TIF: A Time-series-based Image Fusion Algorithm

Kexin Song[1*], Zhe Zhu[1*], Shi Qiu[1], Pontus Olofsson[2], Christopher S.R. Neigh[3], Junchang Ju[3,4], Qiang Zhou[3,5]

[1]Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 06269, USA
[2]NASA Marshall Space Flight Center, Huntsville, AL, USA
[3]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA 20771
[4]University of Maryland, Earth System Science Interdisciplinary Center, 5825 University Research Ct, suite 4001, College Park, Maryland 20740
[5]Science Systems and Applications, Inc (SSAI), 10210 Greenbelt Rd, Suite 600, Lanham, Maryland, USA 20706

Abstract

We developed a Time-series-based Image Fusion (TIF) algorithm to generate 10-m surface reflectance time series by synthesizing Landsats 8/9 and Sentinel-2 A/B data. Unlike traditional methods that rely on image pairs or thematic maps, TIF extracts all valid pixel-level observation pairs across time to build per-pixel linear regression models. This approach captures the spectral relationships between sensors while accounting for land surface dynamics. A temporal weighting scheme and an iterative refinement strategy improves the fusion process, yielding reusable coefficients that support efficient, scalable 10-m time-series generation. TIF was applied to all Landsat multispectral bands, using native 10-m Sentinel-2 bands (Blue, Green, Red) and resampled bands (NIR and SWIR1/2) for visual assessment, with quantitative accuracy evaluated at the original Sentinel-2 resolutions. Experiments across five U.S. sites show TIF consistently outperforms state-of-the-art methods like STARFM, FSDAF 2.0, Sen2Like, and ESRCNN. For instance, TIF demonstrated a reduction in RMSE by 24% and an increase in SSIM by 6% compared to FSDAF 2.0 and ESRCNN, and outclassed STARFM and Sen2Like, which showed weaker results across all metrics. In multi-date change detection, TIF-predicted images achieved a mean F1 score of 0.70 and a mean disagreement rate of 0.05 against reference maps. TIF offers

a potential practical and efficient pathway for creating 10-m versions of NASA's HLS products, opening new opportunities for fine-scale, time-sensitive Earth observations.

Keywords: Landsat, Sentinel-2, HLS, TIF, data fusion, time series, change detection, harmonization

Highlights

- **A novel satellite spatial-temporal fusion algorithm called TIF is proposed.**
- **TIF creates 10-m harmonized Landsat and Sentinel-2 surface reflectance time series.**
- **TIF outperforms other state-of-the-art methods in spectral and spatial accuracy.**
- **TIF operates purely on time-series data, without requiring image pairs.**
- **TIF predictions can be used for detecting land changes.**

1. Introduction

Monitoring Earth's dynamic surface requires observing changes across various spatial and temporal scales (Zhu et al., 2022). Satellite time series are an invaluable tool for this purpose, supporting diverse applications in agriculture (Weiss et al., 2020), forestry (Fassnacht et al., 2024), ecological monitoring (Kennedy et al., 2014, Zeng et al., 2020), and urbanization (Zhu et al., 2019). While temporally dense datasets like MODIS (250/500/1000-m) and VIIRS (375/750-m) provide daily global coverage, their coarse spatial resolution hinders detailed analysis of heterogeneous landscapes, limiting their utility for applications such as precision agriculture, land-cover mapping, and quantifying localized land-use change (Chen et al., 2016; Radeloff et al., 2024; Thornton et al., 2023; Suh et al., 2024). Conversely, medium-resolution sensors like Landsats 8-9 (30-m) offer substantially greater spatial detail, but their revisit cycle (16 days with one satellite and 8 days with two) presents challenges for monitoring rapid changes like floods, irrigation dynamics, and insect outbreaks, which often unfold over periods of days to weeks (Zhu et al., 2022).

The NASA Harmonized Landsat and Sentinel-2 dataset (HLS) is a deliverable of the Satellite Needs Working Group, an interagency effort of the US Government dedicated to identifying and addressing Earth observation needs across US civilian federal agencies (Olofsson and Virts, 2022). HLS provides dense time series by combining data from Landsats 8/9 Operational Land Imager (OLI) and Sentinel-2 Multi-Spectral Instrument (MSI). Both top-of-atmosphere (TOA) and surface reflectance products are available at ~2 day revisit interval (Claverie et al. 2018, Ju et al., 2025). While HLS offers consistent and temporally dense datasets, its spatial resolution remains limited to 30-m for both sensors. The 10-m native resolution Sentinel-2 bands were not included as an operational component in the HLS version 2.0 release, primarily due to data volume and logistical considerations. This limitation restricts its applicability for many

63 fine-scale, time-sensitive applications. For example, Tulbure et al. (2022) emphasized the need of
64 higher resolution HLS data to map short-lived flooding. Tian et al. (2020) suggested that 10-m
65 Normalized Difference Vegetation Index (NDVI) time series is required to provide accurate
66 vegetation phenology in urban areas. Additionally, 10-m HLS data hold promise for improving
67 land use classification (Falanga Bolognesi et al. 2020; Parreiras et al. 2022), small-scale forest
68 disturbance detection (Chen et al. 2021), phenology monitoring of heterogeneous forests (Bolton
69 et al. 2020), and near real-time, sub-field level crop monitoring (Gao and Zhang, 2021; Liao et al.,
70 2023; Duan et al., 2024). As a result, a potential solution to address this resolution gap is to blend
71 30-m Landsats 8/9 to the 10-m Sentinel-2 pixels through data fusion. However, doing so at a large
72 scale requires robust fusion approaches that can overcome challenges of data availability and land
73 surface dynamics by leveraging the full temporal depth of the satellite archives.

74 Spatial-temporal image fusion is an effective way to enhance both spatial and temporal
75 resolutions of satellite imagery (Wu et al. 2022). Here, temporal resolution refers to the frequency
76 at which satellite observations are available for a given location. Many algorithms have been
77 developed over the past two decades, broadly categorized as weight-function-based, unmixing-
78 based, learning-based, and hybrid methods (Ghamisi et al. 2019; Wang et al. 2023; Xiao et al.
79 2023; Zhu et al. 2018). The field was pioneered by weight-function-based approaches like the
80 Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al., 2006), which
81 blended 500-m MODIS and 30-m Landsat imagery. While revolutionary, STARFM's performance
82 was limited in landscapes with significant heterogeneity or when predicting abrupt land cover
83 changes (Hilker et al., 2009a). To address these shortcomings, several enhancements were
84 developed, including the Enhanced STARFM (ESTARFM) (Zhu et al., 2010) and the Spatial
85 Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH) (Hilker et al.,

2009b). These methods improved the handling of land changes but still fundamentally relied on the availability of high-quality, temporally-close image pairs and could be constrained by the ability of the coarser data (e.g., MODIS) to accurately represent fine-scale change trajectories. Another class of image fusion methods relies on unmixing approaches (Zhukov et al., 1999; Xu et al., 2015). Fine-resolution land-cover and land-use maps are usually required to create endmembers and estimate the sub-pixel surface reflectance. Moreover, abrupt changes in surface properties between the input and prediction dates often degrade the performance of unmixing approaches.

In recent years, learning-based approaches—including Bayesian methods (Li et al., 2013), sparse representation, and particularly machine and deep learning—have gained significant interest (Xiao et al., 2023). Deep learning methods, for instance, like the extended super-resolution convolutional neural network (ESRCNN) (Shao et al., 2019) and the degradation-term constrained spatiotemporal fusion network (DSTFN) (Wu et al., 2022), often employ a two-step process. First, they sharpen Sentinel-2's native 20-m SWIR bands to 10-m using a convolutional neural network (CNN), guided by the native 10-m bands. Second, another CNN fuses the 30-m Landsat data with the resulting complete 10-m Sentinel-2 dataset. While these approaches can perform well in heterogeneous and dynamic landscapes, they have significant limitations. They require clear-sky Sentinel-2 imagery near the prediction date and rely on substantial GPU resources for model training and inference. Furthermore, a major challenge of deep learning methods is their "black box" nature. The complex, non-linear transformations learned by the network can make it difficult to interpret how a result was generated, and they can sometimes produce unpredictable artifacts or smooth over critical spatial details in ways that are not physically intuitive, potentially affecting the reliability of the fused imagery for scientific analysis. To overcome the limitations of any single

approach, hybrid methods have been developed, which strategically combine components from different fusion categories. A prominent example is the Flexible Spatiotemporal Data Fusion (FSDAF) model (Zhu et al., 2016) and its successors like FSDAF 2.0 (Guo et al., 2020) and SFSDAF (Li et al., 2020). FSDAF-like methods integrate spectral unmixing to handle heterogeneous landscapes with weight-function-based techniques to predict changes, proving effective in complex areas. Other hybrid innovations include the Fit-FC model (Wang and Atkinson, 2018), which enhances regression-based predictions by incorporating spatial filtering and residual compensation to correct errors arising from temporal changes. Similarly, the FIRST model (Liu et al., 2022) improves regression accuracy by leveraging the principle of spectral autocorrelation. These hybrid approaches represent a significant effort to create more robust and accurate fusion models by synergizing the strengths of different theoretical foundations.

Recognizing the limitations of single-pair methods, recent research has increasingly incorporated time-series data. One group of methods eases the strict image pair requirement by operating on spatiotemporal patches. Approaches like the ROBust OpTimization-based (ROBOT) model (Chen et al., 2023), the Unpaired Spatio-Temporal Fusion of Image Patches (USTFIP) (Goyena et al., 2023), and the Quick Spatiotemporal Fusion with Coarse- and Fine-Resolution Scale Transformation Errors and Pixel-Based Synthesis Base Image Pair (STEPSBI) (Ma et al., 2023) divide the region of interest into smaller image patches. By learning relationships from multiple patches over time, they can relax the need for perfectly aligned, full-scene image pairs, but may still face scaling challenges for long-term, large-area applications. A second group focuses on modeling the temporal evolution of pixels to generate synthetic fine-resolution time series, often for specific variables like Vegetation Indices (VIs). These include geostatistical approaches like STGDFM (Kim et al., 2020), which uses land-cover-specific temporal trends and kriging, and

shape-matching methods like SSMM (Zhang et al., 2021; Shen et al., 2023) and TSFM (Liu et al., 2022). These shape-matching models function by identifying pixels with similar temporal profiles (e.g., phenological curves) between coarse- and fine-resolution datasets. While powerful, their core assumption—that the temporal shapes of both time series are consistent—can be violated during abrupt land cover changes within a coarse pixel. Furthermore, methods like SSMM often require searching for similar pixels within a large spatial window, which introduces spatial dependencies and can be computationally inefficient.

Despite these advancements, the creation of large-scale, long-term, operational 10-m surface reflectance time series remains a significant challenge, and to date, no such product exists. Three key limitations in existing spatial-temporal fusion methods hinder this goal. First, many algorithms are constrained by a dependence on high-quality, temporally-aligned image pairs. Weight-function-based and hybrid methods, often designed for MODIS-Landsat fusion, struggle when applied to Landsat and Sentinel-2 because obtaining frequent, cloud-free, near-coincident image pairs is difficult. Even recent methods that incorporate time series can be computationally impractical for processing many years of dense data at large scales. Second, most models exhibit degraded performance in dynamic and heterogeneous landscapes. They often rely on simplifying assumptions, such as the absence of land cover change or that surface reflectance changes linearly between observation dates (Xiao et al., 2023). These assumptions break down in real-world scenarios, reducing algorithm accuracy, particularly when abrupt land changes occur (Zhu et al., 2018). Finally, high computational cost and scalability remain a major hurdle. Traditional methods like weight-function-based and unmixing approaches often employ moving-window searches for similar pixels, which are computationally intensive and difficult to scale (Liu et al., 2019). While deep learning and hybrid approaches leverage GPU acceleration and parallel processing

techniques, scaling these models for continuous, global-scale operational processing is still a formidable challenge (Goyena et al., 2023).

To address these challenges, this paper introduces Time-series-based Image Fusion (TIF), a novel algorithm designed to generate high-spatiotemporal-resolution surface reflectance data. TIF's innovative approach introduces a fully pixel-based framework that learns relationships directly from the entire historical record of observation pairs, thereby avoiding reliance on clear-sky image pairs or computationally expensive searches for similar pixels. By doing so, it generates reusable, pixel-level coefficients that eliminate the need to re-run or retrain the model for each new prediction date. This design explicitly accounts for abrupt land surface changes through adaptive temporal clustering, removing the common assumption of temporal shape consistency. Furthermore, TIF focuses on predicting fundamental surface reflectance rather than derived indices, which enables more flexible, scalable, and robust prediction of 10-m land surface dynamics. We demonstrate that TIF can accurately produce 10-m surface reflectance predictions for any date with a corresponding 30-m Landsat observation, achieving high computational efficiency. This capability paves the way for creating an operational 10-m global HLS product, significantly advancing our capacity to monitor and analyze Earth surface phenomena at unprecedented spatiotemporal scales.

2. Study Area and Datasets

2.1. Study Area

Five HLS tiles across the Conterminous United States (CONUS) were selected for TIF algorithm calibration and validation (Figure 1). These tiles include Far West (T10SFG), Rocky Mountains (T13TCF), Great Plains (T14SPJ), Southeast (T15RXQ), and New England (T18TXM). This selection provides a range of topographic features, land cover/land use types,

historical disturbance regimes, and cloud cover frequencies (Table S1), enabling a thorough assessment of TIF's performance under varied conditions.

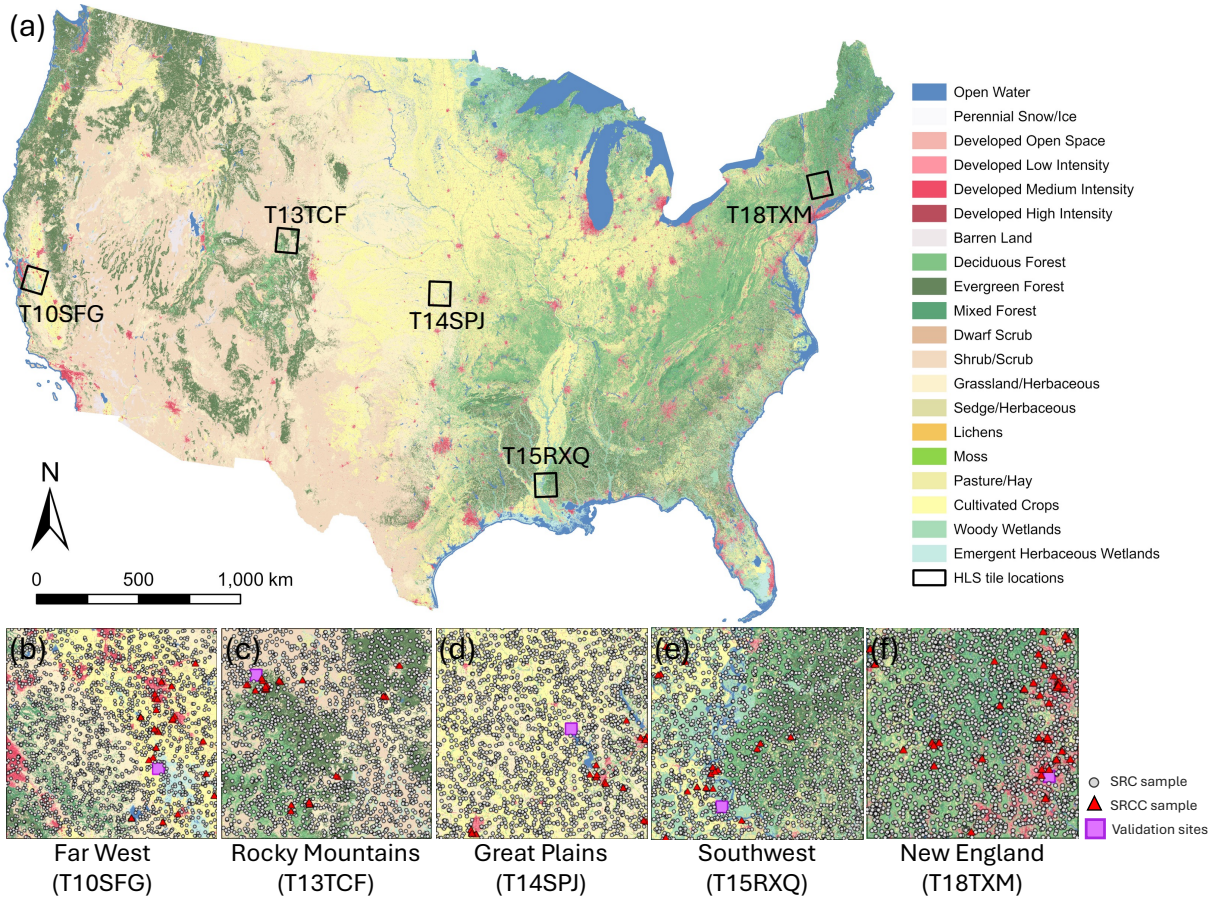


Figure 1. (a) Black polygons: The five HLS tiles used for this study. Background: 2021 NLCD land cover maps (Dewitz 2023). The enlarged views of these tiles are shown at (b) Far West (T10SFG), (c) Rocky Mountains (T13TCF), (d) Great Plains (T14SPJ), (e) Southwest (T15RXQ), and (f) New England (T18TXM). Each tile covers surface areas with 109.8 km by 109.8 km. TIF calibration data includes 10,000 simple random calibration (SRC) sample (grey circles) and 5,000 stratified random change calibration (SRCC) sample (red triangles). Independent validation sites used for evaluating algorithm performance are highlighted in purple polygons. TIF: Time-series-based Image Fusion.

2.2. Landsat Surface Reflectance (30-m)

The lower-resolution input data, relative to the target 10-m resolution, consisted of 30-m Landsat 8 and 9 surface reflectance from the NASA Harmonized Landsat-Sentinel (HLS) Version

2.0 product (Ju et al., 2025), hereinafter referred to as L30. The NASA HLS products were selected due to their comprehensive preprocessing: atmospheric correction using the Land Surface Reflectance Code (LaSRC) v3.5.5 (Vermote et al., 2016), cloud and cloud-shadow masking with Fmask 4.6 (Qiu et al., 2019), and precise co-registration and gridding of Landsat data using the Automated Registration and Orthorectification Package (AROP) (Gao et al., 2009) over the Military Grid Reference System (MGRS) tiling system to align with Sentinel-2 data. Additionally, the C-factor bidirectional reflectance distribution function (BRDF) normalization is applied (Roy et al., 2016). The view angle is set to nadir. The illumination angle is calculated by averaging the solar zenith angles observed during the overpasses of Landsat-8 and Sentinel-2 at the center latitude of the tile (Zhang et al., 2015). We accessed L30 from 2013 to 2021 to ensure comprehensive temporal coverage for our analysis. The details of spatial and spectral characteristics of Landsat 8-9/OLI are shown in Table 1.

Table 1. Spatial and spectral characteristics of Landsat 8-9/OLI and Sentinel-2/MSI used in this study.

Band name	Landsat 8-9/OLI			Sentinel-2/MSI		
	Band ID	Resolution (m)	Bandwidth (nm)	Band ID	Resolution (m)	Bandwidth (nm)
Blue	Band 2	30	450–515	Band 2	10	458-523
Green	Band 3	30	525–600	Band 3	10	543-578
Red	Band 4	30	630-680	Band 4	10	650-680
Broad NIR (BNIR)	-	-	-	Band 8	10	785-899
Narrow NIR (NNIR)	Band 5	30	845-885	Band 8A	20	855-875
SWIR1	Band 6	30	1560-1660	Band 11	20	1565-1655
SWIR2	Band 7	30	2100-2300	Band 12	20	2100-2280

2.3. Sentinel-2 Surface Reflectance (10-m)

For the high-resolution image inputs, we utilized 10-m BRDF-corrected Sentinel-2 surface reflectance, hereinafter referred to as S10. As HLS version 2.0 does not include S10, we derived and preprocessed our own from Sentinel-2 TOA reflectance downloaded from the Copernicus Open Access Hub (now replaced by the Copernicus Data Space Ecosystem). The preprocessing of S10 consists of the following steps: Sen2cor 2.10 (Main-Knorn et al. 2017) for atmospheric correction, Fmask 4.6 (Qiu et al. 2019) for cloud and cloud-shadow screening, nearest interpolation to downscale the cloud mask into 10-m resolution, C-factor BRDF correction (Roy et al. 2017), and bicubic interpolation to resample 20-m Bands 8A, 11 and 12 to 10-m grids. Note that we chose Sentinel-2 Band 8A with a narrower bandwidth to fuse the Landsat NIR band due to their similar spectral characteristics. We used S10 acquired between 2015 and 2021 in this study. Details of Sentinel-2/MSI spectral bands range and spatial resolutions are listed in Table 1.

2.4. Multi-resolution Benchmark Datasets

While TIF can predict 10-m Landsat NIR and SWIR data using downscaled (20-m to 10-m) Sentinel-2 Bands 8A, 11 and 12 as inputs, these resampled Sentinel-2 bands cannot serve as a reliable reference for evaluating fusion performance at 10 m. This is because they do not represent true 10-m observations. We used a multi-resolution benchmarking strategy to address the lack of "ground-truth" 10-m NIR and SWIR Sentinel-2 bands. First, Landsat Bands 5-7 were downsampled to 60-m using a Gaussian model-based degradation function, also known as the point spread function (PSF) (Williams and Becklund, 2002) to simulate coarser-resolution imagery. Next, the TIF framework was used to fuse this 60-m Landsat time series with the original 20-m Sentinel-2 Bands 8A, 11, and 12. This generated a 20-m fused product for Landsat Bands 5-7. Finally, we compared this 20-m fused products with the original 20-m Sentinel-2 bands (serving

as the reference) to quantitatively assess TIF's fusion accuracy using spectral metrics. This scale-degradation experiment is consistent with the principles of Wald's protocol (Wald et al., 1997), a foundational validation framework in image fusion. The protocol relies on the principle of scale invariance, positing that an algorithm's ability to accurately reconstruct a sensor's characteristics at an intermediate resolution (e.g., 60-m to 20-m) is a strong indicator of its capability to perform a similar downscaling task at other resolutions where direct ground truth is unavailable (e.g., 30-m to 10-m). This controlled experiment provides an objective measure of TIF's performance in resolving finer spatial details while preserving spectral fidelity for these bands (Wang et al., 2016, 2017; Shao et al., 2019).

2.5. Calibration Data

To develop and calibrate the TIF algorithm, we employed a dual-sampling strategy designed to capture both typical land cover conditions and less frequent but important land surface dynamics. First, a simple random calibration (SRC) sample of 10,000 pixels was selected in total, with 2,000 points allocated to each of the five study tiles. This sample represents the dominant land cover types and their characteristic spectral behavior. Second, a stratified random change calibration (SRCC) sample of 5,000 pixels was selected, with 1,000 points allocated to each study tile. These points were randomly drawn from manually delineated areas known to have experienced land surface changes, such as those resulting from land disturbances or vegetation recovery. Figure 1 displays the spatial distribution of both sample classes. By incorporating both the SRC sample, representing common conditions, and the SRCC sample, focusing on land surface changes, we enhanced the robustness of the TIF algorithm and ensured its reliable performance across a wide range of scenarios. This approach allows the algorithm to learn both the general

spectral relationships between Landsat and Sentinel-2 and the specific adjustments needed to accurately capture land surface changes.

2.6. Validation Data

To ensure a rigorous evaluation of the TIF algorithm, we designed the validation process to be entirely independent of the calibration data. Specifically, we identified five 6 km × 6 km test sites that were not used during the algorithm's calibration phase (Figure 1). For each site, two growing-season images were designated to perform both qualitative and quantitative assessments of fusion performance. These test areas were specifically chosen to include diverse land cover types and surface changes, ensuring a robust and unbiased evaluation. While the TIF method is universally applicable, the selection of validation datasets was constrained by the requirements of the benchmark methods, Sen2Like (S2L) (Saunier et al., 2022), STARFM, FSDAF 2.0, and ESRCNN, used for comparison. Specifically, S2L requires at least one Sentinel-2 image prior to the prediction date, while ESRCNN requires at least one completely cloud-free Sentinel-2 image acquired near the prediction date for model training, and the other two approaches require one clear-sky Landsat and Sentinel-2 image pairs.

3. Methodology

3.1. The TIF Algorithm

The conceptual basis of the TIF algorithm is the understanding that while the reconstruction of fine-resolution imagery from coarser data is an inherently ill-posed problem, robust relationships can be established empirically between co-located 30-m Landsat (L30) and 10-m Sentinel-2 (S10) pixels through the analysis of long-term time series. These relationships, which can be approximated by a set of linear models, are critically dependent on land surface dynamics. Instead of assuming a single, static linear relationship, TIF leverages the full temporal depth of all available L30-S10 observation pairs to learn the characteristic spectral co-variations between the

sensors for each 10-m pixel over time. To account for sub-pixel heterogeneity and dynamic processes, it employs K-means clustering to identify and model distinct spectral states or change trajectories. This allows the algorithm to refine the linear relationships for different conditions by then utilizing robust regression and adaptive model selection to derive the most accurate TIF coefficients for each state. Thus, the resulting linear TIF coefficients are not a simplistic assumption but rather the outcome of an adaptive, data-driven modeling process designed to capture these land surface dynamics. The entire TIF process is outlined in Figure 2.

The TIF algorithm begins by assembling a time series of co-located Landsat (L30) and Sentinel-2 (S10) observation pairs for each 10-m pixel, matching observations within a 16-day temporal window (Qiu et al., 2018; Shang et al., 2019). To minimize variability from atmospheric or surface changes, these pairs are temporally weighted, giving higher influence to observations acquired closer in time (details in Section 3.1.1). The core assumption is that a consistent linear relationship exists between L30 and S10 reflectance. As demonstrated in previous work and our examples (Figures 3-5), this assumption holds for stable pixels and for homogeneous changes where both sensors' reflectance shifts synchronously, yielding a single set of regression coefficients (hereafter, TIF coefficients). However, this assumption is violated in cases of heterogeneous change—for instance, when land cover converts within only a portion of an L30 pixel (Figure 5). In such scenarios, TIF coefficients derived from pre- and post-change observations can differ substantially for the changed (and sometimes for the unchanged) S10 pixels within that Landsat pixel. Such a condition is rare but will be more apparent in heterogeneous areas with fragmented change features. To address this, TIF employs K-means clustering (temporally applied; Section 3.1.2) on the time series observation pairs for each pixel. This identifies distinct temporal-spectral states, allowing robust regression (Section 3.1.3) to calculate separate, optimized

TIF coefficients for each state. Finally, these adaptively derived coefficients are applied to the input L30 data to predict the 10-m Sentinel-2-like surface reflectance for any given date (Section 3.1.4). To further illustrate the algorithm's application and performance, we present three representative examples: a stable land surface pixel (Figure 3), a pixel undergoing homogeneous land change (Figure 4), and a pixel experiencing heterogeneous land change (Figure 5).

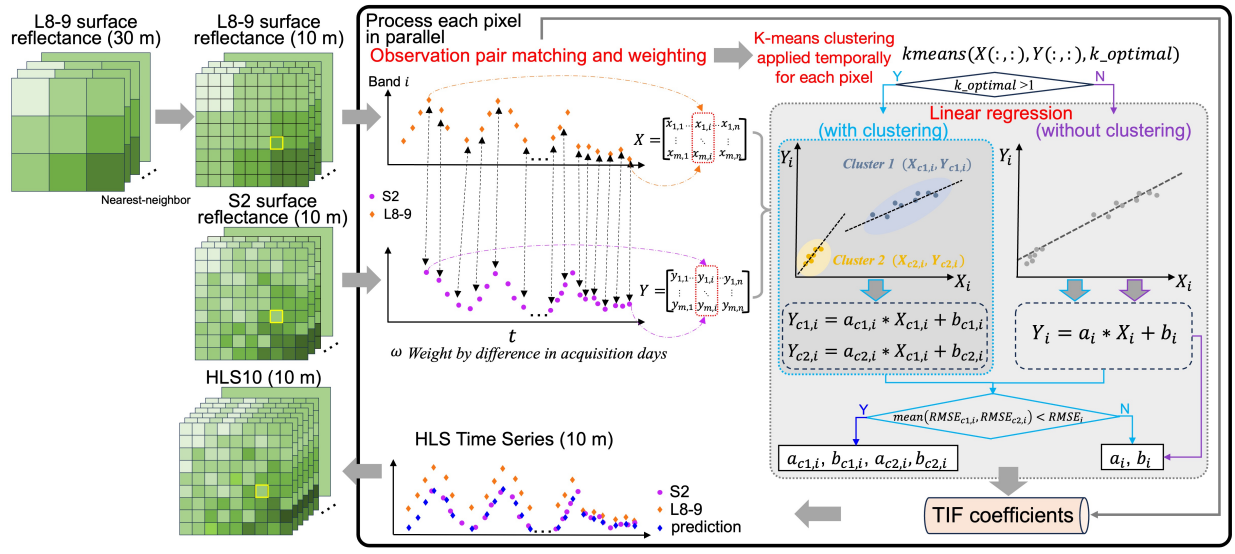


Figure 2. Illustration of 10 m HLS imagery time series obtained from the TIF approach. The key steps of TIF with respect to the target pixel (in yellow square) are highlighted in the black box. Observation pairs from L8-9 and S2 are first matched and weighted, after which pixel-wise K-means clustering is optionally applied. In the "without clustering" branch, a single regression line is fit across all matched pairs, producing two coefficients (slope a and intercept b) that are passed downstream. In the "with clustering" branch, separate regression lines are fit for each cluster, in addition to a simple regression line across all pairs. For each spectral band, the regression RMSEs are compared, and the coefficients from the model with lower error are retained as the final TIF coefficients. TIF: Time-series-based Image Fusion. L8-9: Landsat 8 and 9. S2: Sentinel-2. m is the number of matched observation pairs. n is the total number of spectral bands (n=6).

3.1.1. Selection of L30-S10 Pairs

TIF operates at the pixel level, estimating linear regression coefficients for each pixel and each spectral band using L30-S10 observation pairs. Each L30 pixel is divided into nine S10 pixels. This algorithm searches for valid L30 observations (excluding cloud, cloud shadow, and snow)

collected within ±16 days of the valid S10 acquisition date for observation matching. The 16-day window ensures sufficient L30-S10 observation pairs for regression while reducing misalignments due to temporal variations (Shang et al., 2019). If an L30 matches an S10 on the same day, all other Landsat observations will not be considered. In the absence of a same-day match, the search window is expanded to locate the closest L30 observation (still needs to be within ±16 days) to the S10 date.

To mitigate the influence of temporal variations between Landsat (L30) and Sentinel-2 (S10) observation pairs, TIF employs a weighted regression approach. Unlike previous methods that treat all observation pairs equally (e.g., Shang et al., 2019), TIF assigns each pair an initial weight (ω), which is inversely proportional to the temporal offset between the acquisition date of L30 and S10 (**Error! Reference source not found.**). The smaller the temporal offset, the higher the weighting, resulting in a greater influence of the observation pair on the regression model. The highest weight is given to the observation pairs acquired on the same day. In this study, we tested three weighting functions (Equations. 1-3): 'Fair', 'Cauchy', and 'Sqrt'. According to the sensitive analyses with the SRC sample (Figure S2) and the change sample (Figure S3), we selected the 'Fair' function in TIF due to its high accuracy and simplicity. We also explored the scenario without any weighting functions. The RMSE with the 'Fair' weighting function is 183.77, which is a slight improvement over the RMSE of 185.23 achieved without weighting, thus justifying its inclusion in the TIF algorithm. This weighting strategy ensures that the TIF coefficients are primarily derived from observation pairs with the highest temporal correspondence, minimizing the influence of phenological changes and other temporal variations on the derived spectral relationships.

$$\text{Fair: } \omega = \frac{1}{(1+|r|)} \quad (1)$$

$$\text{Cauchy: } \omega = \frac{1}{(1+r^2)} \quad (2)$$

$$\text{Sqrt: } \omega = \frac{1}{(1+\sqrt{|r|})} \quad (3)$$

where r , ranging from [-16,16], is the temporal offset between the acquisition date of L30 and S10.

3.1.2. K-means Clustering Applied Temporally for Each Pixel

To account for the shifts in the L30-S10 spectral relationships caused by land changes within mixed 30-m pixels, TIF employs K-means clustering across all spectral bands. Importantly, this clustering is performed temporally on the historical observation pairs for each individual 10-m pixel, not spatially across neighboring pixels. For a given 10-m pixel, the full time series of its L30-S10 observation pairs across all six spectral bands is grouped into categories. Each resulting cluster represents a distinct temporal-spectral state, often corresponding to pre-change, post-change, or different phenological conditions. This allows for more nuanced and accurate model fitting by identifying and separating these different states. Such clustering is particularly important for modeling heterogeneous changes, where sub-pixels within an L30 footprint change at varying rates or in different ways.

The algorithm is governed by two key clustering parameters. The first, *k_max*, is a predefined upper limit on the number of clusters to test. Based on sensitivity analysis (Figures S2 and S3), we set *k_max* = 3, as this value provides the best balance between capturing landscape heterogeneity and maintaining sufficient observations within each potential cluster for robust regression. The second parameter, *k_optimal*, is the optimal number of clusters determined empirically for each individual pixel's time series (where $1 \leq k\_optimal \leq k\_max$). We determine

k_optimal using the gap statistic method (Tibshirani et al., 2001), selecting the smallest value of k for which the gap statistic is within one standard deviation of the maximum gap value. A larger gap statistic indicates a more defined, non-random clustering structure in the data. While clustering is performed jointly using all spectral bands to capture the overall spectral state, the resulting L30-S10 patterns can still vary by band within the same temporal cluster (Figure 5c). To handle this, a final adaptive selection step is performed after regression, as detailed in the next section.

3.1.3. Regression Model Fitting and Adaptive Coefficient Selection

To derive the TIF coefficients for each spectral band, a robust linear regression is employed (DuMouchel and O'Brien, 1989). This method is less sensitive to outliers than ordinary least squares (OLS) because it uses an iteratively reweighted least squares (IRLS) approach (Holland and Welsch, 1977). The core principle of IRLS, as implemented in TIF, is to iteratively assign lower weights to observations that have large residuals from the current model fit, thereby reducing their influence on the parameter estimation in subsequent iterations. This process continues until the model coefficients converge. Specifically, in each iteration, weights are calculated based on the residuals from the previous fit using a chosen weighting function. Observations identified as outliers receive progressively smaller weights, effectively downweighing their contribution to the final regression model. This iterative refinement makes the resulting TIF coefficients more robust to noise and improves the overall robustness of the fusion process. Our sensitivity tests confirmed this, showing that robust fitting consistently yielded better accuracies than OLS for both calibration samples (Figures S2 and S3).

A key innovation in TIF is the adaptive selection of coefficients for pixels where K-means has identified multiple temporal clusters. For each spectral band in such a case, TIF fits regression models both with clustering (i.e., a separate model for each cluster) and without clustering (a single

model using all data). The final TIF coefficients are then chosen by comparing the Root Mean Square Error (RMSE) of these competing models, evaluated against same-day matching reference observations. For example, Figure 5 illustrates a pixel experiencing heterogeneous change where the NIR and SWIR2 bands show divergent spectral patterns. For the NIR band, the model with two clusters yields a lower RMSE (404) than the unclustered model (415), so the coefficients from the two clusters are retained. Conversely, for the SWIR2 band, a single unclustered model provides better accuracy, so those coefficients are used. This band-by-band adaptive process ensures that clustering is only used when it demonstrably improves fusion accuracy. This flexibility is crucial because the magnitude and spectral signature of land cover change can vary significantly across bands, and this method optimizes the prediction accuracy for each spectral component individually.

3.1.4. Surface Reflectance Prediction

The final step of the TIF algorithm is to predict the 10-m Sentinel-2-like surface reflectance (y_i) for a given 30-m Landsat observation (x_i). The specific set of TIF coefficients (a_i and b_i) applied depends on the temporal clustering results for that pixel. For a pixel with only one identified temporal cluster ($k\_optimal$ = 1), the prediction is a direct application of the derived linear model for each spectral band i, as shown in Equation (4). For a pixel where multiple temporal clusters (and thus multiple sets of TIF coefficients) exist, a cluster assignment step is required. To determine which set of coefficients to use for the input L30 observation, we calculate the Euclidean distance (D_k) between the multi-band L30 observation vector and the centroid of each available cluster ($c_{i,k}$), as shown in Equation (5). The cluster assignment is performed jointly using all six spectral bands to capture the overall spectral characteristics of the L30 observation. This approach ensures robust assignment, as some land surface changes may only be prominent in

specific bands. The set of TIF coefficients corresponding to the cluster with the minimum Euclidean distance (D_k) is then selected and applied using Equation (4) to generate the final surface reflectance prediction. In rare instances where a robust regression model could not be derived for a pixel due to an insufficient number of historical observation pairs, the missing prediction value is filled using a simple bicubic interpolation of the original 30-m Landsat image.

$$y_i = a_i * x_i + b_i \quad (4)$$

$$D_k = \sqrt{\sum_{i=2}^{7}(x_i - c_{i,k})^2} \quad (5)$$

where

y_i is the predicted Sentinel-2-like 10-m surface reflectance for the i^{th} band.

x_i is the original L30 for the i^{th} band.

a_i is the slope of the TIF coefficients for the i^{th} band.

b_i is the intercept of the TIF coefficients for the i^{th} band.

D_k is the Euclidean distance between L30 and the cluster k's centroids across six bands.

$c_{i,k}$ is the surface reflectance of Cluster k's centroids for the i^{th} band.

i is the spectral band, corresponding to Landsats 8/9 bands 2, 3, 4, 5, 6, and 7.

k is the k^{th} cluster, ranging from 1 to *k_optimal*.

3.1.5. Algorithm Calibration

To calibrate the TIF algorithm, we explored three key components: the weighting function for each matching pair, the maximum number of K-means clustering, and the regression method. Using the L30-S10 observation pairs as TIF input, we applied the holdout validation strategy to randomly split the same-day matching pairs into calibration (60%) and validation (holdout, 40%) sets. For instance, if there were 35 valid pairs (16-day matching) and 15 of them were same-day matches, six of the same-day pairs would be randomly selected and withheld (from model regression) as reference points. This process was repeated five times to ensure robustness. Such a strategy helps ensure that both calibration and validation sets include spatially and temporally diverse observations, reducing potential biases associated with any specific location or time.

Performance was evaluated using a Root Mean Square Error (RMSE) and Average Absolute Difference (AAD), alongside the missing observation rate. The missing observation rate is defined as the proportion of Landsat observations for which TIF could not generate a prediction due to insufficient valid Landsat-Sentinel-2 (L30-S10) observation pairs for regression, compared to a benchmark scenario where the maximum number of clusters (*k_max*) is set to 1 (no clustering). This rate serves as a proxy for the algorithm's applicability to time series data, as it reflects the trade-off between increasing the flexibility of the model (by allowing more clusters) and reducing the number of usable observation pairs.

To derive TIF coefficients, a robust linear regression model is estimated for each spectral band using the available L30-S10 observation pairs within each identified temporal cluster (or for all pairs if no clustering is applied). A minimum of four valid observation pairs is required to proceed with the regression for any given cluster and spectral band. This threshold is set to ensure a sufficiently robust estimation of the two model parameters (slope and intercept). While two

449 points theoretically define a line, using at least four pairs provides a degree of redundancy (two
450 degrees of freedom), making the derived coefficients less sensitive to noise or individual outlying
451 observations and leading to more stable and reliable inter-sensor relationships. Increasing *k_max*
452 generally leads to a higher missing observation rate because dividing the data into more clusters
453 can result in some clusters having fewer than the required four pairs. While higher *k_max* values
454 can potentially improve accuracy by capturing finer-scale variations, an excessively high missing
455 observation rate can negatively impact overall performance by limiting the data available for
456 model training. Therefore, the optimal *k_max* value balances the benefits of increased model
457 flexibility with the need for sufficient training data. To assess the algorithm's accuracy and
458 robustness, we used the mean RMSE and AAD values calculated over five iterations of the
459 optimization process. This iterative approach ensures that the chosen parameters consistently
460 produce reliable results.

461 Calibration results indicate that the optimal configuration for the TIF algorithm is achieved
462 with a maximum of three clusters (*k_max* = 3), utilizing the robustfit regression method and the
463 'Fair' weighting function (Figures S2 and S3). While increasing *k_max* beyond 3 yields marginal
464 improvements in accuracy, it also leads to a noticeable increase in the missing observation rate.
465 Specifically, setting *k_max* to 4 results in a 2% missing observation rate, indicating a significant
466 reduction in the usable time series data. This loss of data outweighs the marginal gain in accuracy.
467 Therefore, *k_max* = 3 strikes the optimal balance between maximizing prediction accuracy and
468 maintaining sufficient observation availability for robust time series analysis. This configuration
469 ensures that the TIF algorithm effectively captures spectral variations while minimizing data loss
470 due to insufficient L30-S10 pairs within individual clusters.

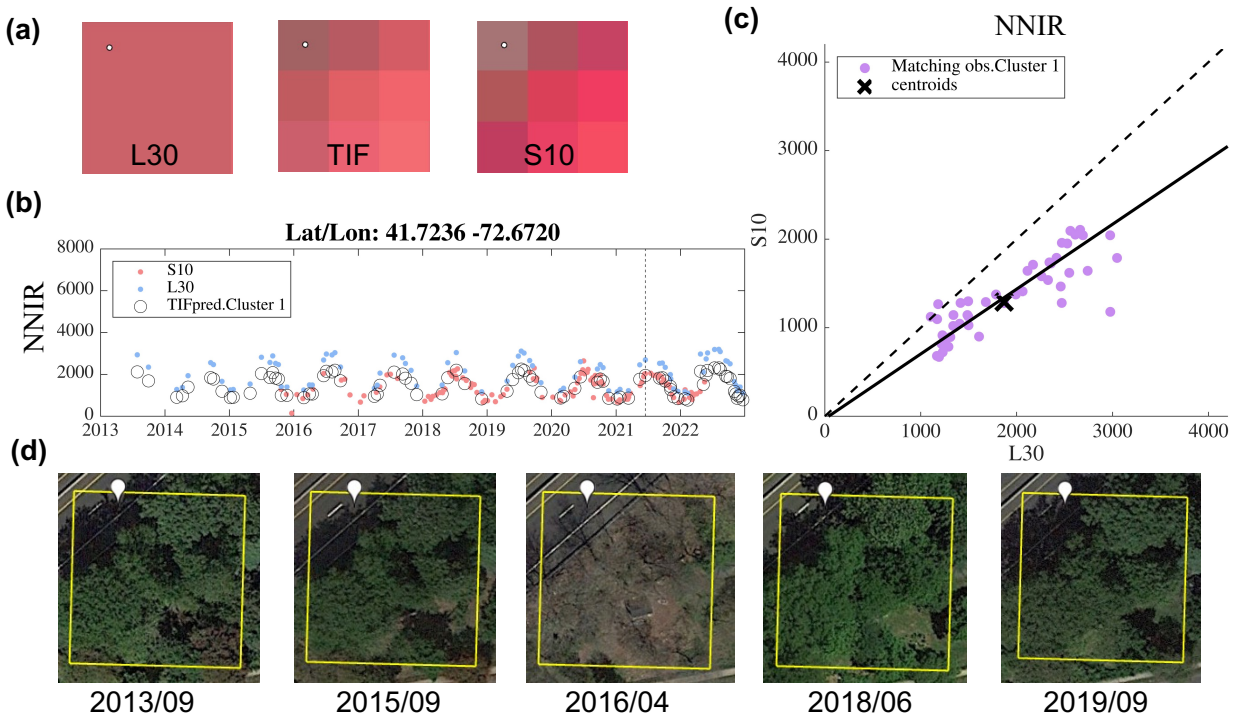


Figure 3. Illustration of TIF execution with a stable land pixel covered by extensive urban trees and a small portion of roads from Wethersfield, CT. (a) False color composite (R/G/B=NNIR/Red/Green) of the L30, TIF predictions, and hidden S10 data for the same geographic area. Both L30 and S10 were acquired on 2021-06-16. The white circle indicates the point of interest (upper left corner) within the 30 m by 30 m pixel. (b) Time series of NNIR reflectance for the point of interest (Lat: 41.7236, Lon: -72.6720). The solid red dots represent S10 data, the blue dots represent L30 data, and the black circles are TIF predictions. The dashed line represents the date of withheld Sentinel-2 image. (c) Linear relationship of the NNIR reflectance at the point of interest (solid line), built by TIF, between matching L30-S10 pairs (purple points), compared to the 1:1 line (dashed line). The black cross denotes the centroid of all matching pairs. (d) Google Earth images of the area of interest captured at different timestamps. The yellow box indicates the extent of the L30 pixel shown in (a). The white marker highlighted the location of the investigated point in (b) and (c).

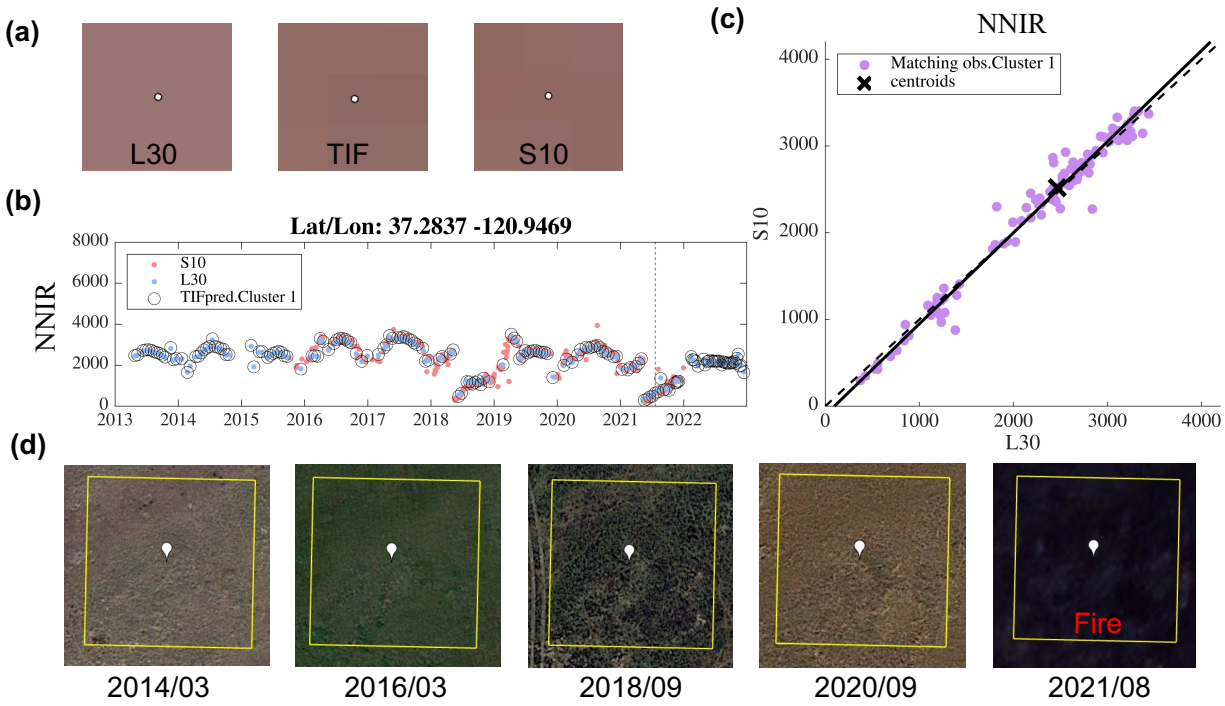


Figure 4. Illustration of TIF execution with a homogeneous land change pixel covered by crops in Gustine, CA. (a) False color composite (R/G/B=NNIR/Red/Green) of the L30 surface reflectance, TIF predictions, and hidden S10 data for the same geographic area. Both L30 and S10 were acquired on 2021-07-20. The white circle indicates the point of interest (middle one) within the 30 m by 30 m pixel. (b) Time series of NNIR reflectance for the point of interest (Lat: 37.2837, Lon: -120.9469) The solid red dots represent S10 data, the blue dots represent L30 data, and the black circles are TIF predictions. The dashed line represents the date of withheld Sentinel-2 image. (c) Linear relationship of the NNIR reflectance at the point of interest (solid line), built by TIF, between matching L30-S10 pairs (purple points), compared to the 1:1 line (dashed line). The black cross denotes the centroid of all matching pairs. (d) Google Earth images of the area of interest captured at different timestamps. The yellow box indicates the extent of the L30 pixel shown in (a). The white marker highlighted the location of the investigated point in (b) and (c).

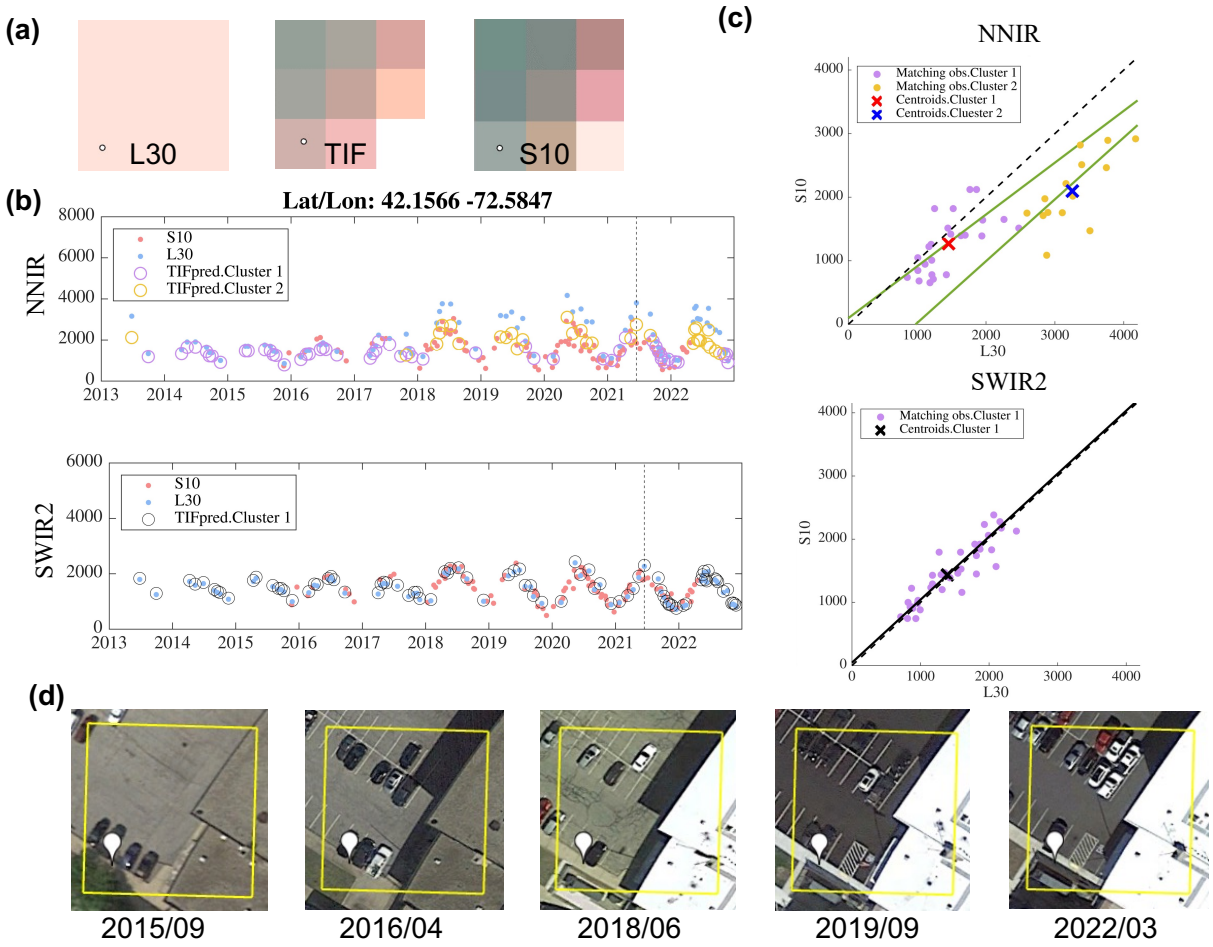


Figure 5. Illustration of TIF execution with a heterogeneous land change pixel from an urban area in the north of Springfield, MA. (a) False color composite (R/G/B= NNIR/Red/Green) of the L30 surface reflectance, TIF predictions, and hidden S10 data for the same geographic area. Both L30 and S10 were acquired on 2021-06-16. The white circle indicates the point of interest (lower left corner) within the 30 m by 30 m pixel. (b) Time series of NNIR and SWIR2 reflectance for the point of interest (Lat: 42.1565, Lon: -72.5847). The solid red dots represent S10 data, and the blue dots represent L30 data. For NNIR, purple and yellow circles indicate TIF predictions for cluster 1 and 2, respectively. For SWIR2, black circles are TIF predictions. The dashed line represents the date of withheld Sentinel-2 image. (c) Linear relationships of surface reflectance at the point of interest (solid lines), built by TIF, compared to the 1:1 line (dashed line). NNIR used clustered matching L30-S10 pairs (purple and yellow points). The red and blue cross denotes the centroids of different clusters. SWIR2 used all matching L30-S10 pairs (black points) with the centroids highlighted in black cross. (d) Google Earth images of the area of interest captured at different timestamps. The yellow box indicates the extent of the L30 pixel shown in (a). The white marker highlighted the location of the investigated point in (b) and (c).

3.2. Sen2Like (S2L)

S2L is an open-source tool designed to produce harmonized 10 and 30-m surface reflectance data from Landsat and Sentinel-2. The data fusion module of S2L utilizes interpolation methods and historical Sentinel-2 data to construct a 10-m Landsat image on the target date. As shown in Equation (6), the predicted 10-m Landsat image (L_{10m}) consists of two parts, the resampled image using bilinear interpolation ($L_{30\to10m}$) and the resolution difference ($D_{30\to10m}$) that describes the details seen by Sentinel-2 but not by Landsat. Following interpolation, it is assumed that the difference in resolution can be estimated from recent Sentinel-2 images with a low pass filter. Two modes are provided to process the Sentinel-2 images and to estimate the difference. (1) The predict mode utilizes linear interpolation to reconstruct a simulated Sentinel-2 image on the target date for calculation $D_{30\to10m}$. On the contrary, (2) the composite mode will select the surface reflectance from the most recent valid observations for analysis, so temporal variations were neglected in this approach. Notably, the S2L fusion module does not require coarse-fine image pairs. While this data fusion algorithm is more efficient than weight-based methods, one major concern is the expected increase in artifacts and noise due to the simplicity of this approach.

$$L_{10m} = L_{30\to10m} + D_{30\to10m} \quad (6)$$

where L_{10m} is the predicted 10-m Landsat image, $L_{30\to10m}$ is the resampled 10-m Landsat image, and $D_{30\to10m}$ refers to the resolution difference between 30-m and 10-m imagery, estimated from Sentinel-2 data using a low-pass filter to recover spatial details not present in the coarser Landsat data.

For this study, we utilized S2L version 4.4.0 and ingested Sentinel-2 L2A and Landsat 8 Collection 2 Level 2 data to generate the 10-m Landsat imagery for all the test sites. As S2L only provides NNIR and SWIR bands at 20-m spacing, we employed bicubic interpolation to obtain

538 10-m images for the visual comparison. Currently, the maximum number of historical Sentinel-2
539 images used to derive the resolution difference is limited to two. Details regarding the S2L
540 procedures can be found at Saunier et al. (2022).

541 **3.3. ESRCNN**

542 ESRCNN (Shao et al. 2019) was developed to predict 10-m Landsat 8 images using a deep
543 learning-based convolutional neural network (CNN). ESRCNN relies on at least one clear
544 Sentinel-2 image acquired close to the target Landsat image. This approach involves two fusion
545 networks with three layers each. The self-adoption fusion of Sentinel-2 is performed to predict 20
546 m SWIR bands (Bands 11 and 12) at 10-m resolution by feeding Sentinel-2 visible (Bands 2-4)
547 and BNIR (Band 8) bands. Then, the multi-temporal fusion of 30-m Landsat 8 Bands 1-7 to 10-m
548 is achieved by using the 15-m Landsat 8 panchromatic band and the six 10-m Sentinel-2 bands
549 (Bands 2, 3, 4, 8, and fused Bands 11 and 12) from the first step.

550 In this study, we kept the original self-adoption fusion and conducted the multi-temporal
551 fusion framework for Landsat 8 Bands 2-7 solely based on Sentinel-2 multispectral bands (i.e.
552 without the 15-m panchromatic band) to keep the input data consistent. We trained and tested
553 ESRCNN models for each validation site (purple polygons in Figure 1) at 30-m to 10-m scale to
554 qualitatively evaluate the accuracies of prediction images. We also adopted the model using the
555 60-m benchmark data to quantitatively assess the spectral fidelity for the predicted 20-m Landsat
556 Bands 5-7.

557 **3.4. STARFM**

558 The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al.,
559 2006) is a foundational weight-function-based algorithm. It predicts a fine-resolution (e.g.,

Landsat) image on a target date by blending information from one or more fine-coarse image pairs acquired on different dates and the coarse-resolution (e.g., MODIS) image from the target date. The core principle involves identifying spectrally similar neighboring pixels within a moving window in the fine-resolution image. These similar neighboring pixels are then weighted based on their spectral similarity, temporal proximity to the prediction date, and spatial distance from the central pixel. These weights are used to calculate the change in reflectance from the coarse-resolution imagery and apply it to the fine-resolution pixels. For our study, we adapted STARFM for Landsat-Sentinel-2 fusion, using one complete Landsat-Sentinel-2 image pair from a base date and the 30-m Landsat image on the prediction date.

3.5. FSDAF 2.0

The Flexible Spatiotemporal Data Fusion (FSDAF) 2.0 (Guo et al., 2020) is an advanced hybrid fusion method that builds upon the original FSDAF (Zhu et al., 2016). It is designed to perform well in heterogeneous landscapes and areas with land cover changes. FSDAF 2.0 integrates multiple techniques: it first uses spectral unmixing to estimate the temporal change for each land cover class based on the difference between coarse-resolution images. It then distributes this class-level change to fine-resolution pixels using a thin plate spline interpolator. A final residual compensation step, similar to STARFM's neighborhood weighting, is applied to capture information missed by the initial class-based prediction. This hybrid approach allows it to better preserve spatial detail and handle land cover changes compared to purely weight-function-based methods. For this study, we implemented FSDAF 2.0 using one base Landsat-Sentinel-2 image pair and the Landsat image on the prediction date.

4. Evaluation Metrics

We evaluated the performance of TIF, S2L, ESRCNN, STARFM, and FSDAF 2.0 through three approaches: qualitative comparison (Section 5.1), quantitative assessments of spectral and spatial fidelity (Section 5.2), and change detection applications (Section 5.3). To generate prediction image for a target date, the TIF algorithm was applied, excluding Sentinel-2 observations acquired on that date from training. The derived TIF coefficients were then used with the original 30-m Landsat image to reconstruct fine-resolution imagery. Details of the prediction dates for each validation site are provided in Table S2. Inputs for S2L and ESRCNN are listed in Table S3 and Table S4, respectively. The configurations of ESRCNN training are outlined in Table S5. Inputs for STARFM and FSDAF 2.0 are shown in Table S6.

4.1. Quantitative Evaluation of Spectral and Spatial Fidelity

To quantitatively evaluate the performance of TIF, S2L, STARFM, FSDAF 2.0, and ESRCNN, we employed six widely used metrics for assessment of image fusion results. Figure 6 presents density plots showing the relationship between the TIF predicted surface reflectance and the reference data for each spectral band. From these plots, we calculated the Correlation Coefficient (CC), RMSE, AAD, and Erreur Relative Global Adimensionnelle de Synthese (ERGAS; Wald, 2002) to assess pixel-level spectral fidelity. To assess spatial fidelity, which encompasses overall image structure and radiometric similarities, we utilized the Universal Image Quality Index (UIQI; Wang and Bovik, 2002) and the Structure SIMilarity (SSIM; Wang et al., 2004). In terms of the metric values, the closer CC, UIQI, and SSIM to 1, the more similar the predictions are to the references. On the contrary, the closer RMSE, AAD, and ERGAS, are to zero, the better the predictions. Finally, to provide a comprehensive evaluation of algorithm performance, we calculated the mean values of these metrics across all six spectral bands.

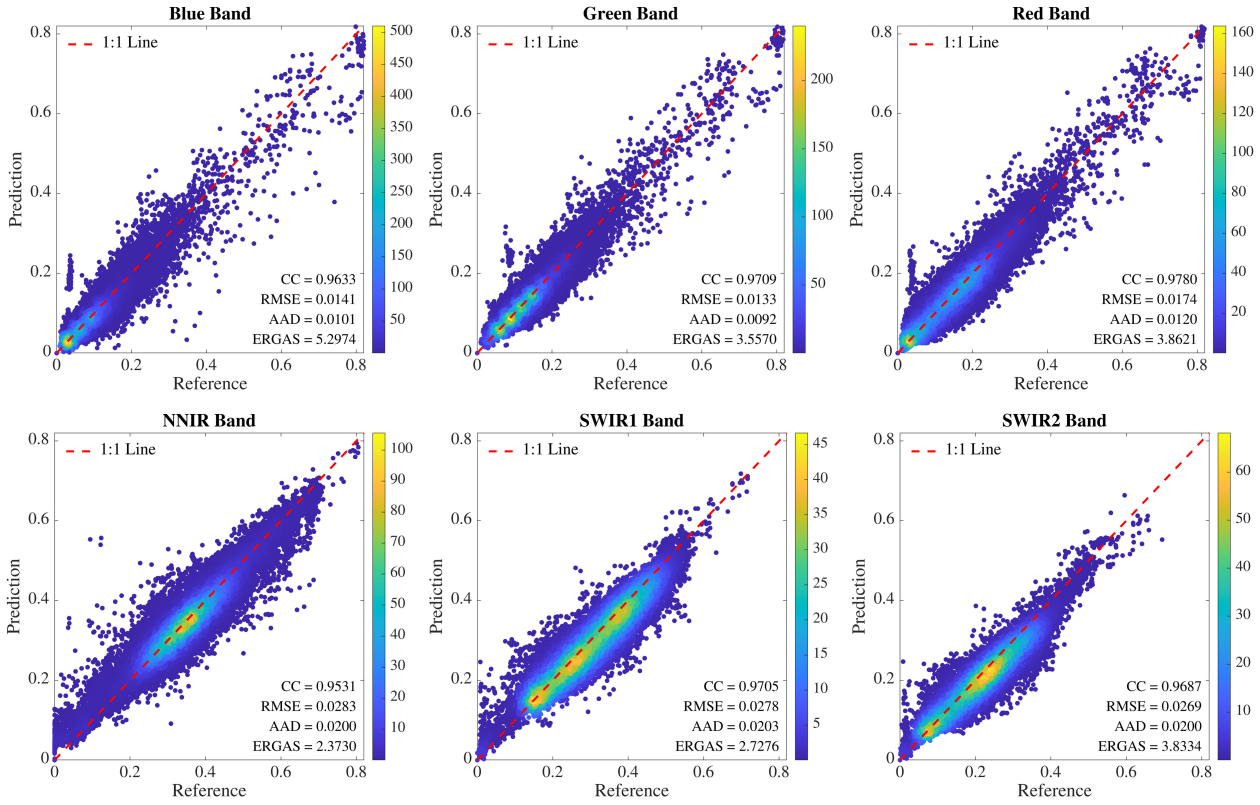


Figure 6. An illustration of TIF predictions (T10SFG, 2021-07-20) versus reference surface reflectance across each spectral band at Far West. TIF: Time-series-based Image Fusion. *CC: Correlation Coefficient; RMSE: Root Mean Square Error; AAD: Absolute Average Difference; ERGAS: Erreur Relative Global Adimensionnelle de Synthese.*

In addition to previously mentioned metrics, we employed the semivariogram (Curran 1988) to measure the spatial fidelity of prediction images. Following the methodology outlined by Qiu et al. (2023), we varied the lag distance h between 0 and 350 m in 10-m increments and analyzed the semivariance for each spectral band. To reduce the computation cost, we randomly selected 1% of the total pixels for each test region (600 × 600 pixels). Figure 7 illustrates the semivariance for a TIF prediction image compared with its corresponding reference image. Higher spatial fidelity was anticipated when the semivariance curve of the predicted image closely aligned with that of the reference image. To quantify these differences, we calculated the absolute

semivariance difference, $\Delta\gamma(h)$, between the prediction image and the reference image at each lag distance (Equation (7) for details and Figure 7 as a visual demonstration). The mean value of these differences across all spectral bands was used to assess algorithm performance in terms of spatial fidelity.

$$\Delta\gamma(h) = \frac{1}{2N}\sum_{i=1}^{N}(x_i - x_{i+h})^2 \quad (7)$$

where,

x_i is the surface reflectance of the ith pixel.

x_{i+h} is the surface reflectance at the pixel with a lag distance of h from the ith pixel.

h is the lag distance between the pixel pair.

N is the total number of paired pixels at the distance h in four directions: upper, lower, left, and right.

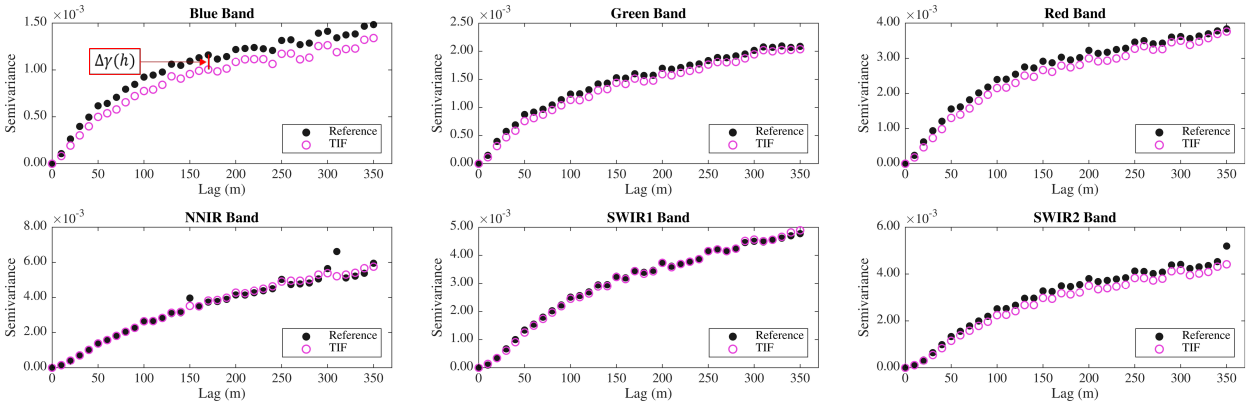


Figure 7. Illustration of semivariograms between the TIF prediction image (T10SFG, 2021-07-20) and the reference image of each spectral band at Far West. TIF: Time-series-based Image Fusion. The $\Delta\gamma(h)$ indicates the semivariance difference between the prediction data and the reference data at lag distance h.

4.2. Supervised Multi-date Change Detection

We created change maps from two growing-season images to assess the applicability of fusion methods in land change detection. Using a multi-date classification change detection approach (Singh, 1989), we classified satellite images from two timestamps to identify changes.

639 The reference change maps were constructed from two actual Sentinel-2 images. The predicted
640 change maps were generated by replacing the second (later) Sentinel-2 image with the predicted
641 10 m image. Agreement between the predicted and reference change maps served as a performance
642 indicator for the fusion methods.

643 We used Random Forest (RF) (Breiman 2001) due to its good performance in remote
644 sensing image classification. We optimized the number of trees to 500 to balance classification
645 accuracy and computation efficiency. For the training data, we collected "stable" and "changed"
646 samples based on the two summer images shown in Table S1. The "changed" category included
647 various types of land changes, such as fire, forest management, water dynamics, agricultural
648 practice, and construction activities etc. The number of training pixels varied across different test
649 areas; the "stable" sample size ranged from 2,334 to 4,351 pixels and the "changed" sample size
650 from 402 to 2,351 pixels. These samples were consistent across different image fusion algorithm
651 evaluations.

652 We calculated the F1 score (for land change) and the disagreement percentage using the
653 reference change maps produced by Sentinel-2 images for quantitative evaluation. The F1 score
654 reflects the classification accuracies over changed places and the disagreement percentage
655 indicates the overall dissimilarity between the reference change map versus the prediction change
656 map. Higher F1 scores and lower percentages of disagreement indicate better fusion performance.

657 **5. Results**

658 **5.1. Visual Assessment**

659 A qualitative visual comparison of TIF, S2L, STARFM, FSDAF 2.0, and ESRCNN outputs
660 (10 m) against withheld reference imagery was conducted across five validation sites. Figure 8

showcases the performance of each method over a heterogenous landscape near Gustine, CA, on 2017-07-09 and 2021-07-20. We could observe that S2L images exhibited clear tonal differences compared to the reference imagery, indicating overall spectral shifts. To highlight fine-scale details, zoom-in views of an agriculture area (Figure 9) and an urban area (Figure 10) are also provided. STARFM showed visible spectral distortions (yellow circles) and speckle-like artifacts (white circles) across both crop fields and urban places. Additionally, it incorrectly generated the shape of a construction on 2017-07-09, as indicated by the yellow arrow. FSDAF 2.0 exhibited similar spectral artifacts to those observed in STARFM and it tended to blur the boundaries between adjacent patches, highlighted by the yellow arrow. ESRCNN exhibited overall good performance, however, inconsistency over edges were observed. For instance, it smeared crop boundaries on 2017-07-09 (yellow arrow) while sharpening them on the other date (white arrow). Both S2L methods struggled to reconstruct spatial details over developed areas (yellow circles). S2L.CompositeMode produced a blurry image on 2017-07-09, whereas S2L.PredictMode generated over-sharpened results with noticeable speckle noise. In contrast, TIF consistently produced high-quality results across all sites and dates. It effectively enhanced spatial details while preserving both spectral and spatial accuracy, resulting in superior visual quality compared to the benchmark methods.

Over homogeneous landscapes, all methods adequately reconstruct the spatial details of roads against the forest background. However, differences emerged in forest management areas (Figure S5 and Figure S6). ESRCNN generated over-sharpened images on 2018-07-04, exhibiting more spatial textures than the reference image, but produced over-smoothed results on 2021-07-28, losing spatial detail in vegetated areas. This suggests that ESRCNN's performance might be unpredictable and image dependent. S2L.CompositeMode and S2L.PredictMode produced blurred

images on 2018-07-04 and noticeable artifacts on 2021-07-28, particularly S2L.PredictMode (Figure S6p). TIF consistently outperformed the benchmark methods in these homogeneous landscapes, enhancing spatial textures while preserving spectral values. The TIF fusion results for remaining validation sites are provided in the Supplementary Data (Figure S11, Figure S13, and Figure S15).

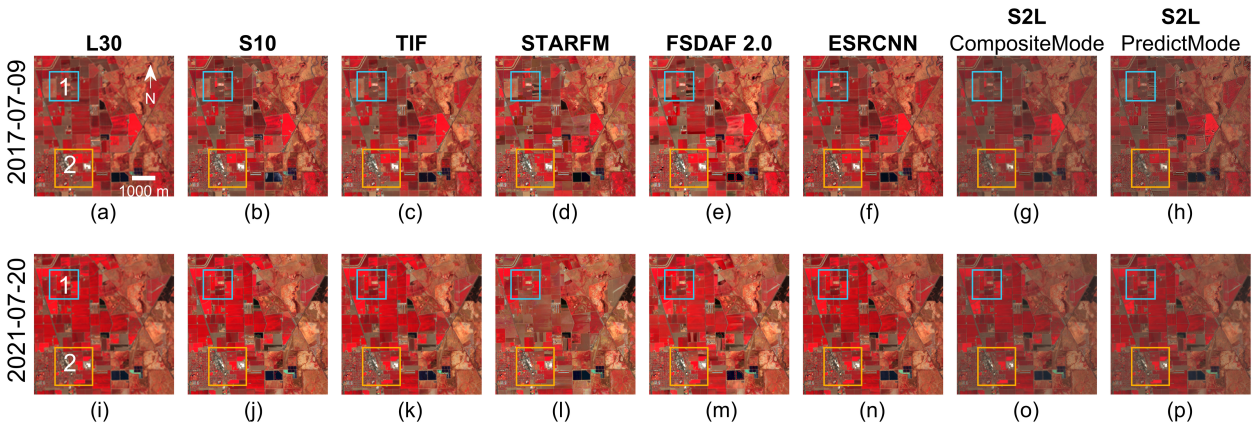


Figure 8. Visual comparison of the fusion results from the T10SFG experiment. The original 30 m Landsat images acquired on 2017-07-09 and 2021-07-20 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2017-07-09 from TIF, STARFM, FSDAF 2.0, ESRCNN, and S2L predict mode and S2L composite mode. The downscaled results on 2021-07-20 are correspondingly (k) – (p). The centroid of the area is Lat = 37.290544, Lon = -120.978904. Images are displayed by NNIR, red, and green bands with the same stretch scales. Two zoom-in areas are highlighted: subarea 1 – blue polygon; subarea 2 – orange polygon.

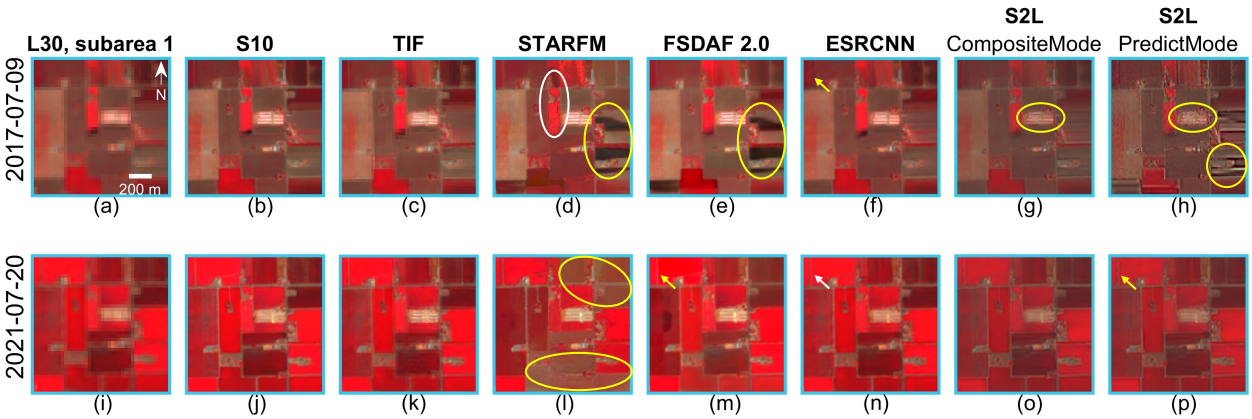


Figure 9. Zoom-in views of subarea 1 marked in **Figure 8**. The original 30 m Landsat images acquired on 2017-07-09 and 2021-07-20 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2017-07-09 from TIF, STARFM, FSDAF 2.0, ESRCNN, and S2L composite mode and S2L predict mode. (k) – (p) are corresponding downscaled images on 2021-07-20. Images are displayed by NNIR, red, and green bands with the same stretch scales.

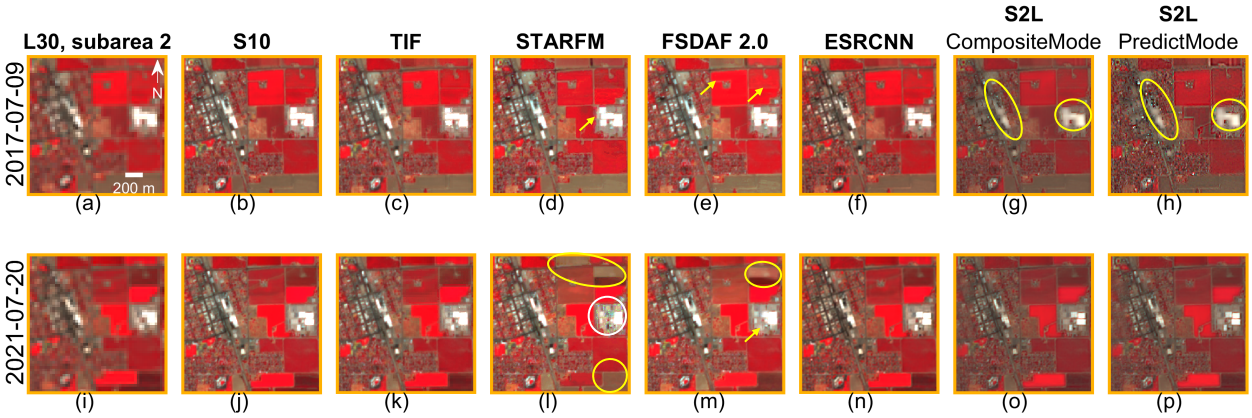


Figure 10. Zoom-in views of subarea 2 marked in **Figure 8**. The original 30 m Landsat images acquired on 2017-07-09 and 2021-07-20 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2017-07-09 from TIF, STARFM, FSDAF 2.0, ESRCNN, and S2L composite mode and S2L predict mode. (k) – (p) are corresponding downscaled images on 2021-07-20. Images are displayed by NNIR, red, and green bands with the same stretch scales.

5.2. Quantitative Evaluation

Figure 11 summarizes the quantitative evaluation across all validation results, with metrics averaged over six spectral bands. TIF achieves the highest accuracy and robustness, consistently delivering top or near-top performance, indicated by structural similarity (UIQI and SSIM), correlation (CC), and the error-based metrics (RMSE, ERGAS, and AAD). FSDAF 2.0 and

ESRCNN also demonstrate strong performance, with competitive results in several metrics. STARFM offers moderate accuracy, generally outperforming the S2L methods. We also notice that the accuracies of STARFM and S2L results varied between day1 and day2 (see Figure S9). The inconsistent performance of these methods is likely due to factors such as the temporal changes that happened between the input image date(s) and the prediction date. These results highlight TIF's robustness and generalizability across diverse surface conditions, with STARFM serving as a baseline and FSDAF 2.0 and ESRCNN as alternatives in specific cases.

The performance of image fusion algorithms varied with land cover types and landscape heterogeneity. TIF achieved the best overall performance, with a mean RMSE of 0.0196 across all validation sites and the lowest RMSE of 0.0132 in homogeneous forested areas (T13TCF). Its performance declined slightly over urban (T18TXM) and agriculture areas (T10SFG). STARFM, FSDAF 2.0, and ESRCNN showed similar patterns. On the contrary, S2L fusion algorithms produced the worst predictions over highly forested areas (T13TCF and T15RXQ) compared to other sites, indicating larger spectral and spatial inaccuracies. The mean RMSEs of S2L fusion methods among all sites were 0.0540 (S2L.PredictMode) and 0.0527 (S2L.CompositeMode). These quantitative results align with the visual comparisons, confirming TIF's superior performance, with the lowest fusion errors and minimal variability across dates and locations.

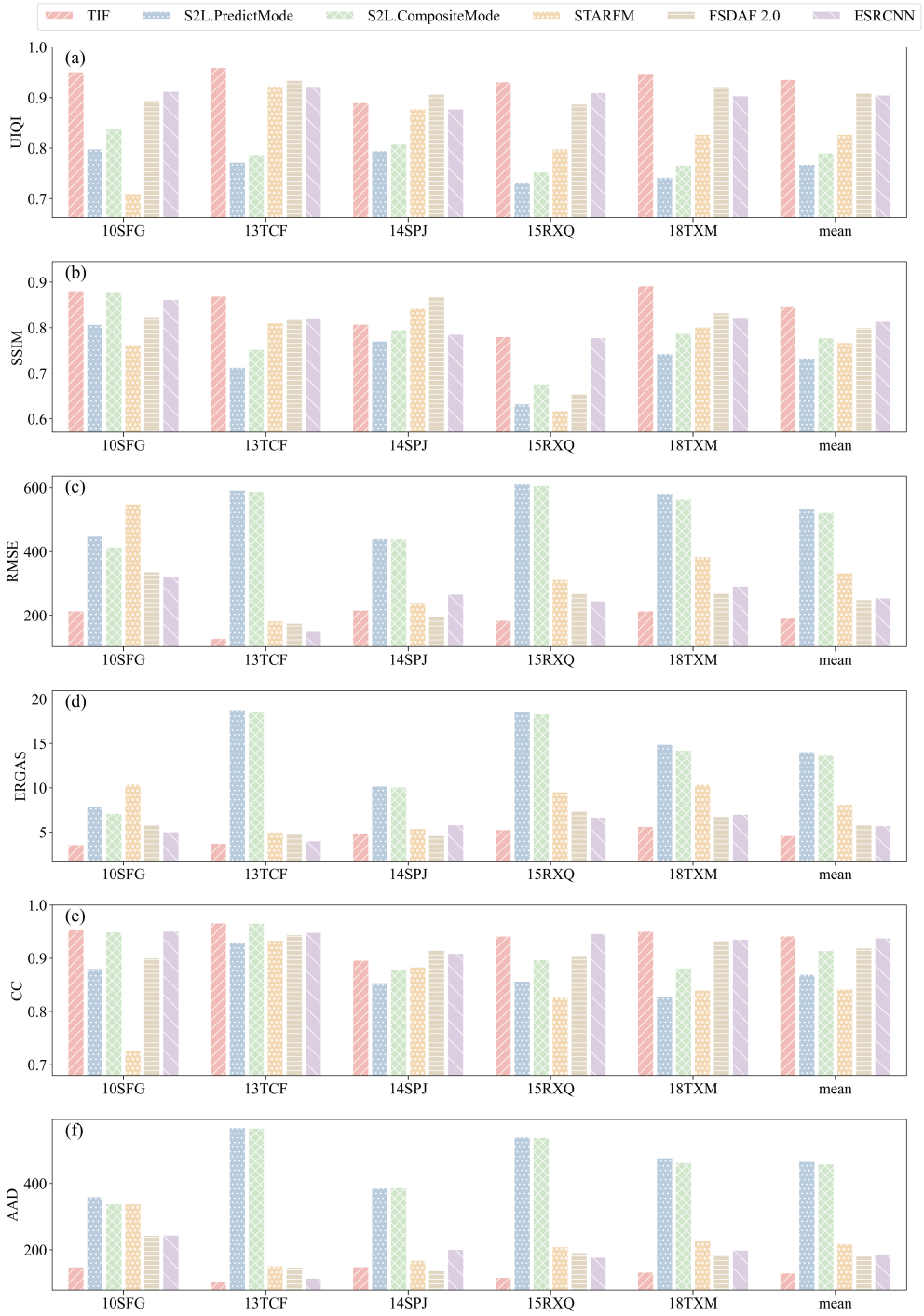


Figure 11. Quantitative spectral and spatial evaluation for all study areas and the mean results using six metrics: (a) UIQI, (b) SSIM, (c) RMSE, (d) ERGAS, (e) CC, (f) AAD. Values averaged over six spectral bands. Colors represent different algorithms: Orange for TIF, Blue for S2L.PredictMode, Green for S2L.CompositeMode, Yellow for STARFM, Brown for FSDAF 2.0, and Purple for ESRCNN. Note: The values of RMSE and AAD are enlarged by 10^4. *UIQI: Universal Image Quality Index. SSIM: Structure SIMilarity. RMSE: Root Mean Square Error;*

ERGAS: Erreur Relative Global Adimensionnelle de Synthese. CC: Correlation Coefficient; AAD: Absolute Average Difference.

Figure 12 shows the quantitative evaluation of the spatial fidelity of the fusion algorithms based on the averaged absolute value of the semivariance difference. The TIF algorithm demonstrated the smallest error, remaining below 1×10^3 for all lag distances, indicating a higher spatial fidelity compared to other methods.

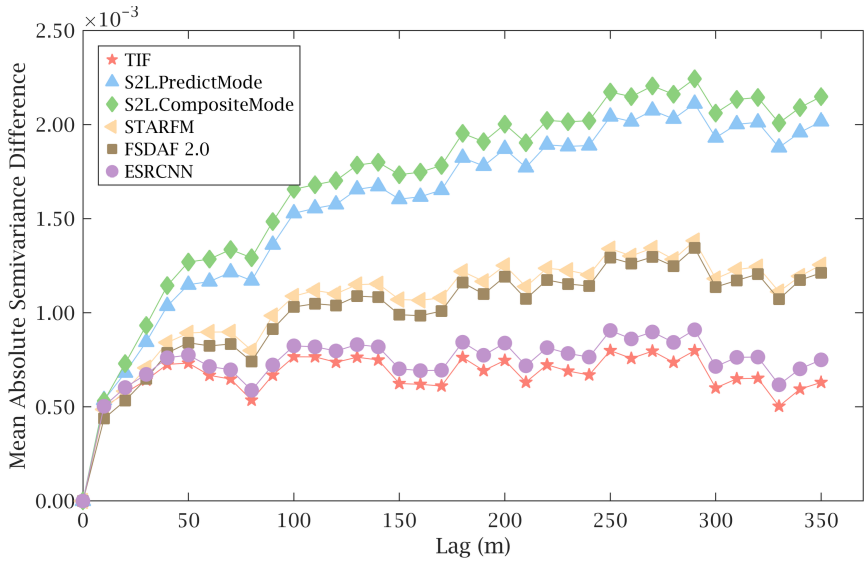


Figure 12. Quantitative spatial assessment of TIF, S2L, STARFM, FSDAF 2.0, and ESRCNN, averaged by all validation sites.

5.3. Change Detection Results

The effectiveness of each fusion method for change detection was assessed by comparing land change maps generated from the predicted and reference surface reflectance data. Figure 13 displays the change maps for a heterogeneous landscape (the same area shown in Figure 8). TIF outperformed the other methods, accurately capturing vegetation changes and small-scale urban modifications (subarea 2). STARFM and FSDAF 2.0 produced more commission errors (red circles). ESRCNN is generally good, but showed more commission errors across heterogeneous landscapes (Figure 13l), including some resulting from artifacts introduced by the deep learning model, such as artificial lines that are not present in the original imagery. S2L exhibited more

omission errors (yellow circles), and S2L.PredictMode introduced noticeable salt-and-pepper noise (purple circles, Figure 13u). For the homogeneous forest harvest area (Figure S7), TIF, STARFM, FSDAF 2.0, and ESRCNN delivered comparable results. However, the S2L-derived change maps underestimated the extent of tree removals, resulting in substantial omission errors. Across the remaining validation sites, TIF demonstrated consistent and accurate performance, effectively identifying diverse surface changes. For example, TIF successfully captured the spatial patterns of subtle vegetation recovery signals after selective logging, which were missed by other methods (Figure S14). Furthermore, TIF provided more accurate delineation of urban growth and modification areas compared to the omission-prone outputs from FSDAF 2.0, ESRCNN, and S2L (Figure S16). These findings highlight TIF's superiority in capturing fine-scale changes across diverse land cover types and landscape conditions.

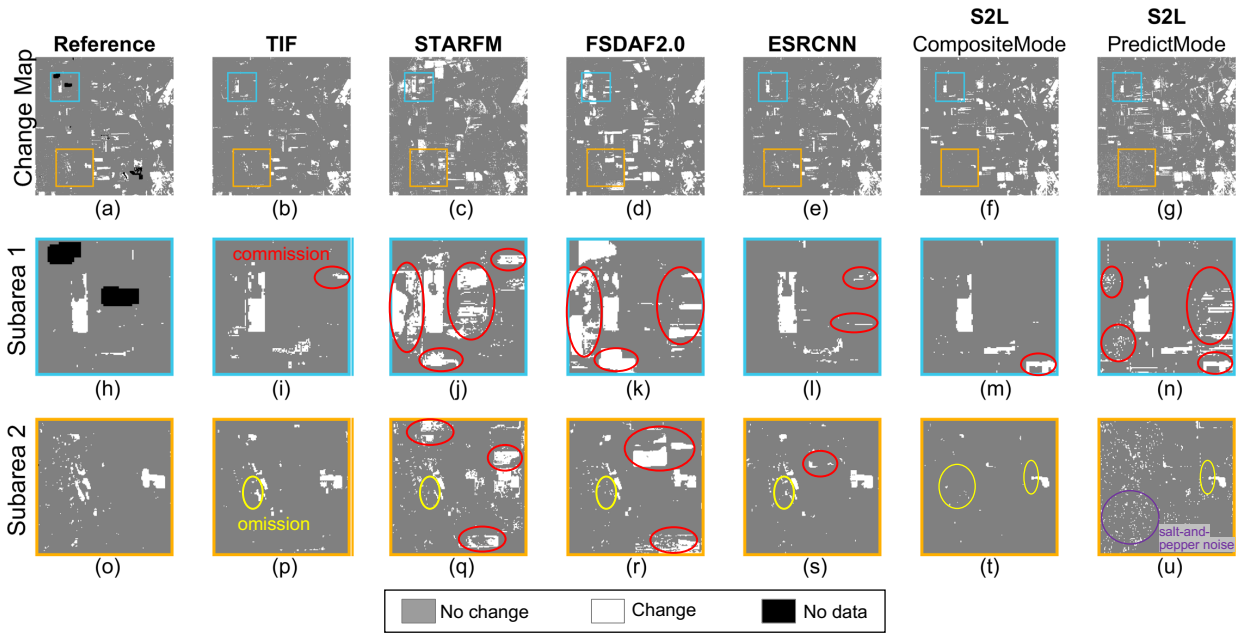


Figure 13. Change map comparisons with the same extent as **Figure 8** are shown in (a) – (g). The zoom-in views of subarea 1 enclosed in blue boxes with the same extent of **Figure 9** are shown in (h) – (n). The magnificent views of subarea 2 enclosed in orange boxes with the same extent of **Figure 10** are shown in (o) – (u). Grey means stable surfaces, white depicts changed areas, and black indicates invalid data from the QA band.

Figure 14 summarizes the quantitative evaluation of the change maps among different sites. Higher F1 scores and lower disagreement rates indicate better similarity to the reference maps. TIF method consistently ranks among the top performers, achieving the highest mean F1 (0.70) and the lowest mean disagreement rate (0.05), followed by ESRCNN (F1 = 0.67, disagreement = 0.07) and FSDAF 2.0 (F1 = 0.66, disagreement = 0.07). Both S2L methods performed worse, particularly S2L.PredicMode, which has the lowest F1 scores and the most omission errors. A detailed change detection comparison including resampling results (nearest, bilinear, and bicubic) can be found in Figure S21. These results demonstrate that TIF predictions align more closely with the reference data, highlighting its effectiveness in capturing diverse land changes across a range of landscapes.

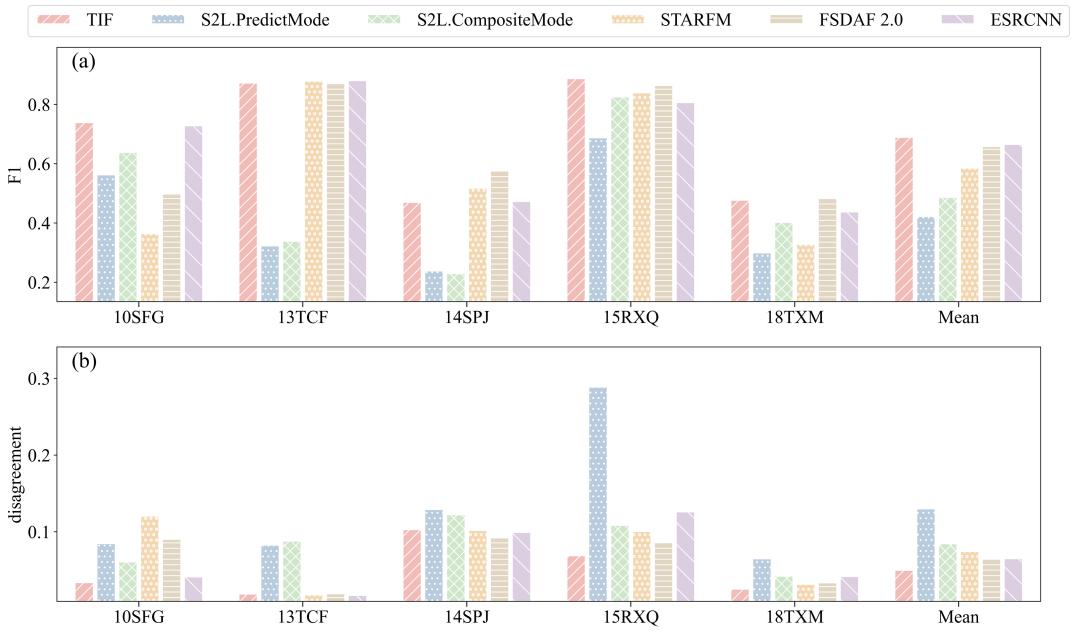


Figure 14. F1 scores and disagreements of change maps for the five test regions and the mean values. Different colors and textures represent results from different fusion methods.

6. Discussion

The Time-series-based Image Fusion (TIF) algorithm presented here offers a novel approach to generating 10-m Harmonized Landsat-Sentinel (HLS) data, consistently outperforming state-of-the-art methods in our experiments. Its success stems from several key innovations that directly address long-standing challenges in spatiaotemporal fusion.

TIF's most significant advantage lies in its full exploitation of the temporal domain. Unlike methods reliant on limited image pairs or short time series (Chen et al., 2023; Goyena et al., 2023), TIF leverages the entire historical record of pixel-level observations. This obviates the need for image-pair matching and allows TIF to uniquely accommodate inherent differences in atmospheric correction (e.g., LaSRC vs. Sen2Cor) and sensor bandpasses without explicit harmonization steps. The core outputs of TIF are reusable coefficients, which enable a highly efficient "predict-mode" (Figure S17-S20). New Landsat observations can be downscaled to 10-m on-demand without retraining the model, a crucial feature for large-scale operational workflows. Our analysis showed that these pre-calculated coefficients maintain high accuracy, even when predicting events like wildfires that occurred outside the initial coefficient development period (Figure S19). We also quantitatively evaluated the fusion performance in the "predict-mode" and found that, although accuracy decreases slightly in these scenarios (Figure S18), likely due to challenges in adapting to novel, heterogeneous sub-pixel change (Figure S20), model continues to yield high performance (e.g. SSIM>0.85, RMSE<0.0025), demonstrating strong spatial and spectral fidelity. This "train once, predict many" paradigm, combined with a pixel-based parallel processing strategy, makes TIF computationally efficient and scalable.

The enhanced spatiotemporal resolution of TIF-generated data also creates new scientific opportunities. The dense 10-m time series can improve the detection of rapid and fine-scale land

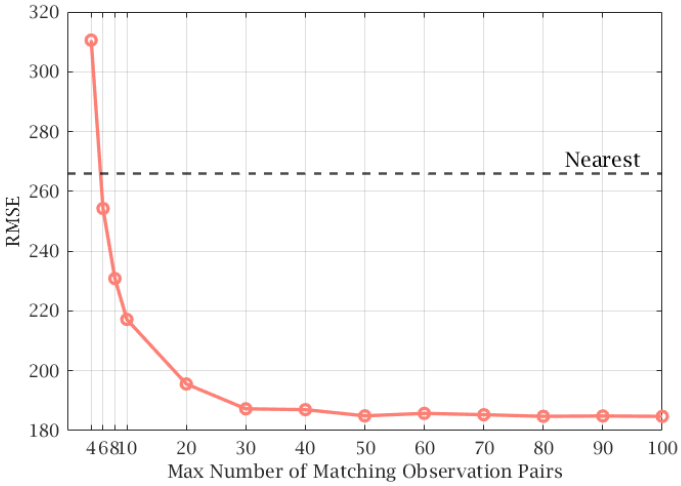
changes often missed by coarser data (Gao et al., 2015) and can serve as a higher-resolution input for time-series analysis algorithms like the Continuous Change Detection and Classification (CCDC) method (Zhu et al., 2014), thereby enhancing their applicability in heterogeneous landscapes.

When compared to other methods, TIF's advantages become clear. Weight-function-based and hybrid methods like STARFM and FSDAF 2.0 perform well when high-quality image pairs are available but struggle in heterogeneous landscapes and with predicting small changes. A deeper analysis of the results reveals why TIF's design leads to superior performance in specific contexts. In regions with high phenological variability like the agricultural T10SFG site, paired methods such as STARFM and FSDAF 2.0 are highly sensitive to the specific state captured in their input pairs. TIF, by contrast, leverages the entire time series to learn a more complete model of the typical phenological cycle for each pixel. Its temporal clustering can identify distinct states, allowing for more robust predictions even during periods of rapid change. This reduced reliance on ideal, synchronous image pairs also makes TIF inherently more resilient to temporal data gaps caused by cloud cover. Furthermore, the use of robust regression mitigates the influence of noise and outliers (e.g., undetected cloud contamination) that can disproportionately affect pair-based methods, while the pixel-level modeling inherently accounts for systematic inter-sensor differences. It is this combination of long-term learning, adaptive change modeling, and noise reduction that explains TIF's consistently strong performance across diverse and challenging conditions. Deep learning approaches like ESRCNN can achieve comparable accuracy to TIF but require significant computational resources (e.g., GPU acceleration) and depend on clear-sky Sentinel-2 imagery near the prediction date, which is often unavailable. Furthermore, CNN-based methods can introduce over-smoothing artifacts along feature boundaries. In contrast, TIF's

841 design, which requires no image pairs or model retraining for prediction, significantly reduces both
842 data dependency and computational demand (Table S7), making it more suitable for large-scale
843 and near-real-time applications. Our test using Matlab R2023b on an AMD EPYC 7713 system
844 with 512 GB memory showed a processing time of 0.4–0.7 s per pixel per CPU core for TIF
845 coefficient generation. The storage requirement for the TIF coefficients depends on the number of
846 temporal clusters (k_optimal) identified for each pixel. For global land coverage (approximately
847 18,500 Sentinel-2 tiles), we estimate the total storage would range from ~49 TB if all pixels were
848 stable (k_optimal=1) to a maximum of ~227 TB if all pixels required the most complex modeling
849 (k_optimal=3), as detailed in Table S7. Based on the cluster distribution observed across our
850 diverse U.S. test sites (where heterogeneous pixels requiring more clusters are common), a more
851 realistic global estimate is approximately 153 TB. This storage demand is highly manageable for
852 an operational system, especially considering that only these small coefficient files need to be
853 archived, rather than the voluminous raw pixel-level time series.

854 Despite its advantages, TIF has limitations that warrant further investigation. Performance
855 is reduced over water bodies and areas with frequent water/land transitions, likely attributable to
856 uncertainties in the Sentinel-2 atmospheric correction over water (Warren et al., 2019). We also
857 acknowledge that blocky patterns can appear in intensively managed agricultural areas, likely due
858 to rapid phenological changes that are challenging for a model that relies solely on temporal
859 information. While TIF's pixel-wise design avoids the need for image pairs and ensures
860 algorithmic simplicity, it entails a trade-off between scalability and local consistency in highly
861 heterogeneous landscapes. To mitigate potential artifacts due to the absence of spatial
862 neighborhood information, future work could explore spatially adaptive strategies, such as
863 integrating neighboring pixels or edge-aware regularization.

The algorithm's accuracy is also dependent on the number of available historical matched observation pairs. Our sensitivity analysis (Figure 15) based on 10,000 simple random calibration sample units (grey circles in Figure 1) demonstrates that TIF's performance (measured by RMSE) improves sharply as the number of pairs increases to approximately 30, after which the accuracy gains stabilize. Notably, TIF outperforms a simple nearest-neighbor baseline with as few as six observation pairs, establishing a minimum threshold for effective performance. To assess the practical applicability of TIF at a global scale, we analyzed the availability of these observation pairs over a 10-year period for 100,000 simple random sample units. This analysis confirmed that over 96% of the world's land surface meets the minimum performance threshold (≥6 pairs) and over 93% meets the optimal threshold (≥30 pairs) (Figure 16a, c). As expected, observation pair availability is primarily limited by cloud cover (Figure 16b). Nonetheless, TIF remains viable even in some of the cloudiest regions (up to 99% cloud cover), where the minimum of six pairs is still often attainable (blue line in Figure 16a). For persistently low-density regions, such as parts of the tropics or high-latitude areas with extensive snow and ice, adaptive strategies like extending the temporal matching window beyond 16 days will be necessary to gather sufficient data for robust coefficient derivation.



880

881 **Figure 15.** Sensitivity of TIF performance to the number of matching observation pairs. The
882 analysis is based on 10,000 simple random calibration sample units (grey circles in **Figure 1**). The
883 curve shows the Root Mean Square Error (RMSE) as a function of the maximum number of pairs
884 used in the TIF regression. The dashed line indicates the RMSE of a simple nearest-neighbor
885 resampling, serving as a baseline for comparison.

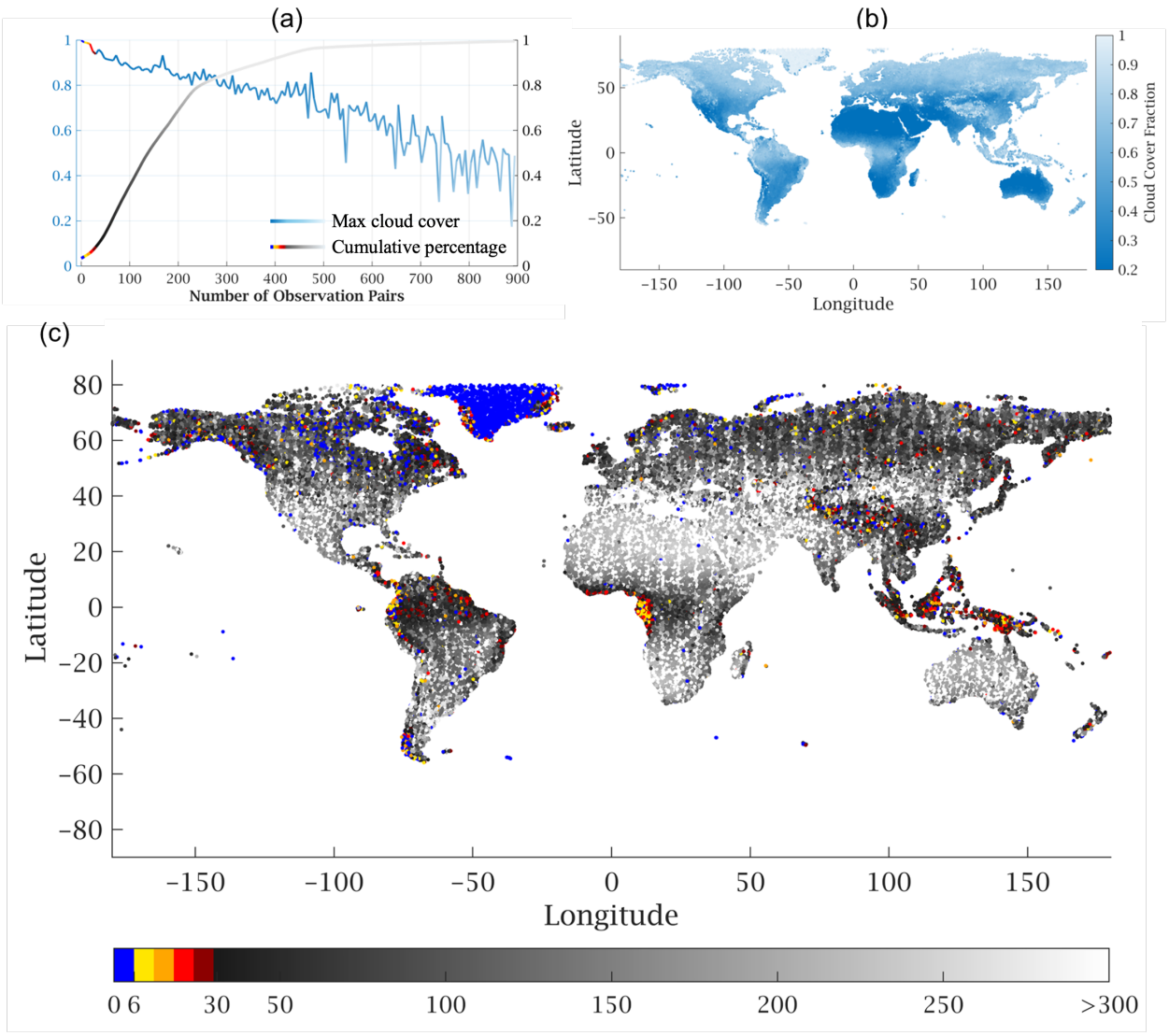


Figure 16. Global analysis of observation pair availability from 100,000 simple random sample units. (a) Relationship between the maximum cloud cover fraction (blue line) and the resulting number of matched observation pairs, alongside the cumulative percentage of observation pairs found globally (gray line). (b) Global map of mean cloud cover fraction (2013–2024). (c) Global spatial distribution of the number of matched L30-S10 observation pairs over a 10-year period. Dot colors in (c) correspond to the observation pair counts shown on the x-axis of (a), with saturation capped at 300 for high-density areas.

Methodologically, several avenues exist for future enhancement. The fixed *k_max* value of 3, while effective, could be replaced by more adaptive approaches like hierarchical clustering to better capture complex change histories in very long time series. For near-real-time applications

using pre-calculated coefficients, prediction accuracy can be reduced in rare cases of novel heterogeneous change (Figure S20). This could be addressed by developing more efficient coefficient updating strategies, such as recursive regression, which would enable fast updates only in areas of detected change. Finally, it is important to acknowledge that our validation of NIR/SWIR bands using downsampled 60-m imagery may underrepresent the algorithm's true performance on fine-scale spatial textures.

Finally, the TIF framework is extensible. Future work should explore its adaptation for other satellite missions (e.g., PlanetScope, WorldView) and the synergistic integration of other data sources. For example, incorporating the Landsat 15-m panchromatic band could enhance spatial detail for Landsat 8/9-focused applications (Li et al., 2017), while integrating SAR data (e.g., Sentinel-1) could improve performance over water and in cloudy regions. While global-scale implementation presents practical challenges related to data management and regional heterogeneity, TIF's efficiency and robust design make it a strong candidate to underpin a future operational 10-m global HLS product.

7. Conclusions

In this study, we presented TIF, a novel time-series-based image fusion method, for generating 10-m resolution surface reflectance imagery by synthesizing Landsat 8-9 and Sentinel-2 A/B data. TIF uniquely leverages long-term pixel-level observations, eliminating reliance on prior image pairs or ancillary data. Our experiment results showed that TIF accurately produces 10-m Sentinel-2-like surface reflectance images for any clear-sky Landsat observation date, effectively reconstructing fine-resolution time series and facilitating improved land change detections. Moreover, TIF's highly efficient pixel-based parallel processing framework enables

operational implementation for large-scale applications. These capabilities highlight TIF's potential as an exploratory approach toward a 10-m HLS product. While further validation is needed for operational use, such an advancement would significantly expand the existing NASA HLS suite and unlock new possibilities for research and monitoring efforts requiring high spatial and temporal resolutions in Earth observations.

Acknowledgment

This study was supported by the UConn Eversource Energy Center (Near Real-time Assessment of Forest Risk to Infrastructure Using Satellite Time Series) and NASA (Improvements of QA Band and New Science Data Layers Proposed for the NASA Harmonized Landsat and Sentinel-2 Products, grant number 80NSSC23K0773).

Data availability

All satellite images are available online from NASA, USGS, and Copernicus. Our Matlab code for the Time-series-based Image Fusion (TIF) method and sample data are available at GitHub repository https://github.com/GERSL/TIF. The test images for reproducing paper results and for future spatial-temporal fusion algorithm evaluation can be obtained via the GitHub page as well. The STARFM source code was obtained from USDA website (https://www.ars.usda.gov/research/software/download/?softwareid=432&modecode=80-42-05-10). The FSDAF 2.0 MATLAB executable was downloaded from https://github.com/Andy-cumt/FSDAF2_GUI. The Sen2Like package can be accessed at https://github.com/senbox-org/sen2like. The ESRCNN code can be found at https://github.com/MrCPlusPlus/ESRCNN-for-Landsat8-Sentinel2-Fusion.

940

Reference

Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., & Friedl, M.A. (2020). Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sensing of Environment, 240*, 111685

Breiman, L. (2001). Random forests. *Machine learning, 45*, 5-32

Chen, S., Wang, J., & Gong, P. (2023). ROBOT: A spatiotemporal fusion model toward seamless data cube for global remote sensing applications. *Remote Sensing of Environment*, *294*, 113616.

Chen, Y., Song, X., Wang, S., Huang, J., & Mansaray, L.R. (2016). Impacts of spatial heterogeneity on crop area mapping in Canada using MODIS data. *ISPRS Journal of Photogrammetry and Remote Sensing, 119*, 451-461

Chen, N., Tsendbazar, N.-E., Hamunyela, E., Verbesselt, J., & Herold, M. (2021). Sub-annual tropical forest disturbance monitoring using harmonized Landsat and Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation, 102*, 102386

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., & Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment, 219*, 145-161

Curran, P.J. (1988). The semivariogram in remote sensing: an introduction. *Remote Sensing of Environment, 24*, 493-507

Dewitz, J., 2023, National Land Cover Database (NLCD) 2021 Products: U.S. Geological Survey data release, [Accessed on 06/03/2024].

Doxani, G., Vermote, E.F., Roger, J.-C., Skakun, S., Gascon, F., Collison, A., De Keukelaere, L., Desjardins, C., Frantz, D., & Hagolle, O. (2023). Atmospheric Correction Inter-comparison eXercise, ACIX-II Land: An assessment of atmospheric correction processors for Landsat 8 and Sentinel-2 over land. *Remote Sensing of Environment, 285*, 113412

Duan, M., Wang, Z., Sun, L., Liu, Y., & Yang, P. (2024). Monitoring apple flowering date at 10 m spatial resolution based on crop reference curves. *Computers and Electronics in Agriculture*, 225, 109260.

DuMouchel, W. H., and F. L. O'Brien. (1989). Integrating a Robust Option into a Multiple Regression Computing Environment. *Computer Science and Statistics*: *Proceedings of the 21st Symposium on the Interface*. Alexandria, VA: American Statistical Association.

Falanga Bolognesi, S., Pasolli, E., Belfiore, O.R., De Michele, C., & D'Urso, G. (2020). Harmonized landsat 8 and sentinel-2 time series data to detect irrigated areas: An application in Southern Italy. *Remote Sensing, 12*, 1275

Fassnacht, F.E., White, J.C., Wulder, M.A. and Næsset, E., 2024. Remote sensing in forestry: Current challenges, considerations and directions. *Forestry: An International Journal of Forest Research*, 97(1), pp.11-37.

Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., & Yang, Y. (2015). Fusing Landsat and MODIS Data for Vegetation Monitoring. *IEEE Geoscience and Remote Sensing Magazine, 3*, 47-60

Gao, F., 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*, 2207-2218

Gao, F., Masek, J.G., & Wolfe, R.E. (2009). Automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing, 3*, 033515

Gao, F., & Zhang, X. (2021). Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. *Journal of Remote Sensing*.

Ghamisi, P., Gloaguen, R., Atkinson, P.M., Benediktsson, J.A., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., & Anders, K. (2019). Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art. *IEEE Geoscience and Remote Sensing Magazine, 7*, 6-39

Goyena, H., Pérez-Goya, U., Montesino-SanMartin, M., Militino, A.F., Wang, Q., Atkinson, P.M., & Ugarte, M.D. (2023). Unpaired spatio-temporal fusion of image patches (USTFIP) from cloud covered images. *Remote Sensing of Environment, 295*, 113709

Guo, D., Shi, W., Hao, M., & Zhu, X. (2020). FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details. *Remote Sensing of Environment, 248*, 111973

Hilker, T., Wulder, M. A., Coops, N. C., Seitz, N., White, J. C., Gao, F., ... & Stenhouse, G. (2009a). Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, *113*(9), 1988-1999.

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., & White, J.C. (2009b). A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment, 113*, 1613-1627

Holland, Paul W., and Roy E. Welsch. "Robust regression using iteratively reweighted least-squares." Communications in Statistics-theory and Methods 6.9 (1977): 813-827.

Ju, J., Zhou, Q., Freitag, B., Roy, D. P., Zhang, H. K., Sridhar, M., ... & Neigh, C. S. (2025). The Harmonized Landsat and Sentinel-2 version 2.0 surface reflectance dataset. *Remote Sensing of Environment*, *324*, 114723.

Kennedy, R. E., Andréfouët, S., Cohen, W. B., Gómez, C., Griffiths, P., Hais, M., ... & Zhu, Z. (2014). Bringing an ecological view of change to Landsat-based remote sensing. *Frontiers in Ecology and the Environment*, *12*(6), 339-346.

Kim, Y., Kyriakidis, P. C., & Park, N. W. (2020). A cross-resolution, spatiotemporal geostatistical fusion model for combining satellite image time-series of different spatial and temporal resolutions. *Remote Sensing*, 12(10), 1553.

Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., & He, Y. (2013). Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. *Remote Sensing of Environment, 135*, 52-63

Li, X., Foody, G.M., Boyd, D.S., Ge, Y., Zhang, Y., Du, Y., & Ling, F. (2020). SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. *Remote Sensing of Environment, 237*, 111537

Li, Z., Zhang, H. K., Roy, D. P., Yan, L., Huang, H., & Li, J. (2017). Landsat 15-m panchromatic-assisted downscaling (LPAD) of the 30-m reflective wavelength bands to Sentinel-2 20-m resolution. *Remote Sensing*, 9(7), 755.

Liao, C., Wang, J., Shan, B., Shang, J., Dong, T., & He, Y. (2023). Near real-time detection and forecasting of within-field phenology of winter wheat and corn using Sentinel-2 time-series data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 105-119.

Liu, M., Ke, Y., Yin, Q., Chen, X., & Im, J. (2019). Comparison of five spatio-temporal satellite image fusion models over landscapes with various spatial heterogeneity and temporal variation. *Remote Sensing*, 11(22), 2612.

Liu, S., Zhou, J., Qiu, Y., Chen, J., Zhu, X., & Chen, H. (2022). The FIRST model: Spatiotemporal fusion incorrporting spectral autocorrelation. *Remote Sensing of Environment*, *279*, 113111.

Liu, Y., Gu, X., Cheng, T., Zhan, Y., Zhang, H., Li, J., ... & Zhang, Y. (2022). Temporal Shape-Based Fusion Method to Generate Continuous Vegetation Index at Fine Spatial Resolution. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-14.

Ma, Y., Shen, Y., Shen, G., Wang, J., Xiao, W., He, H., ... & Qin, K. (2023). STEPSBI: Quick spatiotemporal fusion with coarse-and fine-resolution scale transformation errors and pixel-based synthesis base image pair. *ISPRS Journal of Photogrammetry and Remote Sensing*, *206*, 1-15.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). Sen2Cor for sentinel-2. In, *Image and signal processing for remote sensing XXIII* (pp. 37-48): SPIE

Olofsson, P. and Virts, K. (2022). Leveraging Satellite Assets for Better Decision-Making: the Satellite Needs Working Group. *AGU Fall Meeting Abstracts* (Vol. 2022, pp. SY12B-0393).

Parreiras, T.C., Bolfe, É.L., Chaves, M.E.D., Sanches, I.D.A., Sano, E.E., Victoria, D.d.C., Bettiol, G.M., & Vicente, L.E. (2022). Hierarchical classification of soybean in the Brazilian savanna based on Harmonized Landsat Sentinel data. *Remote Sensing, 14*, 3736

Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L., & Zhu, Z. (2018). Making Landsat time series consistent: Evaluating and improving Landsat analysis ready data. Remote Sensing, 11(1), 51.

Qiu, S., Zhu, Z., & He, B. (2019). Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment, 231*, 111205

Qiu, S., Zhu, Z., Olofsson, P., Woodcock, C.E., & Jin, S. (2023). Evaluation of Landsat image compositing algorithms. *Remote Sensing of Environment, 285*, 113375

Radeloff, V. C., Roy, D. P., Wulder, M. A., Anderson, M., Cook, B., Crawford, C. J., ... & Zhu, Z. (2024). Need and vision for global medium-resolution Landsat and Sentinel-2 data products. Remote Sensing of Environment, 300, 113918.

Roy, D.P., Li, J., Zhang, H.K., Yan, L., Huang, H., & Li, Z. (2017). Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment, 199*, 25-38

Roy, D.P., Zhang, H., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C., Sun, Q., Li, J., Huang, H., & Kovalskyy, V. (2016). A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment, 176*, 255-271

Saunier, S., Pflug, B., Lobos, I. M., Franch, B., Louis, J., De Los Reyes, R., ... & Kocaman, S. (2022). Sen2Like: Paving the way towards harmonization and fusion of optical data. *Remote Sensing*, *14*(16), 3855.

Shang, R., & Zhu, Z. (2019). Harmonizing Landsat 8 and Sentinel-2: A time-series-based reflectance adjustment approach. Remote Sensing of Environment, 235, 111439.

Shang, R., Zhu, Z., Zhang, J., Qiu, S., Yang, Z., Li, T., & Yang, X. (2022). Near-real-time monitoring of land disturbance with harmonized Landsats 7–8 and Sentinel-2 data. Remote Sensing of Environment, 278, 113073.

Shao, Z., Cai, J., Fu, P., Hu, L., & Liu, T. (2019). Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sensing of Environment, 235*, 111425

Shen, Y., Zhang, X., Yang, Z., Ye, Y., Wang, J., Gao, S., ... & Ju, J. (2023). Developing an operational algorithm for near-real-time monitoring of crop progress at field scales by fusing harmonized Landsat and Sentinel-2 time series with geostationary satellite observations. *Remote Sensing of Environment*, *296*, 113729.

Singh, A. (1989). Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing, 10*, 989-1003

Suh, J. W., Zhu, Z., & Zhao, Y. (2024). Monitoring construction changes using dense satellite time series and deep learning. Remote Sensing of Environment, 309, 114207.

Thornton, P. E., Reed, B. C., Xian, G. Z., Chini, L., East, A. E., Field, J. L., ... & Zhu, Z. (2023). Land cover and land-use change. Crimmins, AR, Avery, CW, Easterling, DR, Kunkel, KE, Stewart, BC, Maycock, TK (E ds.), Fifth National Climate Assessment. US Global Change Research Program, Washington, DC, USA. doi, 10.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411-423.

Tulbure, M.G., Broich, M., Perin, V., Gaines, M., Ju, J., Stehman, S.V., Pavelsky, T., Masek, J.G., Yin, S., & Mai, J. (2022). Can we detect more ephemeral floods with higher density harmonized Landsat Sentinel 2 data compared to Landsat 8 alone? *ISPRS Journal of Photogrammetry and Remote Sensing, 185*, 232-246

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens Environ, Volume 185*, 46-56

Wald, L. 2002. Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions. Paris, France: Presses des MINES.

Wald, L., Ranchin, T., & Mangolini, M. (1997). Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric engineering and remote sensing, 63*, 691-699

Wang, Q., Shi, W., Li, Z., & Atkinson, P. M. (2016). Fusion of Sentinel-2 images. *Remote sensing of environment*, 187, 241-252.

Wang, Q., Blackburn, G. A., Onojeghuo, A. O., Dash, J., Zhou, L., Zhang, Y., & Atkinson, P. M. (2017). Fusion of Landsat 8 OLI and Sentinel-2 MSI data. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), 3885-3899.

Wang, Q., & Atkinson, P.M. (2018). Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment, 204*, 31-42

Wang, Q., Tang, Y., Ge, Y., Xie, H., Tong, X., & Atkinson, P. M. (2023). A comprehensive review of spatial-temporal-spectral information reconstruction techniques. *Science of Remote Sensing*, *8*, 100102.

Wang, Z., and A. C. Bovik. (2002). "A Universal Image Quality Index." IEEE Signal Processing Letters 9 (3): 81–84.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. (2004). "Image Quality Assessment: From Error Visibility to Structural Similarity." IEEE Transactions on Image Processing 13 (4): 600–612.

Warren, M.A., Simis, S.G., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., & Ansper, A. (2019). Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sensing of Environment, 225*, 267-289

Weiss, M., Jacob, F. and Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote sensing of environment*, 236, p.111402.

Williams, C. S., & Becklund, O. A. (2002). Introduction to the optical transfer function (Vol. 112). SPIE Press.

Wu, J., Lin, L., Li, T., Cheng, Q., Zhang, C., & Shen, H. (2022). Fusing Landsat 8 and Sentinel-2 data for 10-m dense time-series imagery using a degradation-term constrained deep network. *International Journal of Applied Earth Observation and Geoinformation, 108*, 102738

Xiao, J., Aggarwal, A.K., Duc, N.H., Arya, A., Rage, U.K., & Avtar, R. (2023). A review of remote sensing image spatiotemporal fusion: Challenges, applications and recent trends. *Remote Sensing Applications: Society and Environment, 32*

Xu, Y., Huang, B., Xu, Y., Cao, K., Guo, C. and Meng, D. (2015). Spatial and temporal image fusion via regularized spatial unmixing. *IEEE Geoscience and Remote Sensing Letters*, 12(6), pp.1362-1366.

Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, *237*, 111511.

Zhang, H. K., Roy, D. P., & Kovalskyy, V. (2015). Optimal solar geometry definition for global long-term Landsat time-series bidirectional reflectance normalization. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(3), 1410-1418.

Zhang, X., Gao, F., Wang, J., & Ye, Y. (2021). Evaluating a spatiotemporal shape-matching model for the generation of synthetic high spatiotemporal resolution time series of multiple satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102545.

Zhang, Y., Woodcock, C. E., Arévalo, P., Olofsson, P., Tang, X., Stanimirova, R., ... & Friedl, M. A. (2022). A global analysis of the spatial and temporal variability of usable Landsat observations at the pixel scale. *Frontiers in Remote Sensing*, *3*, 894618.

Zhu, X., Cai, F., Tian, J., & Williams, T. (2018). Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. *Remote Sensing, 10*

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*, 2610-2623

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

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote sensing of Environment*, *144*, 152-171.

Zhu, Z., Qiu, S., & Ye, S. (2022). Remote sensing of land change: A multifaceted perspective. *Remote Sensing of Environment*, 282, 113266.

Zhu, Z., Zhou, Y., Seto, K. C., Stokes, E. C., Deng, C., Pickett, S. T., & Taubenböck, H. (2019). Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote Sensing of Environment*, *228*, 164-182.

Zhukov, B., Oertel, D., Lanzl, F., & Reinhackel, G. (1999). Unmixing-based multisensor multiresolution image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(3), 1212-1226.

List of Figure Captions

- **Figure 1**. (a) Black polygons: The five HLS tiles used for this study. Background: 2021 NLCD land cover maps (Dewitz 2023). The enlarged views of these tiles are shown at (b) Far West (T10SFG), (c) Rocky Mountains (T13TCF), (d) Great Plains (T14SPJ), (e) Southwest (T15RXQ), and (f) New England (T18TXM). Each tile covers surface areas with 109.8 km by 109.8 km. TIF calibration data includes 10,000 simple random calibration (SRC) sample (grey circles) and 5,000 stratified random change calibration (SRCC) sample (red triangles). Independent validation sites used for evaluating algorithm performance are highlighted in purple polygons. TIF: Time-series-based Image Fusion.

- **Figure 2**. Illustration of 10 m HLS imagery time series obtained from the TIF approach. The key steps of TIF with respect to the target pixel (in yellow square) are highlighted in the black box. Observation pairs from L8-9 and S2 are first matched and weighted, after which pixel-wise K-means clustering is optionally applied. In the "without clustering" branch, a single regression line is fit across all matched pairs, producing two coefficients (slope a and intercept b) that are passed downstream. In the "with clustering" branch, separate regression lines are fit for each cluster, in addition to a simple regression line across all pairs. For each spectral band, the regression RMSEs are compared, and the coefficients from the model with lower error are retained as the final TIF coefficients. TIF: Time-series-based Image Fusion. L8-9: Landsat 8 and 9. S2: Sentinel-2. m is the number of matched observation pairs. n is the total number of spectral bands (n=6).

- **Figure 3**. Illustration of TIF execution with a stable land pixel covered by extensive urban trees and a small portion of roads from Wethersfield, CT. (a) False color composite (R/G/B=NNIR/Red/Green) of the L30, TIF predictions, and hidden S10 data for the same geographic area. Both L30 and S10 were acquired on 2021-06-16. The white circle indicates the point of interest (upper left corner) within the 30 m by 30 m pixel. (b) Time series of NNIR reflectance for the point of interest (Lat: 41.7236, Lon: -72.6720). The solid red dots represent S10 data, the blue dots represent L30 data, and the black circles are TIF predictions. The dashed line represents the date of withheld Sentinel-2 image. (c) Linear relationship of the NNIR reflectance at the point of interest (solid line), built by TIF, between matching L30-S10 pairs (purple points), compared to the 1:1 line (dashed line). The black cross denotes the centroid of all matching pairs. (d) Google Earth images of the area of interest captured at different timestamps. The yellow box indicates the extent of the L30 pixel shown in (a). The white marker highlighted the location of the investigated point in (b) and (c).

- **Figure 4**. Illustration of TIF execution with a homogeneous land change pixel covered by crops in Gustine, CA. (a) False color composite (R/G/B=NNIR/Red/Green) of the L30 surface reflectance, TIF predictions, and hidden S10 data for the same geographic area. Both L30 and S10 were acquired on 2021-07-20. The white circle indicates the point of interest (middle one) within the 30 m by 30 m pixel. (b) Time series of NNIR reflectance for the point of interest (Lat: 37.2837, Lon: -120.9469) The solid red dots represent S10 data, the blue dots represent L30 data, and the black circles are TIF predictions. The dashed line represents the date of withheld Sentinel-2 image. (c) Linear relationship of the NNIR reflectance at the point of interest (solid line), built by TIF, between matching L30-S10 pairs (purple points), compared

to the 1:1 line (dashed line). The black cross denotes the centroid of all matching pairs. (d) Google Earth images of the area of interest captured at different timestamps. The yellow box indicates the extent of the L30 pixel shown in (a). The white marker highlighted the location of the investigated point in (b) and (c).

- **Figure 5**. Illustration of TIF execution with a heterogeneous land change pixel from an urban area in the north of Springfield, MA. (a) False color composite (R/G/B= NNIR/Red/Green) of the L30 surface reflectance, TIF predictions, and hidden S10 data for the same geographic area. Both L30 and S10 were acquired on 2021-06-16. The white circle indicates the point of interest (lower left corner) within the 30 m by 30 m pixel. (b) Time series of NNIR and SWIR2 reflectance for the point of interest (Lat: 42.1565, Lon: -72.5847). The solid red dots represent S10 data, and the blue dots represent L30 data. For NNIR, purple and yellow circles indicate TIF predictions for cluster 1 and 2, respectively. For SWIR2, black circles are TIF predictions. The dashed line represents the date of withheld Sentinel-2 image. (c) Linear relationships of surface reflectance at the point of interest (solid lines), built by TIF, compared to the 1:1 line (dashed line). NNIR used clustered matching L30-S10 pairs (purple and yellow points). The red and blue cross denotes the centroids of different clusters. SWIR2 used all matching L30-S10 pairs (black points) with the centroids highlighted in black cross. (d) Google Earth images of the area of interest captured at different timestamps. The yellow box indicates the extent of the L30 pixel shown in (a). The white marker highlighted the location of the investigated point in (b) and (c).

- **Figure 6**. An illustration of TIF predictions (T10SFG, 2021-07-20) versus reference surface reflectance across each spectral band at Far West. TIF: Time-series-based Image Fusion. *CC: Correlation Coefficient; RMSE: Root Mean Square Error; AAD: Absolute Average Difference; ERGAS: Erreur Relative Global Adimensionnelle de Synthese.*

- **Figure 7**. Illustration of semivariograms between the TIF prediction image (T10SFG, 2021-07-20) and the reference image of each spectral band at Far West. TIF: Time-series-based Image Fusion. The $\Delta\gamma(h)$ indicates the semivariance difference between the prediction data and the reference data at lag distance h.

- **Figure 8**. Visual comparison of the fusion results from the T10SFG experiment. The original 30 m Landsat images acquired on 2017-07-09 and 2021-07-20 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2017-07-09 from TIF, STARFM, FSDAF 2.0, ESRCNN, and S2L predict mode and S2L composite mode. The downscaled results on 2021-07-20 are correspondingly (k) – (p). The centroid of the area is Lat = 37.290544, Lon = -120.978904. Images are displayed by NNIR, red, and green bands with the same stretch scales. Two zoom-in areas are highlighted: subarea 1 – blue polygon; subarea 2 – orange polygon.

- **Figure 9**. Zoom-in views of subarea 1 marked in **Figure 8**. The original 30 m Landsat images acquired on 2017-07-09 and 2021-07-20 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2017-07-09 from TIF, STARFM, FSDAF 2.0, ESRCNN, and S2L composite mode and S2L predict

mode. (k) – (p) are corresponding downscaled images on 2021-07-20. Images are displayed by NNIR, red, and green bands with the same stretch scales.

- **Figure 10**. Zoom-in views of subarea 2 marked in **Figure 8**. The original 30 m Landsat images acquired on 2017-07-09 and 2021-07-20 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2017-07-09 from TIF, STARFM, FSDAF 2.0, ESRCNN, and S2L composite mode and S2L predict mode. (k) – (p) are corresponding downscaled images on 2021-07-20. Images are displayed by NNIR, red, and green bands with the same stretch scales.

- **Figure 11**. Quantitative spectral and spatial evaluation for all study areas and the mean results using six metrics: (a) UIQI, (b) SSIM, (c) RMSE, (d) ERGAS, (e) CC, (f) AAD. Values averaged over six spectral bands. Colors represent different algorithms: Orange for TIF, Blue for S2L.PredictMode, Green for S2L.CompositeMode, Yellow for STARFM, Brown for FSDAF 2.0, and Purple for ESRCNN. Note: The values of RMSE and AAD are enlarged by 10^4. *UIQI: Universal Image Quality Index. SSIM: Structure SIMilarity. RMSE: Root Mean Square Error; ERGAS: Erreur Relative Global Adimensionnelle de Synthese. CC: Correlation Coefficient; AAD: Absolute Average Difference.*

- **Figure 12**. Quantitative spatial assessment of TIF, S2L, STARFM, FSDAF 2.0, and ESRCNN, averaged by all validation sites.

- **Figure 13**. Change map comparisons with the same extent as **Figure 8** are shown in (a) – (g). The zoom-in views of subarea 1 enclosed in blue boxes with the same extent of **Figure 9** are shown in (h) – (n). The magnificent views of subarea 2 enclosed in orange boxes with the same extent of **Figure 10** are shown in (o) – (u). Grey means stable surfaces, white depicts changed areas, and black indicates invalid data from the QA band.

- **Figure 14**. F1 scores and disagreements of change maps for the five test regions and the mean values. Different colors and textures represent results from different fusion methods.

- **Figure 15**. Sensitivity of TIF performance to the number of matching observation pairs. The analysis is based on 10,000 simple random calibration sample units (grey circles in **Figure 1**). The curve shows the Root Mean Square Error (RMSE) as a function of the maximum number of pairs used in the TIF regression. The dashed line indicates the RMSE of a simple nearest-neighbor resampling, serving as a baseline for comparison.

- **Figure 16**. Global analysis of observation pair availability from 100,000 simple random sample units. (a) Relationship between the maximum cloud cover fraction (blue line) and the resulting number of matched observation pairs, alongside the cumulative percentage of observation pairs found globally (gray line). (b) Global map of mean cloud cover fraction (2013–2024). (c) Global spatial distribution of the number of matched L30-S10 observation pairs over a 10-year period. Dot colors in (c) correspond to the observation pair counts shown on the x-axis of (a), with saturation capped at 300 for high-density areas.

Supplementary Materials

Table S1. A summary of land cover and major disturbances for the HLS tiles used in this study.

Evaluation site/ MGRS tiles	Common land cover types	Major disturbances
Far West (T10SFG)	Agriculture, shrubs, and urban areas	Wildfire and urban expansion
Rocky Mountains (T13TCF)	Forest and shrubs	Harvest, wildfires, insect outbreaks, and landslides
Great Plains (T14SPJ)	Agriculture and grassland	Agricultural activities
Southeast (T15RXQ)	Forests and wetland	Timber harvesting, agricultural expansion, and urban development
New England (T18TXM)	Forests and urban areas	Urbanization, insect outbreaks, and windstorms

Table S2. Acquisition dates of test images used for evaluation.

Tile Name	Day 1	Day 2
T10SFG	2017-07-09	2021-07-20
T13TCF	2018-07-04	2021-07-28
T14SPJ	2018-07-19	2021-08-12
T15RXQ	2018-06-30	2021-07-24
T18TXM	2016-06-18	2021-06-16

Table S3. Acquisition dates of S2L input images.

Tile Name	WRS Path/Row	Landsat day 1	Prior Sentinel-2 images used for fusion	Landsat day 2	Prior Sentinel-2 images used for fusion
T10SFG	P43/R34	2017-07-09	2017-06-26, 2017-06-29	2021-07-20	2021-07-15, 2021-07-18
T13TCF	P35/R31	2018-07-04	2018-06-26, 2018-07-01	2021-07-28	2021-07-23, 2021-07-25
T14SPJ	P28/R33	2018-07-19	2018-07-14, 2018-07-17	2021-08-12	2021-08-07, 2021-08-10
T15RXQ	P23/R39	2018-06-30	2018-06-15, 2018-06-25	2021-07-24	2021-07-04, 2021-07-14
T18TXM	P13/R31	2016-06-18	2016-04-18, 2016-05-28	2021-06-16	2021-06-06, 2021-06-11

Table S4. Acquisition dates of ESRCNN input images.

Tile Name	Landsat day 1	Clear Sentinel-2 image used for fusion	Landsat day 2	Clear Sentinel-2 image used for fusion
T10SFG	2017-07-09	2017-06-29	2021-07-20	2021-07-15
T13TCF	2018-07-04	2018-07-01	2021-07-28	2021-07-25
T14SPJ	2018-07-19	2018-07-04	2021-08-12	2021-08-07
T15RXQ	2018-06-30	2018-07-20	2021-07-24	2021-08-23
T18TXM	2016-06-18	2016-05-28	2021-06-16	2021-06-06

Table S5. Hyperparameters and configurations of the ESRCNN training.

	Sentinel-2 self fusion	Landsat and Sentinel-2 fusion
Batch size	64	64
Learning rate	0.001	0.0001
Learning scheme	MultiStepLR	MultiStepLR
Learning rate gamma	0.5	1
Number of iterations	30,000	30,000

Table S6. Acquisition dates of STARFM and FSDAF2.0 input images.

Tile Name	**Landsat day 1**	**Image pair for day 1**	**Landsat day 2**	**Image pair for day 2**
T10SFG	2017-07-09	2017-06-23 (L8) 2017-06-26 (S2)	2021-07-20	2021-06-18 (L8) 2021-06-15 (S2)
T13TCF	2018-07-04	2018-09-06 (L8) 2018-09-04 (S2)	2021-07-28	2021-07-12 (L8) 2021-07-10 (S2)
T14SPJ	2018-07-19	2018-07-03 (L8) 2018-07-04 (S2)	2021-08-12	2021-07-27 (L8) 2021-07-28 (S2)
T15RXQ	2018-06-30	2018-04-11 (L8) 2018-04-16 (S2)	2021-07-24	2021-11-08 (L8) 2021-11-16 (S2)
T18TXM	2016-06-18	2016-09-22 (L8) 2016-09-27 (S2)	2021-06-16	2021-03-12 (L8) 2021-03-13 (S2)

Table S7. Summary of spatial-temporal fusion algorithms employed in this paper.

Name of method	Category	Input requirement	Programming Language	Computing Time (s/pixel)	Reference
TIF	T	Coarse and fine resolution time series	Matlab	Training: (0.4 - 0.7 s) / CPU core * pixel Testing: $2.50*10^{-5}$	This study
S2L	I	One or two fine resolution images prior to the target date	Python	Predict mode: $1.460*10^{-6}$ Composite mode: $1.610*10^{-6}$	Saunier et al. (2022)
STARFM	W	One or more pairs	C++	$2.278*10^{-4}$	Gao et al. (2006)
FSDAF 2.0	H	One pair	Matlab	$2.972*10^{-4}$	Guo et al. (2020)
ESRCNN	D	One clear-sky fine resolution image on a neighboring date	Python	Training: 0.48 (GPU) Testing: 1.667 $*10^{-5}$ (GPU)	Shao et al. (2019)

* T: Time-series-based; I: Interpolation; W: Weight-function-based; H: Hybrid; D: Deep learning.

** All methods need one coarse image at prediction date as input, so the only other required input data are listed in the table.

*** One pair means one clear-sky image with finer spatial resolution and one clear-sky image with coarser spatial resolution acquired on the same or nearly the same date as the finer resolution image.

**** The storage requirement of TIF coefficient files depends on the number of clusters (k) used. The calculations are as follows: k=1: Each pixel requires slopes (int16) and intercepts (int16) for six spectral bands, totaling (2+2) × 6 = 24 bytes per pixel. k = 2: In addition to the base slopes and intercepts, one extra variable (band_k2, int16) is needed to identify the bands requiring a secondary set of coefficients, along with a 6 × 2 centroid array (int16) to determine which cluster applies to new Landsat data. The storage demand is (2+2) × 6 + 2 + 6 × 2 × 2 + (2+2) × n = 50 + 4n bytes, where n (1–6) is the number of bands needing two coefficient sets. This results in 54–74 bytes per pixel. k = 3: An additional variable (band_k3, int16) is introduced to identify bands requiring a third set of coefficients, along with a 6 × 3 centroid array (int16). The total size is (2+2) × 6 + 2 + 2 + 6 × 3 × 2 + (2+2) × n + (2+2) × m = 64 + 4n + 4m bytes, where n and m (1–6) are the numbers of bands requiring two and three sets of coefficients, respectively. This results in 72–112 bytes per pixel.

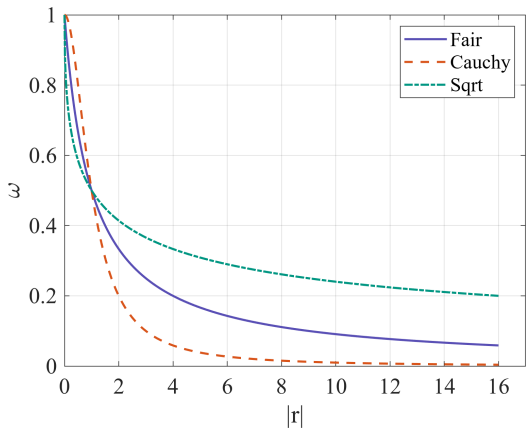


Figure S1. The relationship between $|r|$ and ω in main text Equations 1-3**.**

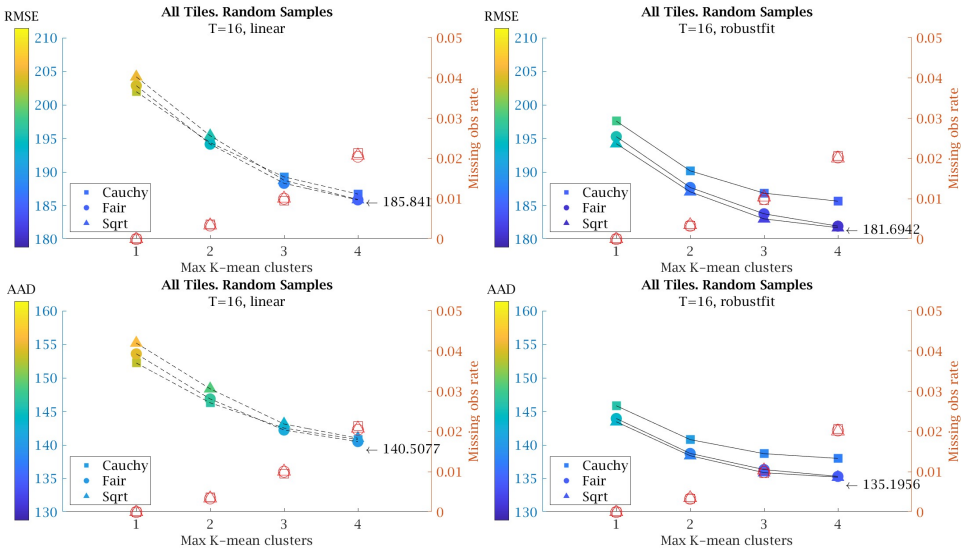


Figure S2. Calibration results of the TIF algorithm using the random sample, with Root Mean Square Error (RMSE) in the top row and Average Absolute Deviation (AAD) in the bottom row. We compared three weighting functions: Cauchy, Fair, and Sqrt across varying Max K-mean clusters (from 1 to 4) and two regression methods: ordinary least square (OLS) regression and robust model fitting (robustfit). In each subplot, an arrow highlights the scenario with the lowest error, and the associated error value is displayed. Unfilled markers represent the missing observation rate due to insufficient matching pairs, with *k_max*=1 used as the benchmark for comparison.

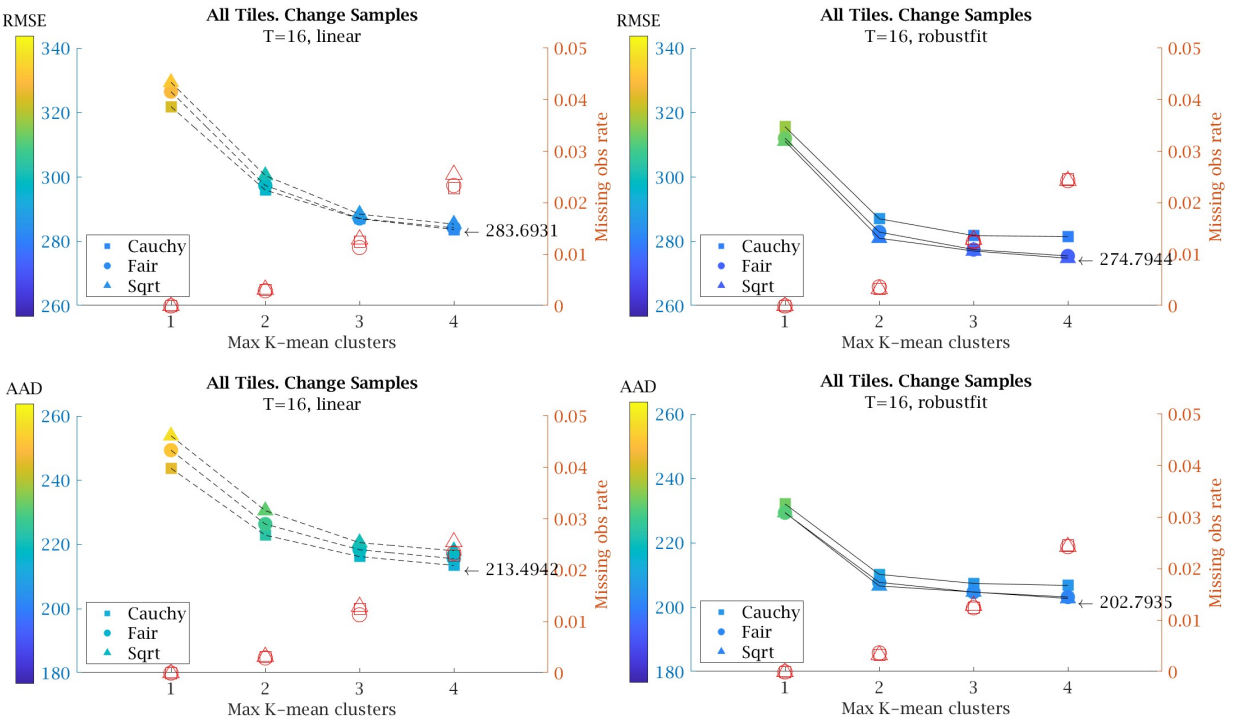


Figure S3. Calibration results of the TIF algorithm using the change sample with the same parameters as S1. In each subplot, an arrow highlights the scenario with the lowest error, and the associated error value is displayed. Unfilled markers represent the missing observation rate due to insufficient matching pairs, with *k_max*=1 used as the benchmark for comparison.

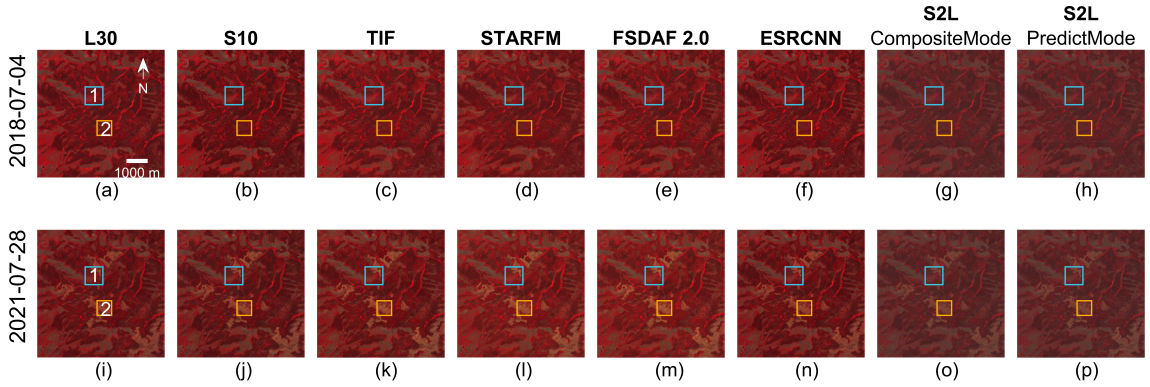


Figure S4. Visual comparison of the fusion results from the T13TCF experiment. The original 30 m Landsat images acquired on 2018-07-04 and 2021-07-28 are shown in (a) and (g). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (h). (c) – (f) are downscaled results on 2018-07-04 from TIF, ESRCNN, and S2L predict mode and S2L composite mode. The prediction results on 2021-07-28 are correspondingly (i) – (l). The centroid of the area is Lat = 41.3113386116, Lon = -107.174246506. Images are displayed by NNIR, red, and green bands with the same stretch scales. Two zoom-in areas are highlighted: subarea 1 – blue polygon; subarea 2 – orange polygon.

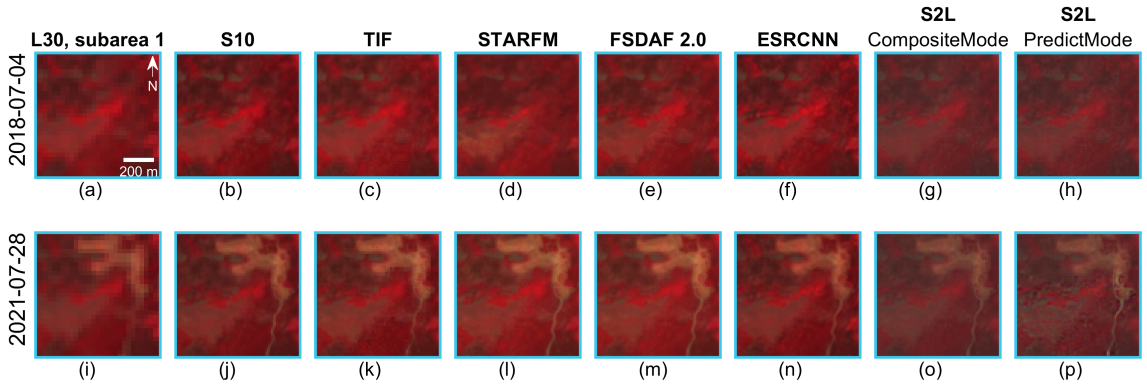


Figure S5. Zoom-in views of subarea 1 marked in **Figure S4**. The original 30 m Landsat images acquired on 2018-07-04 and 2021-07-28 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2018-07-04 from TIF, STARFM, FSDAF 2.0, ESRCNN, and S2L predict mode and S2L composite mode. (k) – (p) are corresponding downscaled images on 2021-07-28. Images are displayed by NNIR, red, and green bands with the same stretch scales.

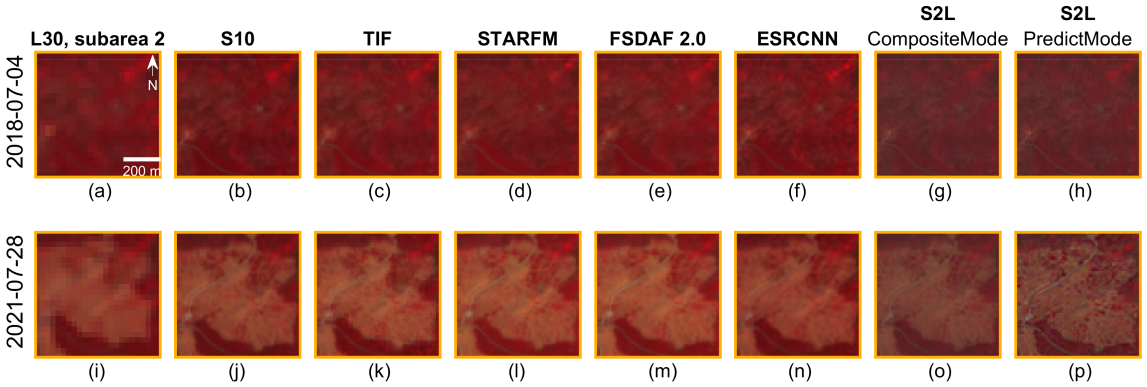


Figure S6. Zoom-in views of subarea 2 marked in **Figure S4**. The original 30 m Landsat images acquired on 2018-07-04 and 2021-07-28 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2018-07-04 from TIF, ESRCNN, and S2L predict mode and S2L composite mode. (k) – (p) are corresponding downscaled images on 2021-07-28. Images are displayed by NNIR, red, and green bands with the same stretch scales.

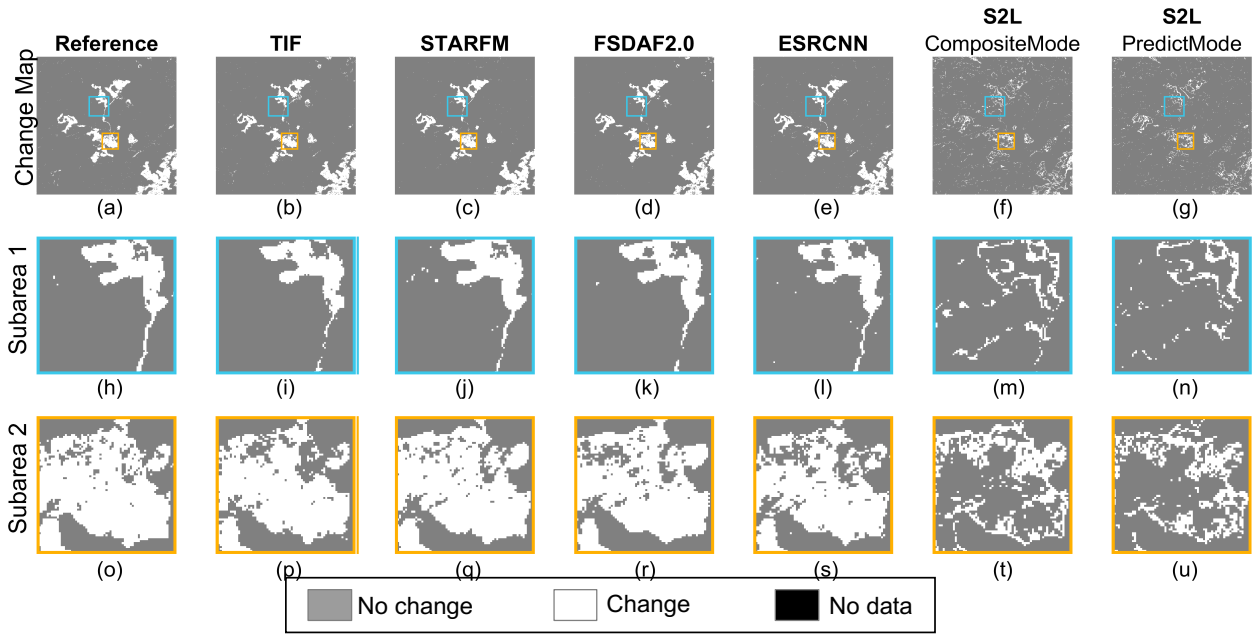


Figure S7. Change map comparisons with the same extent as **Figure S4** are shown in (a) – (g). The zoom-in views of subarea 1 enclosed in blue boxes with the same extent of **Figure S5** are shown in (h) – (n). The magnificent views of subarea 2 enclosed in orange boxes with the same extent of **Figure S6** are shown in (o) – (u). Grey means stable surfaces, white depicts changed areas, and black indicates invalid data from the QA band.

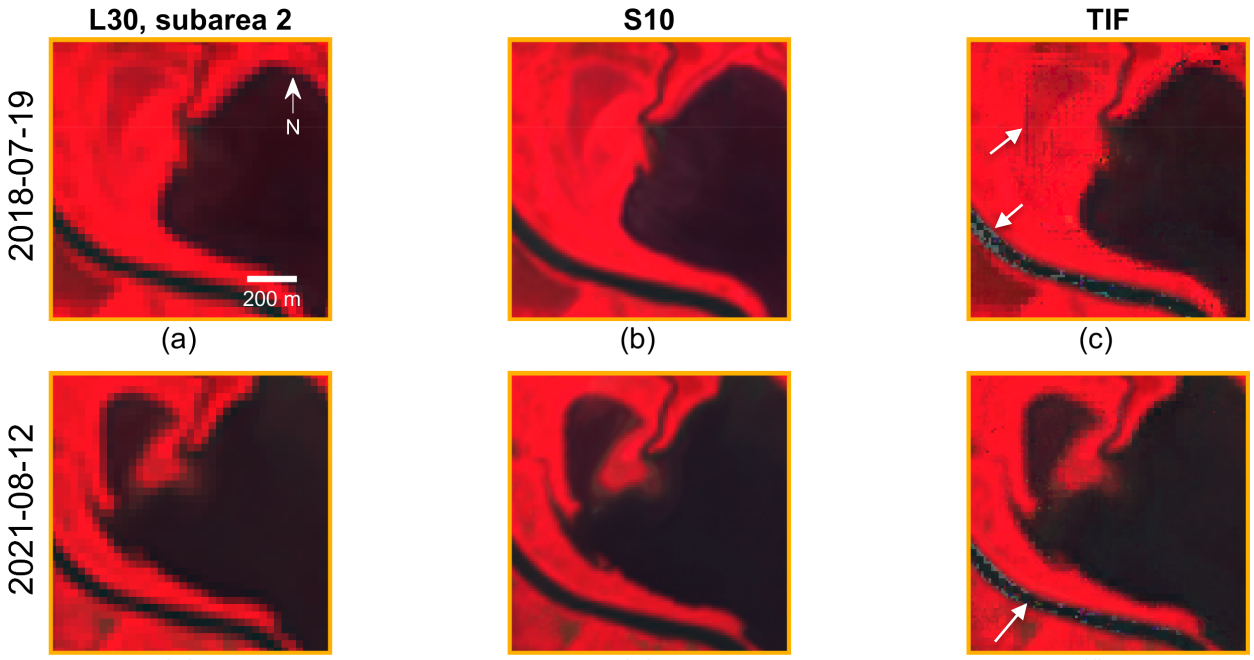


Figure S8. Visual comparison of the fusion results over a watery surface. The original 30 m Landsat images acquired on 2018-07-19 and 2021-08-12 are shown in (a) and (d). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (e). (c) and (f) are corresponding downscaled results on from TIF. Several artifacts are denoted with arrows.

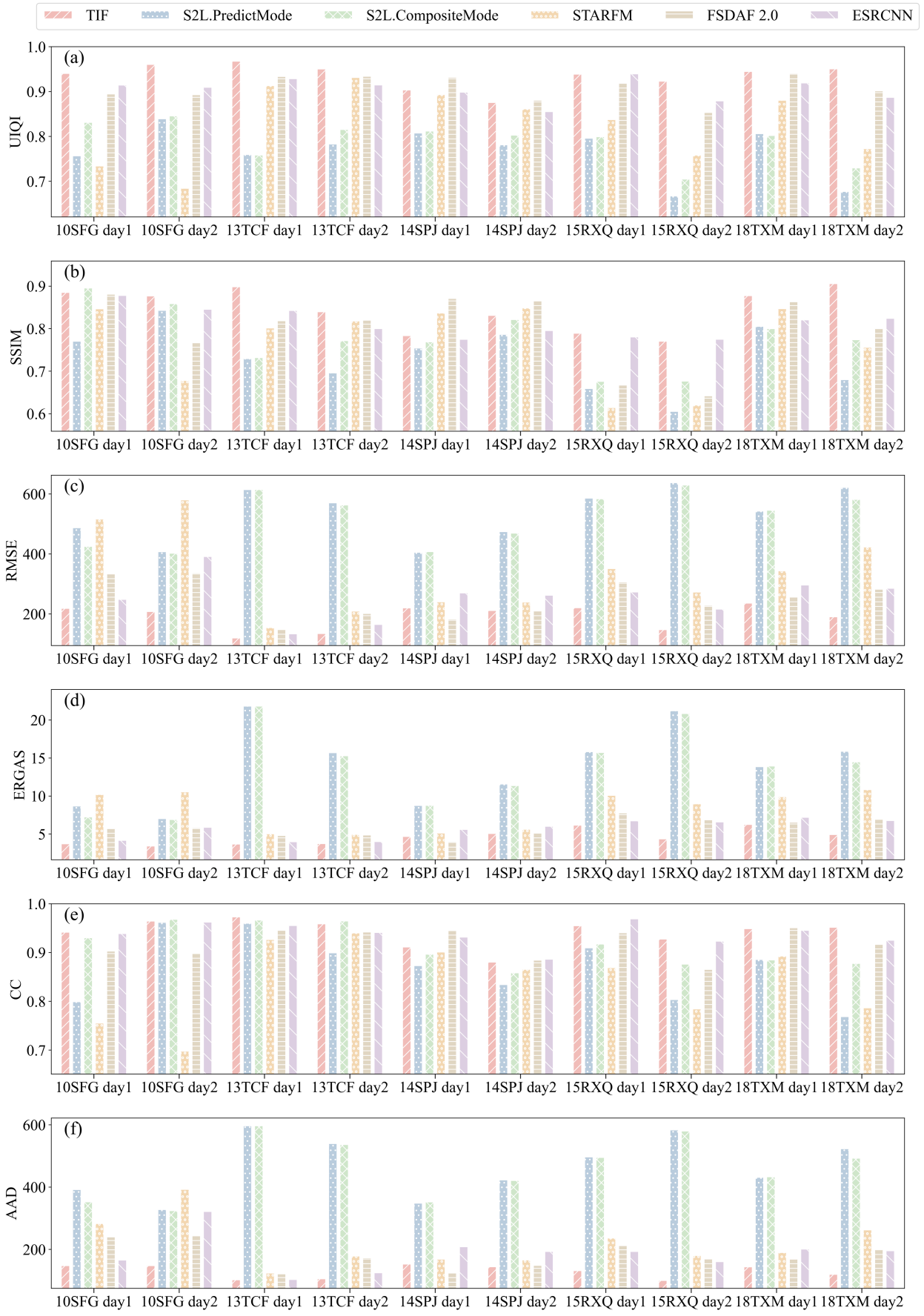

TIF
S2L.PredictMode
S2L.CompositeMode
STARFM
FSDAF 2.0
ESRCNN
(a)
UIQI
(b)
SSIM
(c)
RMSE
(d)
ERGAS
(e)
CC
(f)
AAD
10SFG day1
10SFG day2
13TCF day1
13TCF day2
14SPJ day1
14SPJ day2
15RXQ day1
15RXQ day2
18TXM day1
18TXM day2

Figure S9. Quantitative spectral and spatial evaluation for all study areas and dates using six metrics: (a) UIQI, (b) SSIM (c) RMSE, (d) ERGAS, (e) CC, (f) AAD. Values averaged over six spectral bands. Colors represent different algorithms: Orange for TIF, Yellow for STARFM, Brown for FSDAF 2.0, Blue for S2L.PredictMode, Green for S2L.CompositeMode, and Purple for ESRCNN. Note: The values of RMSE and AAD are enlarged by 10^4. *UIQI: Universal Image Quality Index. SSIM: Structure SIMilarity. RMSE: Root Mean Square Error; ERGAS: Erreur Relative Global Adimensionnelle de Synthese. CC: Correlation Coefficient; AAD: Absolute Average Difference.*

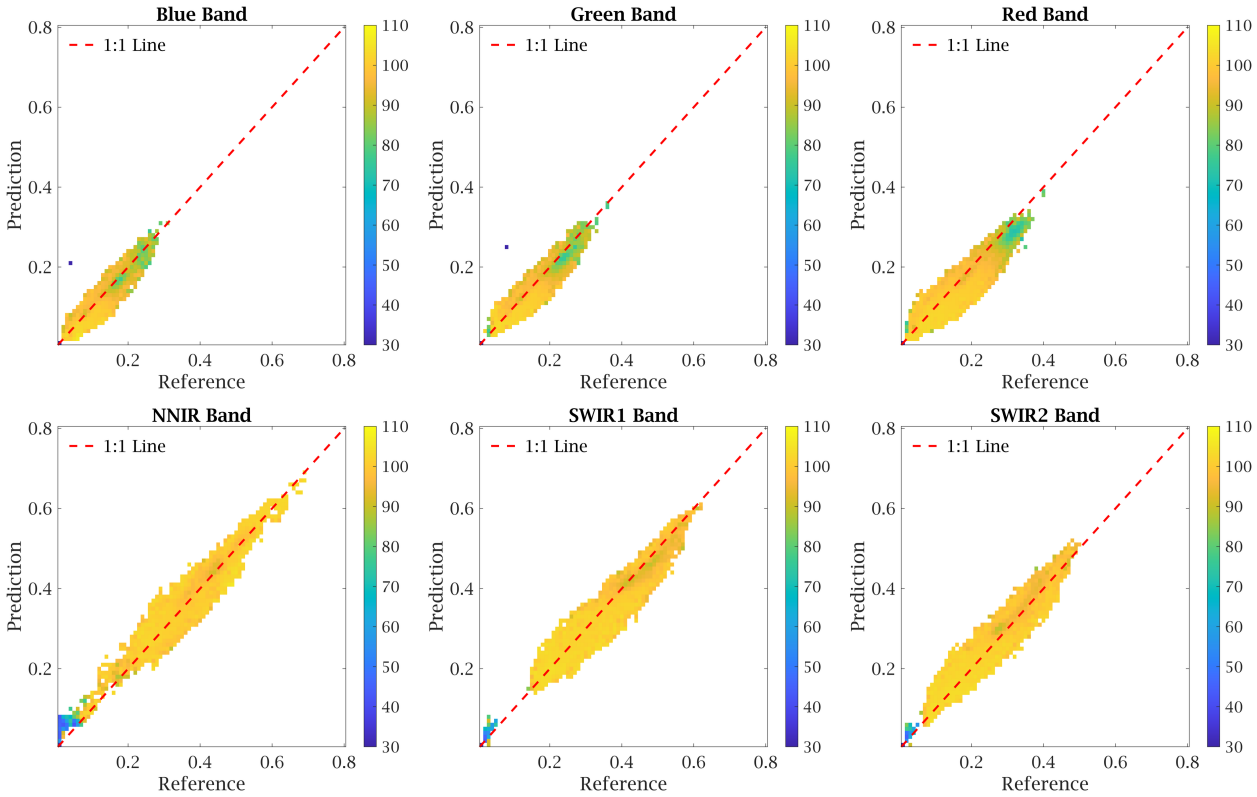


Figure S10. Accuracy of TIF predictions (T10SFG, 2021-07-20) versus number of observation pairs used for regression. Values are binned in 0.01 × 0.01 boxes. The colors represent the average number of observation pairs within each box.

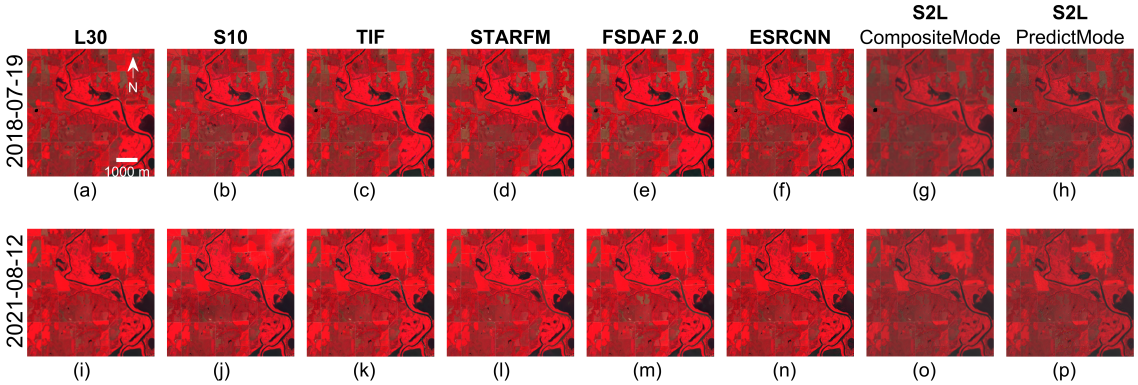


Figure S11. Visual comparison of the fusion results from the T14SPJ experiment. The original 30 m Landsat images acquired on 2018-07-19 and 2021-08-12 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2018-07-19 from TIF, ESRCNN, and S2L predict mode and S2L composite mode. The prediction results on 2021-08-12 are correspondingly (k) – (p). The centroid of the area is Lat = 39.2610397539, Lon = -97.0272638219. Images are displayed by NNIR, red, and green bands with the same stretch scales.

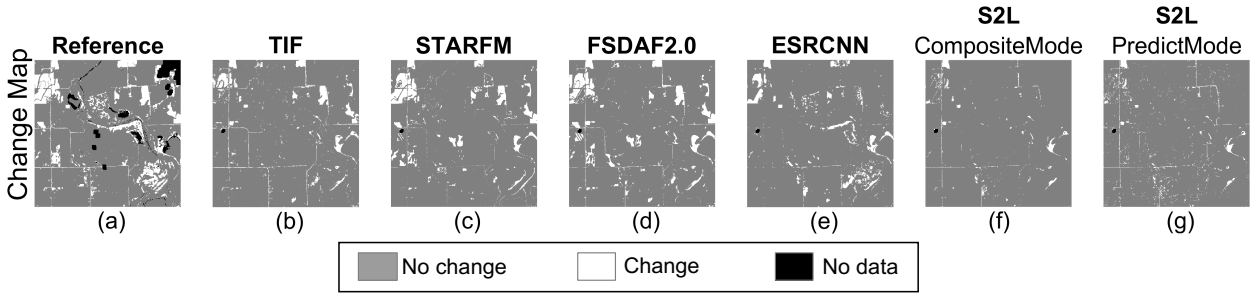


Figure S12. Change map comparisons to the same extent as **Figure S11**.

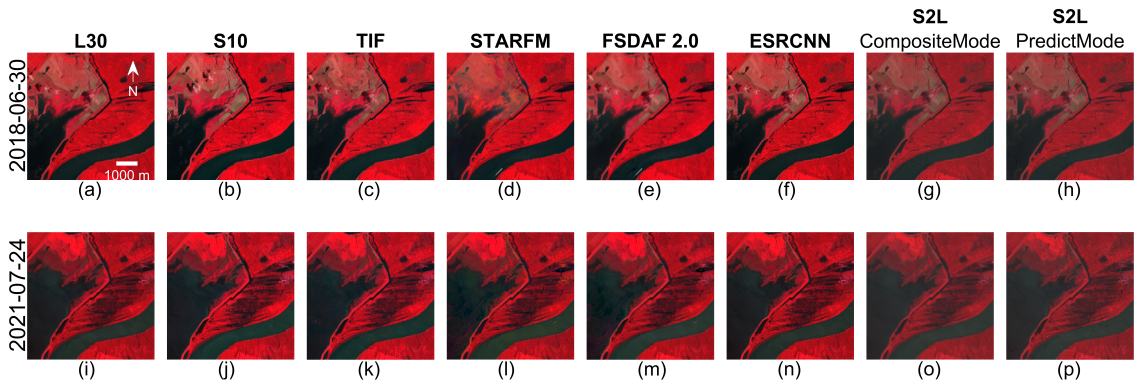


Figure S13. Visual comparison of the fusion results from the T15RXQ experiment. The original 30 m Landsat images acquired on 2018-06-30 and 2021-07-24 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2018-06-30 from TIF, ESRCNN, and S2L predict mode and S2L composite mode. The prediction results on 2021-07-24 are correspondingly (k) – (p). The centroid of the area is Lat =

30.7918687148, Lon = -91.5785579367. Images are displayed by NNIR, red, and green bands with the same stretch scales.

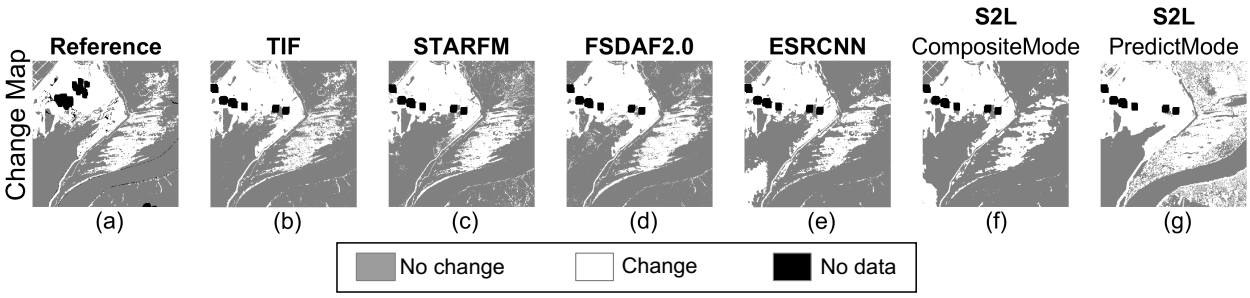


Figure S14. Change map comparisons with the same extent as **Figure S13**.

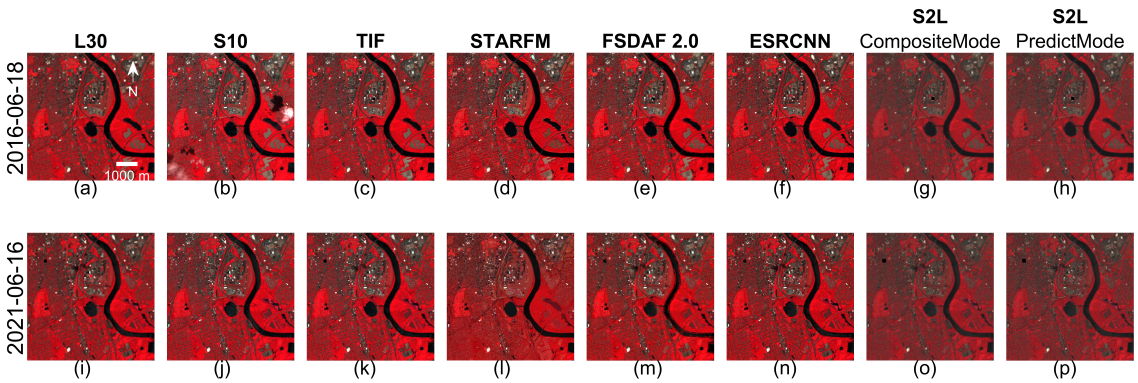


Figure S15. Visual comparison of the fusion results from the T18TXM experiment. The original 30 m Landsat images acquired on 2016-06-18 and 2021-06-16 are shown in (a) and (i). Reference 10 m Sentinel-2 (S10) acquired on the same dates are (b) and (j). (c) – (h) are downscaled results on 2016-06-18 from TIF, ESRCNN, and S2L predict mode and S2L composite mode. The prediction results on 2021-06-16 are correspondingly (k) – (p). The centroid of the area is Lat = 41.7295654684, Lon = -72.657813287. Images are displayed by NNIR, red, and green bands with the same stretch scales.

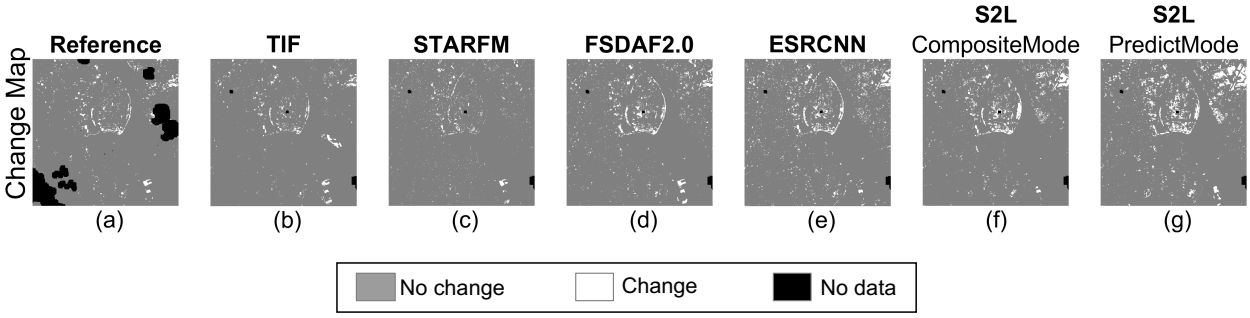


Figure S16. Change map comparisons with the same extent as **Figure S15**.

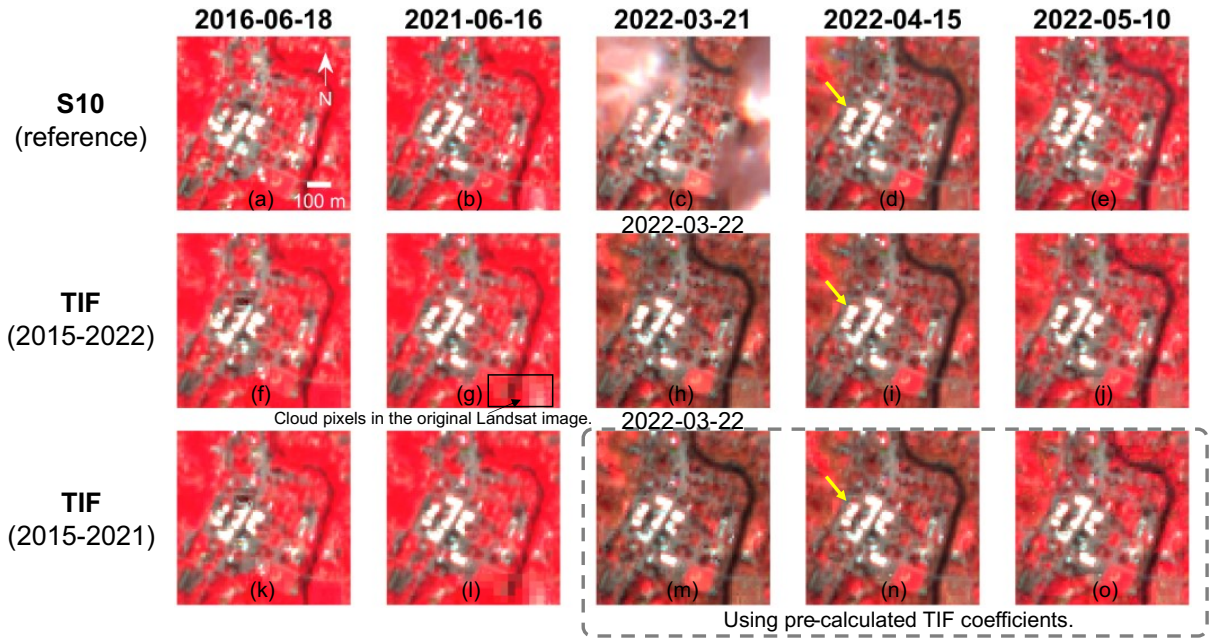


Figure S17. First row: image time series of 10 m Sentinel-2 (reference) with acquisition dates indicated at the top. Second row: TIF predictions using coefficients derived from the full period (2015-2022). Bottom row: TIF predictions using coefficients from 2015-2021. The study area is located in Great Barrington, MA, with a centroid at Lat = 42.19479148, Lon = -73.36087393. The yellow arrow denotes the rare cases where TIF (2015-2021) failed to predict the bright-up signal caused by heterogenous land changes, but TIF (2015-2022) successfully captured them. For pixel-level details and additional illustrations, refer to **Figure S20**.

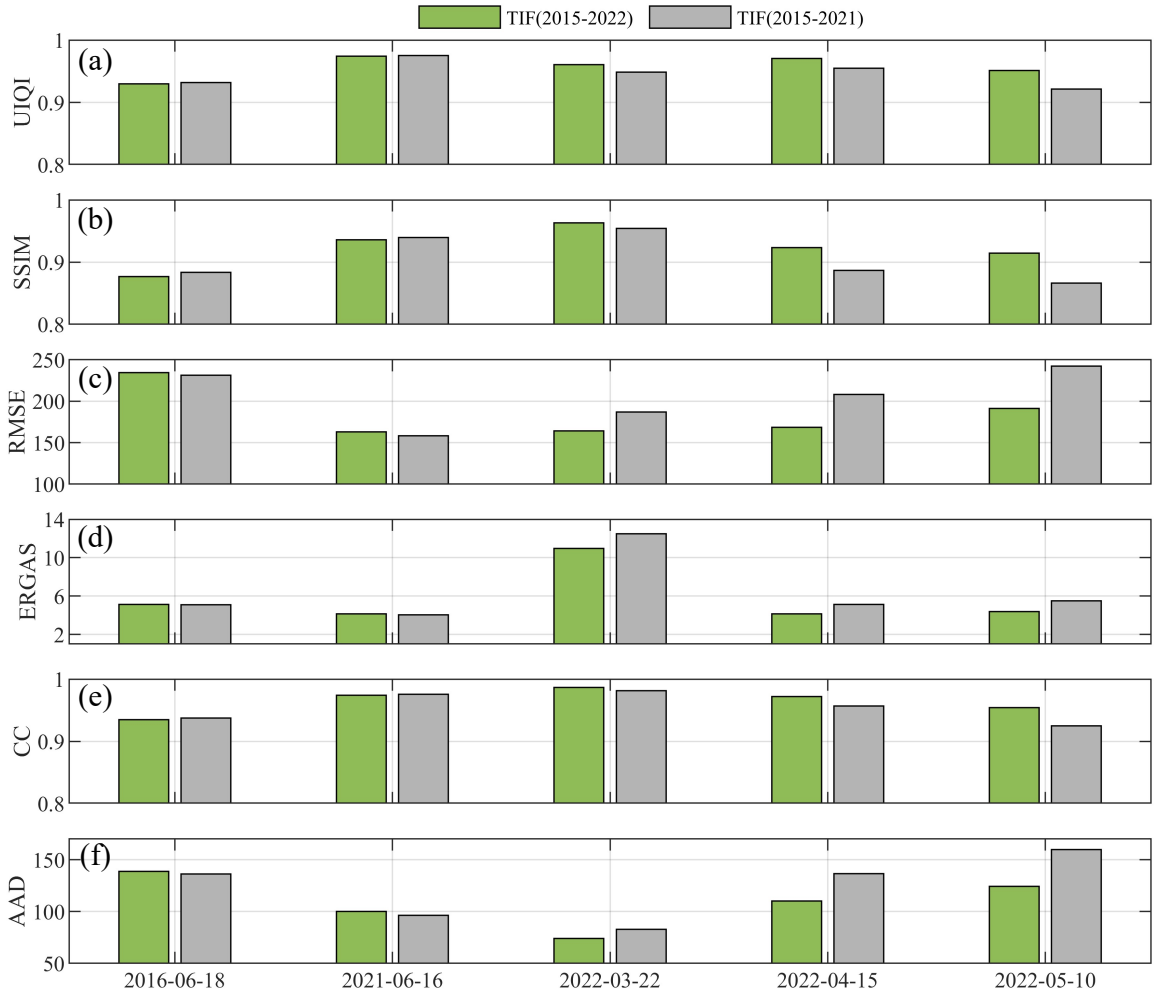


Figure S18. Comparison of quantitative assessment metrics for TIF(2015–2022; green bars) and TIF(2015–2021; grey bars) predictions shown in **Figure S17** for six metrics: (a) Universal Image Quality Index (UIQI), (b) Structural Similarity Index (SSIM), (c) Root Mean Squared Error (RMSE), (d) Erreur Relative Global Adimensionnelle de Synthese (ERGAS), (e) Correlation Coefficient (CC), and (f) Average Absolute Deviation (AAD) for selected dates. Note: The values of RMSE and AAD are enlarged by 10^4.

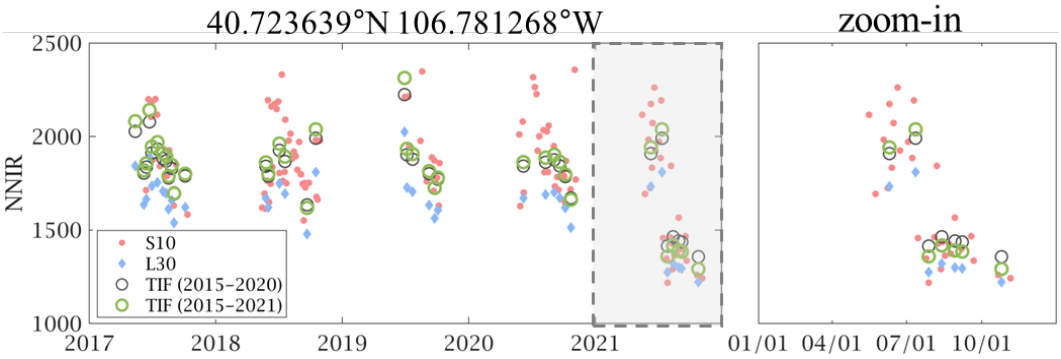


Figure S19. Harmonized Landsat and Sentinel-2 time series (10 m resolution) in NNIR band a forest pixel. Wildfire happened at this pixel in July 2021. The time series includes data from Sentinel-2 (S10, red dots), Landsat (L30, blue diamonds), and harmonized TIF datasets using coefficients derived from different time periods (2015–2020, black circles; 2015–2021, green circles). The zoom-in view of time series from 2021-01-01 to 2021-12-31 is shown on the right.

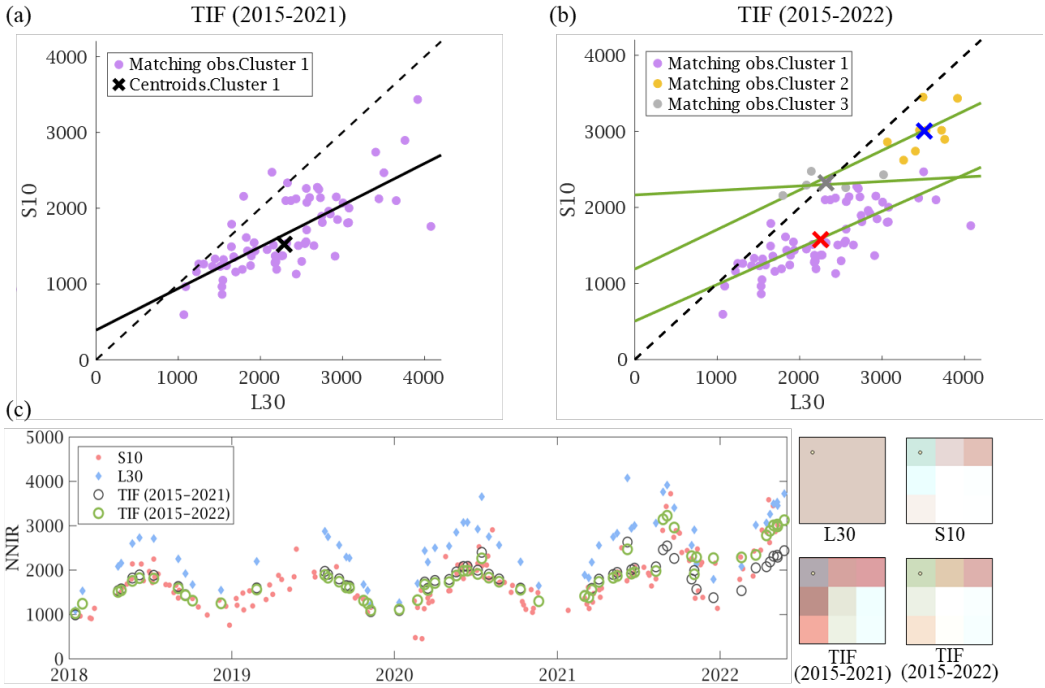


Figure S20. Illustration of TIF execution with a heterogeneous urban pixel (73.3632927°W 42.1953693°N) impacted by rooftop renovations in August 2021. (a) L30-S10 relationship derived from observation pairs spanning 2015-2021 for the NNIR band. (b) Same as (a), but clustered relationships were built using observation pairs from 2015-May 2022. (c) Left: time series of harmonized Landsat and Sentinel-2 observations (10 m resolution) for the target pixel. Right: False-color composite (NNIR/Red/Green) showing the single 30 m Landsat pixel (L30), the 3 by 3 grid of 10 m Sentinel-2 pixels (S10), TIF predictions using coefficients derived from (a) and (b) on 2022-05-10. The target 10 m pixel, located in the upper left corner of the 30 m bounding box, is highlighted with a white dot.

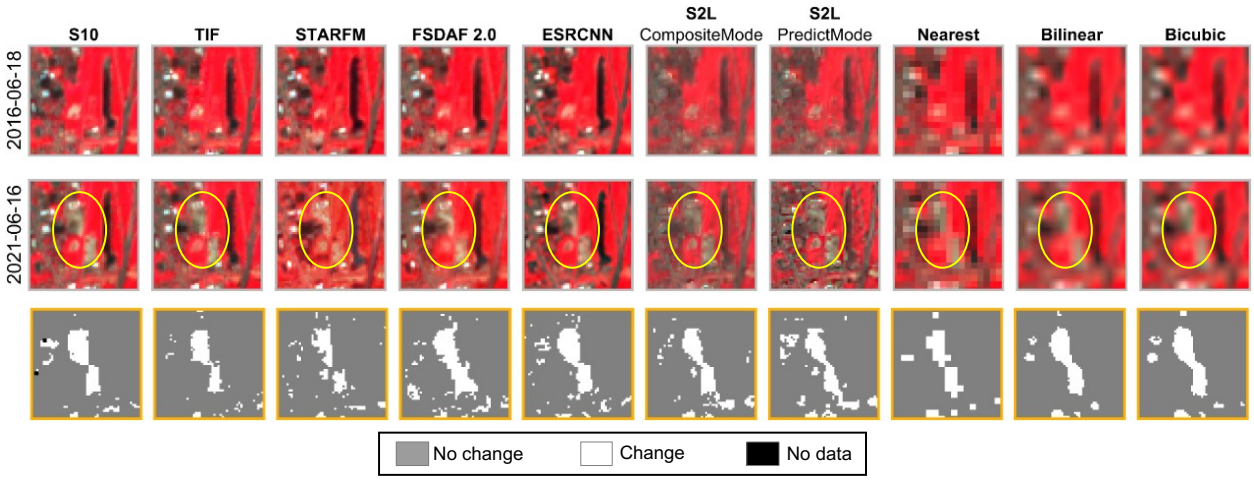


Figure S21. Change map comparisons. Grey means stable surfaces, white depicts changed areas, and black indicates invalid data from the QA band. The land cover conversion (forest to bare land) is indicated by the yellow circle.