled to 0.5 m, considering planting interval of soybean (0.4 – 0.6 m) and stability of the model performance.

Table 1 Spectral details of Planet Superdove satellite and Phantom 4 Multispectral drone. Wavelength width of Planet Superdove indicates Full Width at Half Maximum (FWHM)

Band number	Band name	Central wavelength (nm)	Wavelength width (nm)
Planet Superdove	Coastal Blue	443	20
	Blue	490	50
	Green I	531	36

	Green	565	36
	Yellow	610	20
	Red	665	31
	Red Edge	705	15
	NIR	865	40
Phantom 4 Multispectral	Blue	450	32
	Green	560	32
	Red	650	32
	Red Edge	730	32
	NIR	840	32

Temporal distribution of Planet and UAV datasets is represented in Fig. 2. UAV imagery has a lower temporal resolution than Planet data because acquisitions require site-specific field campaigns under suitable weather and logistical conditions, and are also influenced by researchers' schedules. Whereas Planet satellites provide automated, near-daily global coverage. Total 24 Planet images and 8 UAV images are acquired, and 6 images pairs are selected as anchor pairs among them, which captured almost same date that regarded having temporal consistency.

The validation was performed using a 4-fold cross-validation approach. For example, the fine-resolution image on July 7 was treated as unavailable, and a corresponding fine image was predicted from the coarse image on that date using the remaining anchor pairs. The prediction accuracy was then evaluated against the actual observation, and this procedure was repeated for the remaining validation anchor pairs. Initial and terminal anchor pairs were excluded from validation, since accuracy was highly degraded in case of extrapolation.

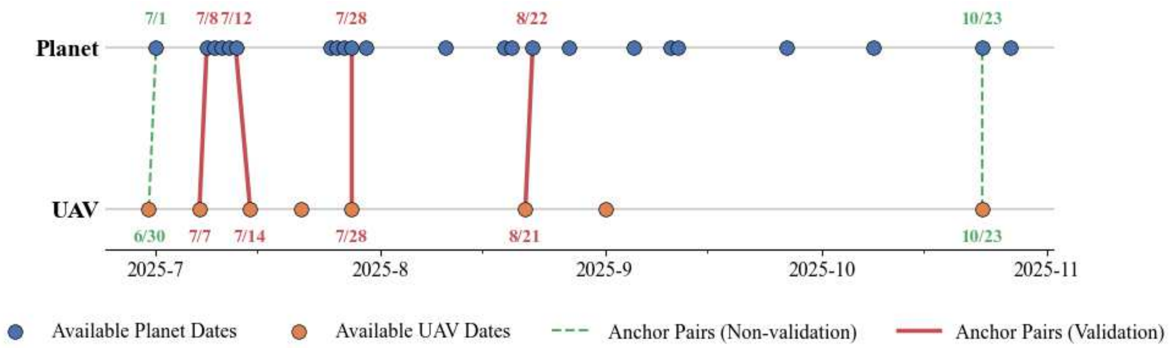


Fig. 2 Temporal distribution of the Planet and UAV datasets. Only cloud-free Planet images were selected as available Planet dates. Anchor pairs refer to Planet–UAV image pairs acquired within ±2 days of each other, and they play an important role in the application of STF algorithms.

2.3 Relative Radiometric Normalization

Differences in sensor characteristics, spectral response functions, and atmospheric conditions, as seen in Table 1, can significantly degrade the accuracy of STF results (Li et al., 2023). To address this issue, Relative Radiometric Normalization (RRN) is applied to align the radiometric properties of multi-source imagery prior to fusion. RRN is applied based on Eq. (1) – (4)

$$F(x) = a(x) \cdot C(x) + b(x) \tag{1}$$

$$a(x) = \frac{\sum_{i \in \Omega(x)} (C_i - \bar{C})(F_i - \bar{F})}{\sum_{i \in \Omega(x)} (C_i - \bar{C})^2 + \epsilon} \tag{2}$$

$$b(x) = \bar{F} - a(x)\bar{C} \tag{3}$$

$$C^{\text{norm}}(x) = a(x) \cdot C(x) + b(x) \tag{4}$$

Where, $C(x)$ and $F(x)$ denote the coarse-resolution (satellite) and fine-resolution (UAV) reflectance at pixel x, respectively, $\Omega(x)$ represents a local window centered at pixel x, $\bar{C}$ and $\bar{F}$ are the mean reflectance values within $\Omega(x)$, $a(x)$ and $b(x)$ are the locally estimated regression coefficients corresponding to slope and intercept, and ϵ is a small constant introduced to ensure numerical stability.

As described in Eq. (1), the reflectance relationship between UAV and satellite imagery is modeled as a linear function, where the fine-resolution reflectance is expressed as a transformed version of the coarse-resolution reflectance. The slope $a(x)$, calculated using local covariance as shown in Eq. (2), captures the relative scaling between the two datasets, while the intercept $b(x)$, defined in Eq. (3), accounts for systematic offsets. By applying these locally adaptive parameters, the normalized coarse reflectance $C^{\text{norm}}(x)$) is obtained through Eq. (4), effectively reducing radiometric discrepancies while preserving spatial variability, which is essential for reliable spatiotemporal fusion.

RRN was implemented based on the anchor pairs. First, for each anchor pair consisting of temporally consistent Planet–UAV observations, the radiometric relationship was established under similar acquisition conditions. These coefficients were derived exclusively from the selected anchor pairs to ensure stability and reliability. For dates without corresponding UAV observations, the radiometric relationships obtained from the temporally adjacent anchor pairs were propagated and combined using a temporally weighted scheme, allowing the entire coarse-resolution time series to be normalized to the UAV radiometric domain.

2.4 STF algorithms and CACAO post processing validation

Among the five category of STF methods, two weighting-based method (ESTARFM, FitFC) and two hybrid STF methods (FSDAF, VSDF) which are computationally effective and verified by various previous studies are validated in this study (Bazrafkan et al., 2025; Li et al., 2023; Xu et al., 2022). In addition, CACAO post processing, developed by Verger et al. (2013), was conducted for each STF method to validate the improvement of CACAO.

The validation process of the STF and CACAO post processed STF (CA-STF) was described in Fig. 3. For each fold, preliminary fine images P_k were first generated at the target coarse acquisition date using the STF algorithm and the remaining anchor pairs. For the CA-STF validation, these preliminary predictions were further refined by applying pixel-wise scale and temporal shift parameters estimated from all fine images except those in the validation fold, generating CP_k. The resulting CA-STF predictions were then compared with the corresponding UAV observations C_k for the same validation dates.

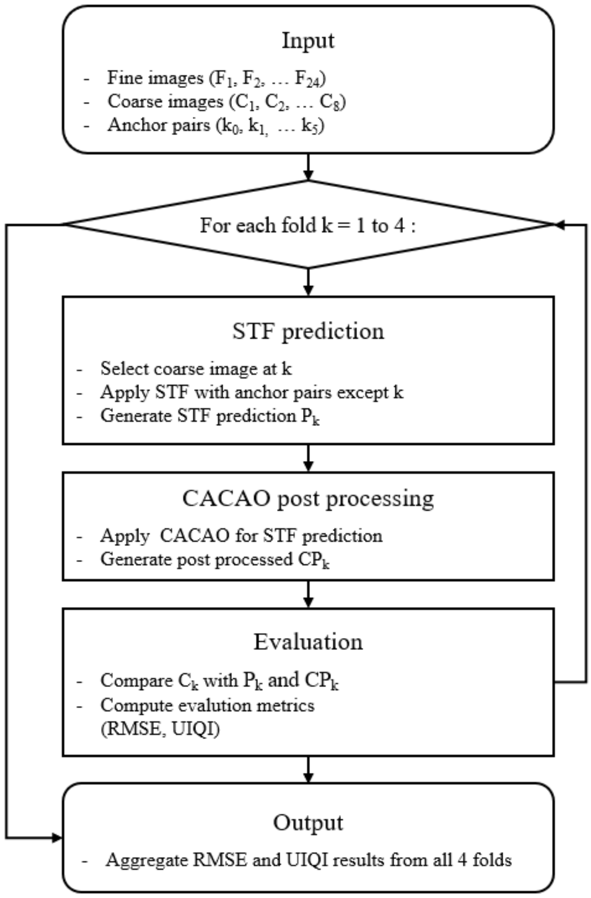


Fig. 3 Workflow of 4-fold cross-validation for STF and CA-STF. Fine images and coarse images indicates UAV and Planet images respectively, and anchor pairs indicates the fine and coarse pairs which are described in Fig. 2

2.4.1 STF algorithms.

ESTARFM, FitFC, FSDAF, and VSDF were applied to generate fine-resolution imagery at the target date by integrating coarse and fine observations using different modeling strategies. ESTARFM estimates fine-resolution reflectance using spectrally similar neighboring pixels, multi-temporal anchor pairs, and a spatial–spectral–temporal weighting scheme (Zhu et al., 2010). FitFC generates fine-resolution imagery by integrating regression modeling (RM), spatial filtering (SF) based on similar pixels, and residual compensation (RC) between coarse and fine images (Wang & Atkinson, 2018). FSDAF reconstructs fine-resolution imagery through class-based temporal change modeling, coarse-to-fine residual redistribution, and thin plate spline interpolation (Zhu et al., 2016). VSDF models fine-resolution reflectance using Relative Reliability Index (RRI), guided change images, variation classification, residual distribution, and edge fusion (Xu et al., 2022).

Algorithms except ESTARFM were originally designed to predict the fine-resolution image at the target date t_2 using a single anchor pair at t_1, together with the coarse-resolution image at t_2. However, in this study, to maintain consistency with ESTARFM and enhance performance of models, the methods were extended to a two-anchor framework by additionally generating a prediction from the subsequent anchor pair at t_3 and combining the two predictions using temporal weighting.

Each algorithms share common basic parameters, which are window size, number of similar pixels, and number of classes. The window size defines the half-size of the moving window, larger window captures more spatial heterogeneity and incorporate neighboring pixel information. The number of similar pixels refers to the number of spectrally similar neighboring pixels used in the prediction, larger number of similar pixels improves robustness by reducing noise but may lead to oversmoothing and loss of fine spatial variability. The number of classes defines level of spectral grouping in the image. FitFC has two types of window size, one for RM stage and one for SF and RC stage, while doesn't need to designate number of classes since it doesn't have clustering process. FSDAF has two types of number of similar pixels, which are minimum class and maximum class, since it uses ISODATA clustering. VSDF does not require a predefined number of classes because it employs an abundant variation classification (AVC) scheme, in which the number of clusters is adaptively determined based on spatiotemporal variation rather than fixed land-cover categories. Other parameters for each algorithms were set to default values. The parameters used and code source in this study are represented in Table 2

Table 2 The parameter details and code source of algorithms used in this study. Each parameter are selected considering target resolution and ROI environment of this study.

	Window size	Number of similar pixels	Number of classes	Code source
ESTARFM	15	20	2	https://xzhu-lab.github.io/post/open-source-code/
FitFC	3 (RM), 15 (SF, RC)	20	2	https://github.com/qunmingwang/Fit-FC
FSDAF	15	20	2 (min), 3 (max)	https://xzhu-lab.github.io/post/open-source-code/
VSDF	15	20	None	https://github.com/ChenXuAxel/VSDF

2.4.2 CACAO post processing

The CACAO method is a temporal smoothing and gap-filling approach based on pixel-wise phenology modeling of satellite time series. The time series is smoothed using a Savitzky–Golay filter with a temporal window of 17 days and a second-order polynomial to construct a continuous phenological trajectory. The phenology model is subsequently adjusted to the available observations by estimating temporal shift and amplitude scaling parameters that minimize the difference between the modeled values and observations. This fitting is conducted independently for each pixel, and the adjusted phenology model is used to reconstruct continuous time-series data and fill missing observations.

In this study, we used CACAO to assess the effectiveness of phenology-based correction and its interpolation capability. The input time series for CACAO was constructed by combining UAV observations with preliminary NDVI predictions for all Planet dates derived from STF algorithms. The CACAO fitting was performed using all available UAV observations, while the image at the target date was excluded from the fitting process.

2.5 NDVI curve generation and vegetation assessment

NDVI, calculated using Eq. (5), is one of the most powerful metrics for vegetation vitality assessment, yield prediction, and evapotranspiration estimation. In this study, we applied the STF method which recorded the highest performance in the validation process generating fine images for all available Planet dates. Then, we applied CACAO post-processing, reconstructing and interpolating temporal gaps between synthesized fine images. By these process, we generated daily NDVI curve for each pixels which represents the phenology of each single crops.

$$NDVI = \frac{NIR-RED}{NIR+RED} \tag{5}$$

Several studies have been conducted to explore the correlation between NDVI and crop growth and to develop NDVI-based metrics. In this study, we applied Vegetation Growth Metrics (VGM)85 and VGMmax, which were reported as effective for yield prediction in the study by Shammi and Meng (2021). VGM85 is the NDVI value at the 85-day growth stage, calculated using Eq. (6), and VGMmax is the maximum NDVI value during the entire growth stage, calculated using Eq. (7).

$$VGM_{85} = \frac{1}{n}\sum_{51 \le DAP_i \le 85} NDVI_i \tag{6}$$

$$VGM_{max} = \max_{i=start}^{end}(NDVI_i) \tag{7}$$

Where, DAP represents Days After Planting.

2.6 Evalutation

To evaluate the performance of the STF models, we employed two complementary metrics, Root Mean Square Error (RMSE) and the Universal Image Quality Index (UIQI). While RMSE quantitatively measures the absolute spectral error at the pixel level, UIQI assesses the structural similarity and visual quality based on the correlation between pixels.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (8)$$

$$UIQI = \left(\frac{\sigma_{xy}}{\sigma_x \sigma_y}\right) \cdot \left(\frac{2\overline{x}\overline{y}}{\overline{x}^2 + \overline{y}^2}\right) \cdot \left(\frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right) \quad (9)$$

Where n is the total number of pixels, y_i and $\hat{y}_i$ are the reference and predicted pixel values, $\bar{x}$ and $\bar{y}$ denote the mean values of the reference and predicted images, σ_x^2 and σ_y^2 represent their respective variances, and σ_{xy} is the covariance between the two images.

3. Result and Discussion

3.1 Accuracy assessment of STF and CACAO-STF

Table 3 summarizes the validation results of STF algorithms, prior to CACAO application. The RMSE values in the NIR band were relatively higher because its absolute reflectance values are generally larger than those of the RGB bands. Meanwhile, NIR band has the highest UIQI compared to other bands. Since NIR reflectance is governed by vegetation structures, its spatial patterns tend to remain highly stable. Furthermore, the absolute reflectance values in the NIR band are higher than those in the visible spectrum, and the STF predictions preserve this dynamic range, leading to superior structural similarity.

Among different algorithms, ESTARFM achieved the highest overall accuracy. In a simplified field environment consisting exclusively of soil and vegetation, the similar-pixel-based weighting approach employed by ESTARFM appears to be more effective than the classification-based or reliability-based change detection mechanisms utilized by FSDAF and VSDF. This observation aligns with previous studies, which reported that ESTARFM provides stable and robust fusion results (Xu et al., 2024). FitFC, however, recorded lower accuracy especially on UIQI. FitFC relies on simplified linear relationships between coarse and fine images, leading to reduced correlation and contrast consistency with the reference observations (Xu et al., 2024; Zhu et al., 2016). VSDF recorded higher accuracy than FSDAF, can be attributed to its RRI-based framework, which enables pixel-wise assessment of temporal changes and reduces dependence on land-cover classification (Xu et al., 2022). Furthermore, VSDF incorporates reliability-guided local correction and edge-preserving mechanisms, resulting in improved preservation of spatial details and even higher UIQI score for NIR and RedEdge than ESTARFM (Xu et al., 2022).

Table 3 RMSE and UIQI metrics of the STF algorithms. All algorithms were validated using a 4-fold cross-validation approach on anchor pairs.

Band	ESTARFM		FitFC		FSDAF		VSDF	
	RMSE	UIQI	RMSE	UIQI	RMSE	UIQI	RMSE	UIQI
Blue	0.005	0.547	0.006	0.251	0.006	0.447	0.005	0.559
Green	0.006	0.629	0.008	0.266	0.008	0.492	0.007	0.599
Red	0.009	0.611	0.010	0.317	0.010	0.544	0.009	0.609
RedEdge	0.015	0.605	0.022	0.034	0.016	0.613	0.013	0.697
NIR	0.028	0.673	0.041	0.363	0.041	0.610	0.031	0.698
NDVI	0.113	0.697	0.186	0.408	0.170	0.616	0.148	0.669

Table 4 presents the performance of the STF algorithms after CACAO post-processing, showing an overall improvement in accuracy across all models. The CACAO method minimized radiometric and temporal discrepancies by adjusting pixel-wise scale and temporal shift parameters to the actual UAV observations, effectively reducing temporal noise in agricultural environments.

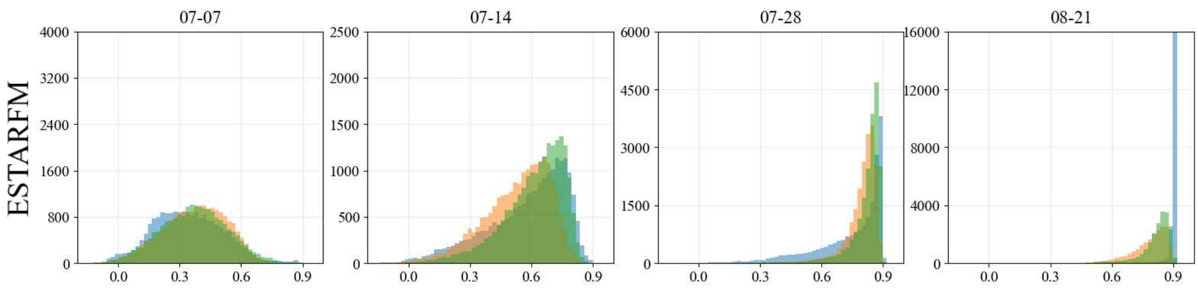
The correction effect was the most notable in the NIR band. Since the NIR band is highly sensitive to temporal changes in canopy structure and LAI during the crop growth cycle, CACAO effectively resolves its fluctuation by reconstructing the underlying seasonal trajectory of vegetation. Comparing the NDVI between different algorithms, CA-ESTARFM achieved the highest accuracy among four CA-STF algorithms, with an RMSE of 0.108 and a UIQI of 0.740. This suggests that the baseline accuracy of the STF algorithm affects the final output of the CACAO processing. CA-FitFC showed a significant accuracy improvement after CACAO application, with its NDVI UIQI increasing from 0.408 to 0.708. These results demonstrates that CACAO can reconstruct reliable phenological trajectories even from less accurate initial fusion results.

Table 4 RMSE and UIQI metrics of CACAO-processed STF algorithms. All algorithms were validated using a 4-fold cross-validation approach on anchor pairs. CA- indicates CACAO post processing application.

Band	CA-ESTARFM		CA-FitFC		CA-FSDAF		CA-VSDF	
	RMSE	UIQI	RMSE	UIQI	RMSE	UIQI	RMSE	UIQI
Blue	0.004	0.566	0.004	0.508	0.005	0.492	0.004	0.567
Green	0.006	0.599	0.007	0.500	0.007	0.535	0.006	0.592
Red	0.008	0.635	0.008	0.605	0.009	0.586	0.008	0.627
RedEdge	0.017	0.597	0.019	0.486	0.013	0.670	0.014	0.706
NIR	0.026	0.714	0.032	0.619	0.032	0.706	0.027	0.741
NDVI	0.108	0.740	0.124	0.708	0.126	0.692	0.121	0.738

Fig. 4 illustrates the NDVI histogram of observed, STF applied, and CA-STF applied images for validation pairs. During the early growth stages (July 7th and 14th), the STF predictions generally aligned with the actual UAV observations especially on ESTARFM. Although slight discrepancies were present in algorithms, the CACAO post-processing effectively mitigated these differences. However, the deviations between the predicted and observed values became more pronounced toward the later growth stages. Specifically, on August 21, robust vegetation activity resulted in a highly skewed distribution with NDVI values approaching 1.0, the baseline STF algorithms demonstrated limitations in accurately reproducing these extreme values.

Comparing among the algorithms, ESTARFM generated histograms that most closely matched the reference data, with the exception of the August 21st observations, corresponding to high accuracy on RMSE and UIQI than other algorithms (Table 3, 4). For FSDAF, while the overall shape of the histograms was similar, there was a noticeable shift in the central tendency, overestimating NDVI on July 28th and underestimating on August 21st. Since FSDAF initially estimates temporal changes from coarse imagery before distributing residuals, the mixed-pixel effect likely reduced the magnitude of temporal variation in the Planet imagery compared to the UAV data, leading to an underestimation of temporal change. Consequently, in predicting July 28th, FSDAF relied more on the subsequent August 21st observation with higher vegetation vitality, generating outputs that resembled this later stage. VSDF exhibited a distinctive bimodal distribution in STF images. As crop growth progresses, bare soil exposure decreases while vegetation cover increases, the RRI mechanism in VSDF failed to adequately reflect this transition, resulting in an overestimation of the soil fraction. These tendency generally relieved by adjusting CACAO post processing. FitFC generated predictions with low inter-pixel variance that were heavily skewed. This indicates that relying solely on simplified linear relationships between coarse and fine images is insufficient to capture the complex variations in vegetation dynamics.



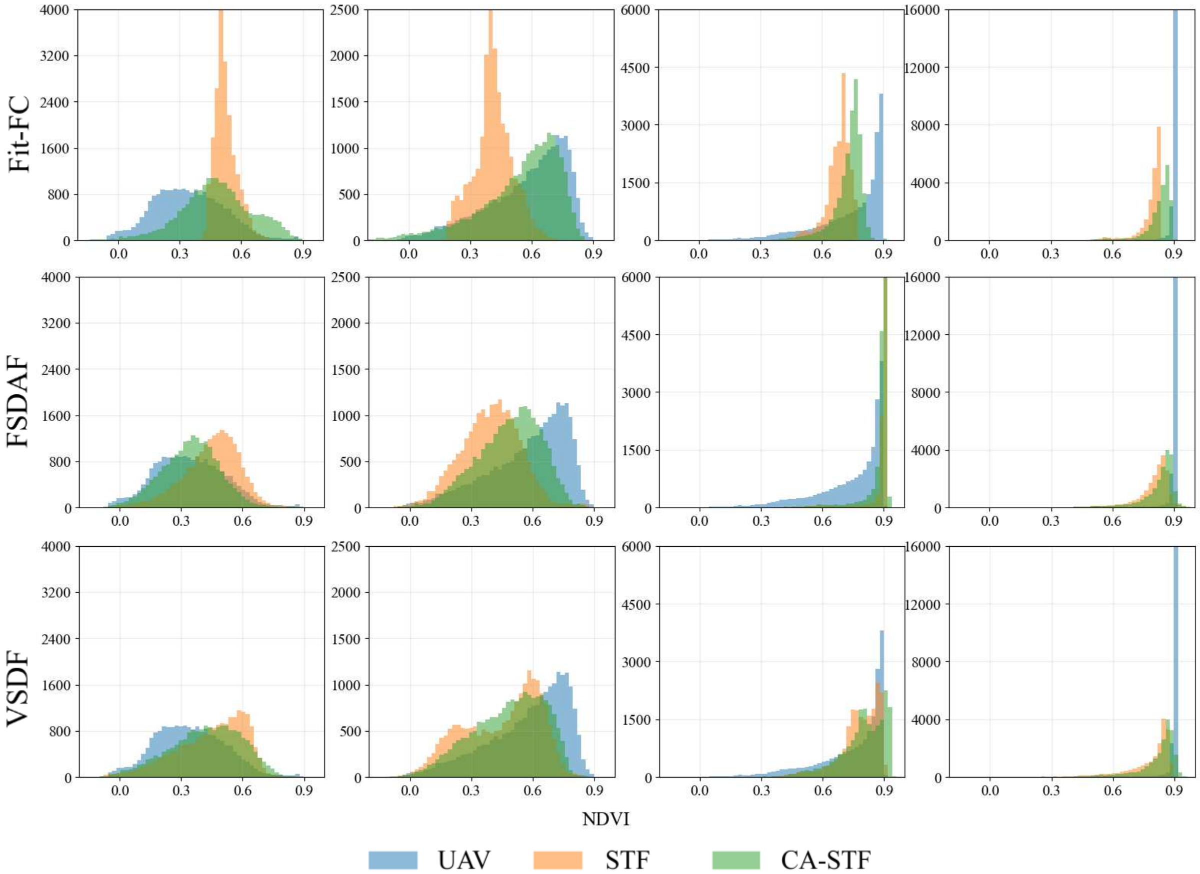
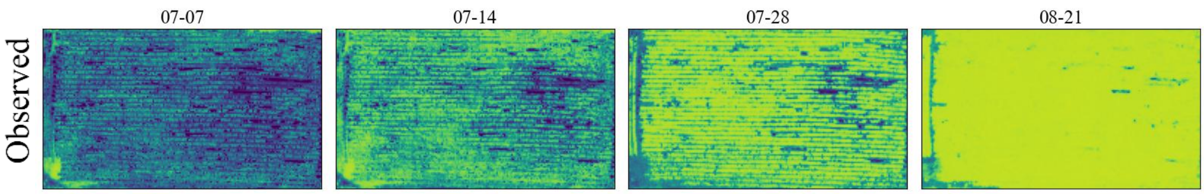


Fig. 4 NDVI histograms of UAV observations, STF predictions, and CA-STF results for four anchor dates across different STF algorithms. CA-STF represents CACAO post-processed NDVI maps.

Fig. 5 displays the spatial NDVI distributions from the four CA-STF algorithms and the UAV observations. Although CACAO post-processing was applied to all outputs, spatial differences among the algorithms remained. This indicates that while CACAO improves radiometric and temporal consistency by scale and shift, it does not change the inherent spatial characteristics of the baseline STF algorithms.

Among the evaluated models, ESTARFM showed the highest spatial agreement with the UAV observations especially on July 7th and 14th, resulting in a low RMSE and a high UIQI in Table 4. Its similar-pixel based weighting effectively reflected gradual phenological changes, while still showing limitation in predicting canopy expansion in the late growing season. Fit-FC generated maps with reduced spatial variability and a relatively uniform NDVI distribution, resulting low UIQI in Table 4, indicating that its regression-based framework tends to oversmooth spatial details. FSDAF made misprediction in the magnitude of NDVI, also confirmed in Fig. 5, likely attributable to the coarse-image-based estimation of temporal change. Furthermore, spatially uniform patterns were observed on July 28th, with reduced intra-field variability compared to the UAV data. This limitation is associated with the class-based framework of FSDAF, which assumes homogeneous temporal dynamics within each class, and may lead to simplified representation of temporal changes in areas with gradual vegetation transitions. VSDF preserved intra-field textures, resulting high UIQI in Table 4, but produced localized noise and overestimated soil signals especially in August 21st. This suggests that the RRI weighting mechanism did not fully capture the transition from bare soil to vegetation canopy.



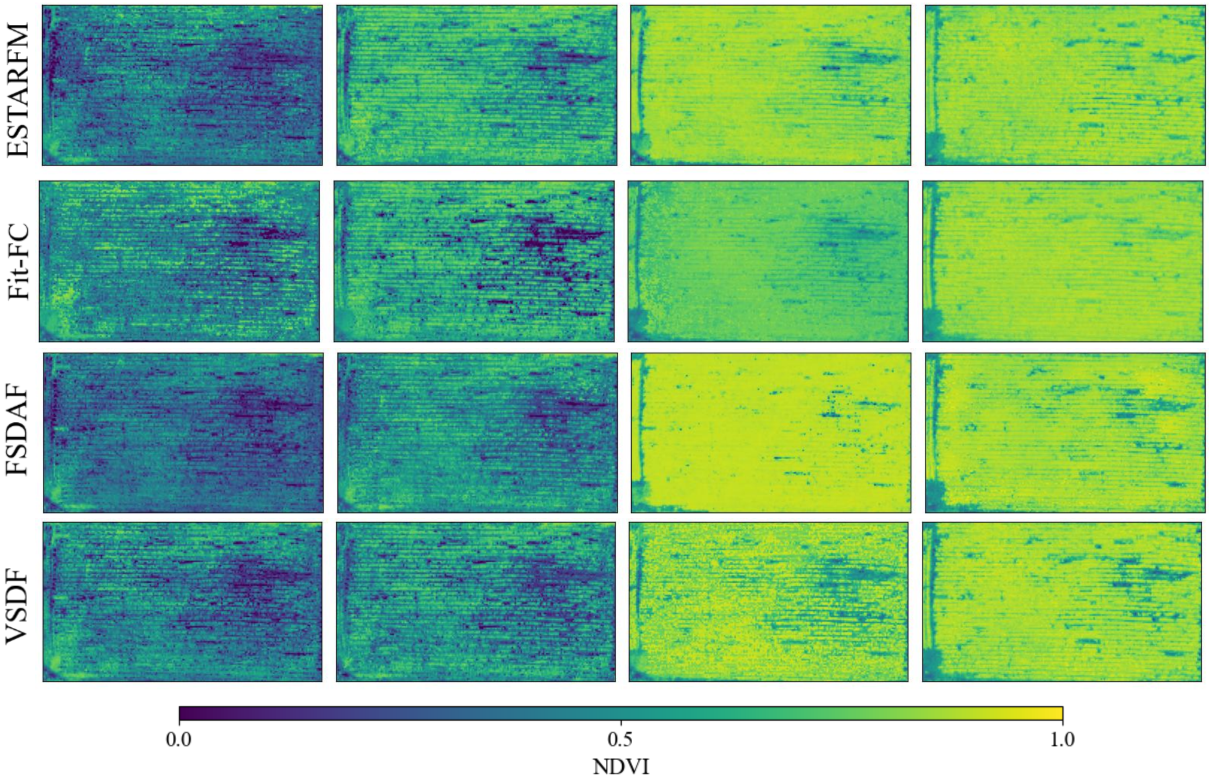


Fig. 5 Spatial comparison of NDVI maps predicted using CA-STF algorithms across four validation dates and the corresponding UAV observations. All NDVI maps were adjusted using CACAO post-processing. Pixels with low NDVI values represent soil, whereas high NDVI values indicate vegetation.

3.2 NDVI curve analysis

Based on validation results using RMSE and UIQI (Table 3 and 4), as well as NDVI patterns (Fig. 4 and 5), ESTARFM showed the most consistent performance among the evaluated models. To further examine continuous intra-field crop growth dynamics, all available UAV and Planet datasets were integrated to generate a daily NDVI trajectory using the CA-ESTARFM framework. Fig. 6 presents the spatially averaged NDVI time series obtained from this pixel-wise analysis. A rapid increase in NDVI was observed from early to mid-July, followed by a plateau during mid-August, and a gradual decline from September to October.

During the early growth stages, the CA-ESTARFM curve closely followed the UAV observations, representing the increase in vegetation vigor. Although the preliminary ESTARFM predictions exhibited minor temporal noise, such as step-like fluctuations during early to mid-July, the CACAO post-processing reduced these inconsistencies, resulting in a continuous NDVI trajectory.

By mid-August, when NDVI values approached saturation, the difference between the CA-ESTARFM trajectory and the preliminary ESTARFM predictions became small. During the late growth stages from September to October, differences were observed between the Planet observations and the ESTARFM predictions. Planet observations showed a narrower range of NDVI variation compared to UAV data, indicating reduced sensitivity to temporal changes. In addition, ESTARFM showed limited ability to capture non-linear phenological transitions during this period, which contributed to the observed differences.

Overall, the CACAO post-processing reduced residual temporal noise and improved the temporal consistency of the NDVI trajectory by adjusting its amplitude and timing based on available UAV observations, enabling the generation of temporally continuous and high-resolution NDVI data throughout the growing season.

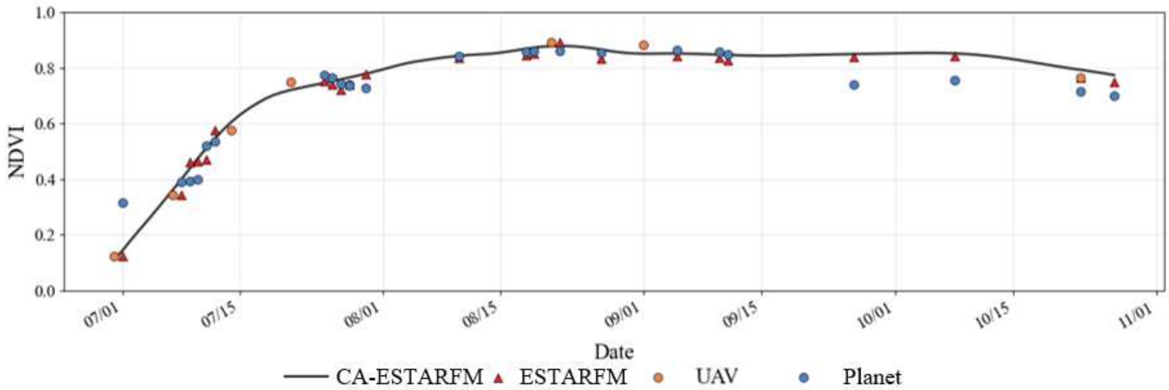


Fig. 6 Temporal variation of averaged NDVI derived from CA-ESTARFM, ESTARFM, UAV observations, and RRN-normalized Planet data. The CA- ESTARFM curve represents the smoothed daily NDVI trajectory, while red triangles indicate preliminary ESTARFM estimates and observation-based NDVI values.

Fig. 6 presents a comparison of the spatial distributions of VGM85 and VGMMAX derived from CA-ESTARFM and linearly interpolated UAV NDVI, along with their difference maps (CA-ESTARFM − UAV). The spatial patterns of CA-ESTARFM and UAV interpolation are generally similar, particularly within the interior of the field where both VGM85 and VGMMAX show relatively uniform distributions. However, the UAV interpolation produces smoother patterns, whereas CA-ESTARFM retains finer spatial variations, including localized zones of relatively lower or higher NDVI.

The difference maps show that discrepancies between the two approaches are generally small within the central field area, with values close to zero. Larger differences are observed along the left boundary and within the regions highlighted by the dotted circles, where both positive and negative deviations appear in VGM85 and VGMMAX. These areas correspond to regions with soil exposure and abrupt spatial transitions, where simple temporal interpolation may not fully represent local variability. By incorporating temporally dense satellite observations and phenology-based adjustments, CA-ESTARFM captures additional spatial variations in these regions, which may improve the representation of spatial transitions compared to UAV-only interpolation.

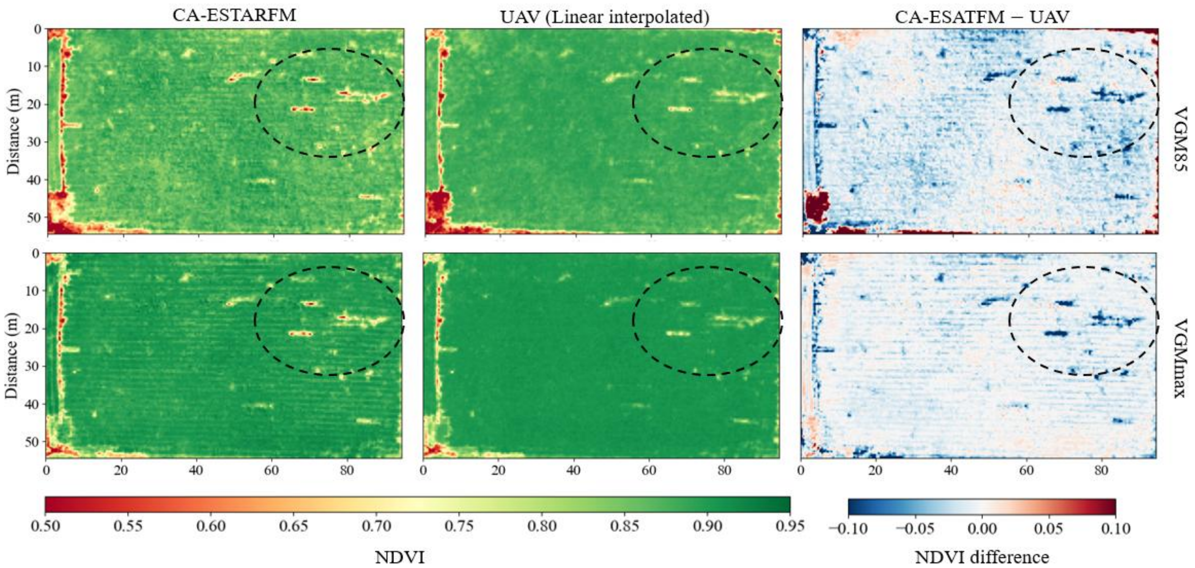


Fig. 7 Spatial comparison of NDVI time series derived metrics, VGM85 and VGMmax, generated from CA-VSDF and linearly interpolated UAV observation. CA-VSDF – UAV represents the subtraction from CA-VSDF to UAV, red pixels indicates that CA-VSDF has higher value and blue pixels indicates that UAV has higher value.

4. Conclusion

This study evaluated the performance of four STF algorithms, ESTARFM, FitFC, FSDAF and VSDF, and CACAO post processing method, in a soybean field composed primarily of vegetation and bare soil. Planet SuperDove imagery was used as the coarse-resolution input, and UAV multispectral imagery, downscaled to target resolution 0.5 m, was used as the fine-resolution reference. RRN was applied reduce radiometric discrepancies between the two sensors. The STF algorithms were assessed using a 4-fold cross-validation, and CACAO post-processing was applied to evaluate the effect of phenology-based temporal correction. Based on the best-performing model, daily NDVI trajectories were reconstructed and further used to derive NDVI-based growth metrics, VGM85 and VGMmax, for intra-field crop growth assessment

The results showed that CA-ESTARFM provided the most consistent performance among the evaluated STF algorithms, recording RMSE 0.108 and UIQI 0.740. CACAO improved the performance of all algorithms by reducing temporal noise and adjusting pixel-wise reflectance. The daily NDVI trajectory generated from CA-ESTARFM followed the observed UAV trend well and produced a smooth phenological curve throughout the growing season. In addition, the spatial distributions of VGM85 and VGMmax derived from CA-ESTARFM were highly consistent with those derived from linearly interpolated UAV observations, while preserving finer intra-field heterogeneity.

The proposed CA-STF framework can be effectively utilized for comprehensive crop monitoring, precision agriculture management, and yield forecasting by providing temporally continuous and spatially detailed information on within-field crop growth. Future research will extend the applicability of this framework to multiple crop types and larger, more heterogeneous agricultural landscapes to further evaluate its robustness under diverse agricultural conditions.

Data availability statement

The data that support the findings of this study are available from the corresponding author, Inhong Song, upon reasonable request.

Author contribution

Jaejun Gou: Conceptualization, Methodology, Formal analysis, Investigation, Software, Writing – Original Draft, Visualization. Dongwon Kang: Conceptualization, Methodology, Investigation, Software. Hyeokjin Lee: Formal analysis, Resources, Writing – Review and Editing. Seongju Jang: Validation, Data Curation, Writing – Review and Editing. Inhong Song: Supervision, Project administration, Funding acquisition, Writing – Review and Editing.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding Statement

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry(IPET) through Intelligent Agricultural Infra Management for Climate Change Development Program, funded by Ministry of Agriculture, Food and Rural Affairs(MAFRA)(RS-2025-02263147)

Reference

Bazrafkan, A., Igathinathane, C., Bandillo, N., & Flores, P. (2025). Optimizing integration techniques for UAS and satellite image data in precision agriculture — a review [Review]. *Frontiers in Remote Sensing*, *Volume 6 - 2025*. https://doi.org/10.3389/frsen.2025.1622884

Feng, G., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(8), 2207-2218. https://doi.org/10.1109/TGRS.2006.872081

Huang, B., Zhang, H., Song, H., Wang, J., & Song, C. (2013). Unified fusion of remote-sensing imagery: generating simultaneously high-resolution synthetic spatial–temporal–spectral earth observations. *Remote Sensing Letters*, *4*(6), 561-569. https://doi.org/10.1080/2150704X.2013.769283

Lacerda, L. N., Snider, J. L., Cohen, Y., Liakos, V., Gobbo, S., & Vellidis, G. (2022). Using UAV-based thermal imagery to detect crop water status variability in cotton. *Smart Agricultural Technology*, *2*, 100029. https://doi.org/https://doi.org/10.1016/j.atech.2021.100029

Li, Y., Yan, W., An, S., Gao, W., Jia, J., Tao, S., & Wang, W. (2023). A Spatio-Temporal Fusion Framework of UAV and Satellite Imagery for Winter Wheat Growth Monitoring. *Drones*, *7*(1), 23. https://www.mdpi.com/2504-446X/7/1/23

Liao, L., Song, J., Wang, J., Xiao, Z., & Wang, J. (2016). Bayesian Method for Building Frequent Landsat-Like NDVI Datasets by Integrating MODIS and Landsat NDVI. *Remote Sensing*, *8*(6), 452. https://www.mdpi.com/2072-4292/8/6/452

Ma, Z., Bao, W., Feng, W., Zhang, X., Ma, X., & Qu, K. (2025). SFT-GAN: Sparse Fast Transformer Fusion Method Based on GAN for Remote Sensing Spatiotemporal Fusion. *Remote Sensing*, *17*(13), 2315. https://www.mdpi.com/2072-4292/17/13/2315

Meng, J., Du, X., & Wu, B. (2013). Generation of high spatial and temporal resolution NDVI and its application in crop biomass estimation. *International Journal of Digital Earth*, *6*(3), 203-218. https://doi.org/10.1080/17538947.2011.623189

Radočaj, D., & Jurišić, M. (2025). A Phenology-Based Evaluation of the Optimal Proxy for Cropland Suitability Based on Crop Yield Correlations from Sentinel-2 Image Time-Series. *Agriculture*, *15*(8), 859. https://www.mdpi.com/2077-0472/15/8/859

Shammi, S. A., & Meng, Q. (2021). Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. *Ecological Indicators*, *121*, 107124. https://doi.org/https://doi.org/10.1016/j.ecolind.2020.107124

Somerscales, S., & Chen, Q. (2026). Crop presence and growth-stage mapping in a tropical smallholder setting in Hawaiʻi with PlanetScope time series. *International journal of remote sensing*, 1-23. https://doi.org/10.1080/01431161.2026.2648095

Swain, R., Paul, A., & Behera, M. D. (2024). Spatio-temporal fusion methods for spectral remote sensing: a comprehensive technical review and comparative analysis. *Tropical Ecology*, *65*(3), 356-375. https://doi.org/10.1007/s42965-023-00318-5

Verger, A., Baret, F., Weiss, M., Kandasamy, S., & Vermote, E. (2013). The CACAO Method for Smoothing, Gap Filling, and Characterizing Seasonal Anomalies in Satellite Time Series. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(4), 1963-1972. https://doi.org/10.1109/TGRS.2012.2228653

Wang, Q., & Atkinson, P. M. (2018). Spatio-temporal fusion for daily Sentinel-2 images. *Remote sensing of Environment*, *204*, 31-42. https://doi.org/https://doi.org/10.1016/j.rse.2017.10.046

Xiao, D., Niu, H., Guo, F., Zhao, S., & Fan, L. (2022). Monitoring irrigation dynamics in paddy fields using spatiotemporal fusion of Sentinel-2 and MODIS. *Agricultural Water Management*, *263*, 107409. https://doi.org/https://doi.org/10.1016/j.agwat.2021.107409

Xiao, J., Suab, S. A., Chen, X., Singh, C. K., Singh, D., Aggarwal, A. K., Korom, A., Widyatmanti, W., Mollah, T. H., Minh, H. V. T., Khedher, K. M., & Avtar, R. (2023). Enhancing assessment of corn growth performance using unmanned aerial vehicles (UAVs) and deep learning. *Measurement*, *214*, 112764. https://doi.org/https://doi.org/10.1016/j.measurement.2023.112764

Xu, C., Du, X., Yan, Z., Zhu, J., Xu, S., & Fan, X. (2022). VSDF: A variation-based spatiotemporal data fusion method. *Remote sensing of Environment*, *283*, 113309. https://doi.org/https://doi.org/10.1016/j.rse.2022.113309

Xu, Q., Guo, Y., Chen, W., Ji, G., Shi, L., Li, Y., Zheng, Z., Sun, C., & Zhu, H. (2024). Comprehensive assessment of Spatiotemporal fusion methods in inland water monitoring. *GIScience & Remote Sensing*, *61*(1), 2343200. https://doi.org/10.1080/15481603.2024.2343200

Yin, G., Li, A., Jin, H., Zhao, W., Bian, J., Qu, Y., Zeng, Y., & Xu, B. (2017). Derivation of temporally continuous LAI reference maps through combining the LAINet observation system with CACAO. *Agricultural and Forest Meteorology*, *233*, 209-221. https://doi.org/https://doi.org/10.1016/j.agrformet.2016.11.267

Yu, W., Li, J., Liu, Q., Zhao, J., Dong, Y., Zhu, X., Lin, S., Zhang, H., & Zhang, Z. (2021). Gap Filling for Historical Landsat NDVI Time Series by Integrating Climate Data. *Remote Sensing*, *13*(3), 484. https://www.mdpi.com/2072-4292/13/3/484

Zhang, W., & Li, Y. (2025). A spatio-temporal-spectral fusion framework for downscaling Sentinel-2 images using UAV images. *International journal of remote sensing*, *46*(3), 1257-1278. https://doi.org/10.1080/01431161.2024.2429781

Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J. G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote sensing of Environment*, *114*(11), 2610-2623. https://doi.org/https://doi.org/10.1016/j.rse.2010.05.032

Zhu, X., Helmer, E. H., Gao, F., Liu, D., Chen, J., & Lefsky, M. A. (2016). A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote sensing of Environment*, *172*, 165-177. https://doi.org/https://doi.org/10.1016/j.rse.2015.11.016

Zurita-Milla, R., Clevers, J. G. P. W., & Schaepman, M. E. (2008). Unmixing-Based Landsat TM and MERIS FR Data Fusion. *IEEE Geoscience and Remote Sensing Letters*, *5*(3), 453-457. https://doi.org/10.1109/LGRS.2008.919685