https://doi.org/10.1016/j.agrformet.2020.108255

Evaluation of four image fusion NDVI products against in-situ spectral-measurements over a heterogeneous rice paddy landscape

Juwon Kong[a], Youngryel Ryu[a,b,c], Yan Huang[b], Benjamin Dechant[c], Rasmus Houborg[d], Kaiyu Guan[e,f], Xiaolin Zhu[g]

a Interdisciplinary Program in Landscape Architecture, Seoul National University, South Korea

b Department of Landscape Architecture and Rural Systems Engineering, Seoul National University, South Korea

c Research Institute of Agriculture and Life Sciences, Seoul National University, South Korea

d Planet Labs, San Francisco, CA, USA

e Department of Natural Resources and Environmental Sciences and National Center for Supercomputing Applications, University of Illinois at Urbana Champaign, Urbana, IL, USA

f Aspiring Universe Corporation, Champaign, IL, USA

g Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong

Submitted to *Agricultural and Forest Meteorology*

Abstract

Satellite image fusion methods that improve spatial and temporal resolution have significant potential to advance understanding of ecosystem dynamics in space and time. However, systematic evaluations of image fusion methods against *in situ* spectral data are lacking. Here, we used a suite of *in situ* spectral data collected at 60 elementary sampling units (10 × 10 m) covering 15 Landsat pixel (30 × 30 m) plots and one Moderate Resolution Imaging Spectroradiometer (MODIS) pixel (250 × 250 m) throughout the entire growing season in a heterogeneous rice paddy landscape to evaluate four state-of-the-art image fusion NDVI products. They include the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM), Flexible Spatiotemporal DAta Fusion (FSDAF), SaTellite dAta IntegRation (STAIR), and the CubeSat Enabled Spatio-Temporal Enhancement Method (CESTEM); the former three blended Landsat and MODIS data, whereas the latter combined CubeSats, Landsat, and MODIS observations. All fusion products showed strong linear relationships against *in situ* data when combining all spatial and temporal observations (R^2: 0.73 to 0.93) although there were partly negative biases (–1% to –9%). These biases resulted from forcing data to image fusion algorithms, such as Landsat (–4%) and MODIS (–7%). Performance difference between fusion methods were considerably larger for spatial than for temporal variation. Furthermore, Landsat NDVI explained only 17–22% of spatial variation against *in situ* spectral data, which can be translated into weak performance of image fusion products to predict spatial variability in NDVI. Image fusion products that relied on spatial interpolation showed large biases (–15% to –30%) for a vegetation plot surrounded by mixed land cover plots. Our results highlight key sources of uncertainty and will be instrumental in improving satellite image fusion methods to monitor land surface phenology in space and time.

1. Introduction

High-spatiotemporal-resolution remote sensing images have enhanced our ability to monitor ecosystem dynamics across different scales. However, a fundamental tradeoff between spatial resolution and temporal revisit frequency limits satellites' ability to observe the Earth's surface at high spatiotemporal resolution (Gao *et al*. 2006). To overcome this, previous studies developed satellite image fusion methods that enhanced spatiotemporal resolution by combining temporally sparse fine-spatial-resolution images with frequent but coarse-spatial-resolution images (Gao *et al*. 2006; Hilker *et al*. 2009a; Luo *et al*. 2018; Zhu *et al*. 2010; Zhu *et al*. 2016). Satellite image fusion products have been widely used to extract spatial and temporal information on changes in land use and land cover (Chen *et al*. 2015b; Schmidt *et al*. 2015; Senf *et al*. 2015), classification of vegetation types (Liu *et al*. 2015), vegetation phenology (Hilker *et al*. 2009b; Walker *et al*. 2014; Zheng *et al*. 2016), and crop growth situations at field scale (Gao *et al*. 2017a; Kimm *et al*. 2020). However, the question as to how much we can trust the quality of fusion products over heterogeneous landscapes remains.

The uncertainties associated with fusion products stem from the inconsistency of the input images and the temporal or spatial interpolation processes (Walker *et al*. 2012). Uncertainties in fusion products due to input images may be caused by the different spectral response functions of satellite sensors and sun-target-view geometries (Gao *et al*. 2014; Roy *et al*. 2016; Wang *et al*. 2014). Moreover, uncertainties surrounding spatial interpolation in fusion products may arise from assumptions that nearby pixels within the same land cover have similar spectral reflectance patterns (Zhu *et al*. 2018). Fusion products may average out drastic changes between individual pixels, leading to blurred images. Furthermore, uncertainties caused by the temporal interpolation process may emerge due to the time lag between the dates of the input pair and the dates of the predicted fusion products (hereafter, "time-lag effects"; Fu *et al*. 2015; Olexa and Lawrence 2014). This affects the performance of temporal interpolation over different land cover types (Emelyanova *et al*. 2013; Olexa and Lawrence 2014).

In situ spectral data over space and time allow us to directly quantify the uncertainties of fusion products as well as assess the uncertainty contribution of input imagery and characterize the time-lag effects. The quality of satellite image fusion products is typically assessed for a given date by comparing original fine spatial resolution satellite images with the fusion product developed using independent data (Emelyanova *et al*. 2013). However, systematically measured *in situ* data—which is obtained temporally close to the overpass time with accurate and precise geolocation—is essential for the comprehensive evaluation of image fusion products. To the best of our knowledge, few efforts to date have tried in this regard (Gao *et al*. 2017a).

Ground-based evaluation of fusion products is particularly important in heterogeneous landscapes.

In homogeneous landscapes, the development and evaluation of fusion products are relatively straightforward, as canopy structure and function change only gradually within a subpixel of coarse spatial resolution imagery. In heterogeneous landscapes, however, abrupt changes in the subpixel over space and time are a considerable challenge for image fusion methods (Fu *et al*. 2015). For example, in crop landscapes, various fields may be managed differently for transplanting and harvest dates, fertilization, or irrigation, which can lead to differences in canopy structure, physiology, and phenology (Ding *et al*. 2014). Only few studies have evaluated image fusion products over heterogeneous landscapes such as dryland (Walker *et al*. 2012) and cropland ecosystems (Gao *et al*. 2017b; Zheng *et al*. 2016) by comparing them to independent satellite images or ground-based visual phenology assessments. To the best of our knowledge, no study has evaluated image fusion products with *in situ* spectral observations based on a systematic spatio-temporal sampling design over heterogeneous landscapes.

Most fusion products rely on weight function–based methods that combine all input data by considering spectral differences in input data, time-lag effects, and the distance between the central pixel in the predicting area and neighboring pixels (Zhu *et al*. 2018). The spatial and temporal adaptive reflectance fusion model (STARFM) was the original weight function–based method (Gao *et al*. 2006). When each pixel of input data includes only one land cover, STARFM considers changes in reflectance to be consistent over time. After this initial study, weight function-based methods were subsequently improved in heterogeneous areas (Zhu *et al*. 2010), in capturing abrupt changes in land cover (Zhu *et al*. 2016). Moreover, (Luo *et al*. 2018) proposed cloud-free/gap-free daily step data. Recently, Houborg and McCabe (2018a) introduced a promising spatiotemporal fusion product by harmonizing the satellite images from CubeSat constellation, albeit with non–weight function–based methods.

The objective of this study was to evaluate the performance of three satellite-based spatiotemporal fusion products and a novel CubeSat constellation-based fusion product with *in situ* measurements over a heterogeneous rice paddy landscape. We evaluated the following four image fusion products: enhanced STARFM (ESTARFM; Zhu *et al*. 2010), Flexible Spatiotemporal DAta Fusion (FSDAF; Zhu *et al*. 2016), SaTellite dAta Integration (STAIR; Luo *et al*. 2018), and the CubeSat Enabled Spatio-Temporal Enhancement Method (CESTEM; Houborg and McCabe 2018a). Of these methods, only CESTEM does not include a spatial gap-filling module. We attempted to answer the following questions: (i) How do the input data and the fusion algorithm affect the accuracy of fusion products? (ii) Under which conditions does the spatial or temporal interpolation process affect the performance of fusion products? (iii) How does heterogeneity affect the performance of fusion products? We report the results of direct comparisons of each product with *in situ* measurements.

2. Methods

2.1 Study site

The study site was a rice paddy landscape in Cheorwon, Republic of Korea (CRK; 38.2013°N, 127.2507°E), which is part of the Korea Flux Network (KoFlux; Huang *et al*. 2018; Dechant *et al*. 2019; Dechant *et al.* 2019; Hwang *et al*. 2020; Figure 1a). As the site is flat and rice canopies have relatively low heights, it is suitable for *in situ* spectral measurements at locations determined by satellite pixel locations. The rice growing season lasts from approximately May to September, and the predominant species in the study area is *Oryza sativa* L. ssp. *Japonica*. The site has a temperate monsoon climate with frequent cloud cover and high precipitation from June to August. The size of the paddy fields ranges from around 2,500 to 4,300 m^2. Cultivation management includes irrigation, fertilization, drainage, and harvesting and varies between rice paddies (Figure 1b).

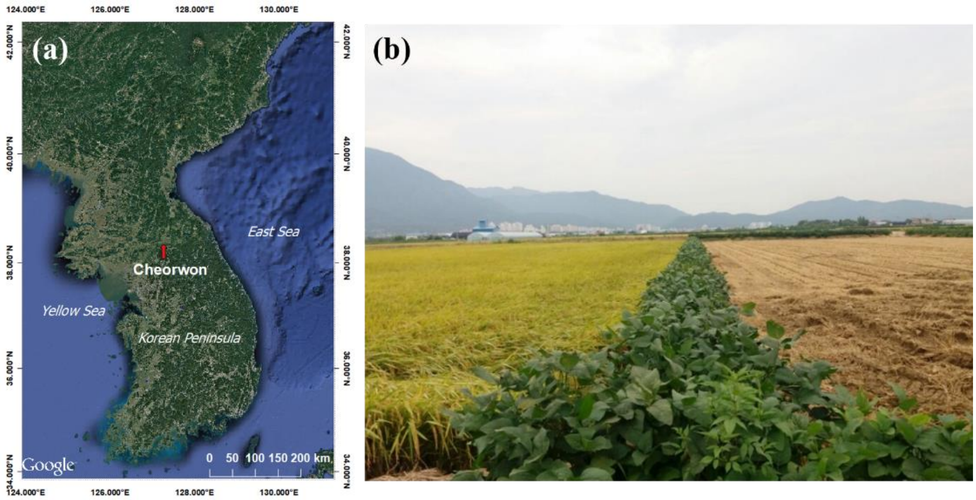


Figure 1 Study site. (a) Map of the Korean Peninsula (image source: Google Earth) with the study site, Cheorwon, marked with a red arrow. (b) This image of the study site was acquired on September 5, 2017, and shows rice paddy fields with different harvest dates bordered by soybean plants.

2.2 In-situ measurements

To conduct a systematic sampling that represented the study landscape, we relied on the elementary sampling unit (ESU), which has been used to evaluate moderate spatial resolution satellite images (Morisette *et al*. 2006). To minimize Global Positioning System (GPS) positioning error, we designed a 10 × 10 m pixel as an ESU to represent a pixel of an independent satellite image (i.e., Sentinel-2; Appendix 1). To evaluate the fusion products and their input satellite products (i.e., Landsat 8; Moderate Resolution Imaging Spectroradiometer [MODIS]), we designed the resolution of the ESUs to represent a pixel of each satellite imagery type: a 30 × 30 m Landsat 8 pixel by 4 ESUs and a 250 × 250 m MODIS pixel by 60 ESUs (Figure 2; Appendix 1).

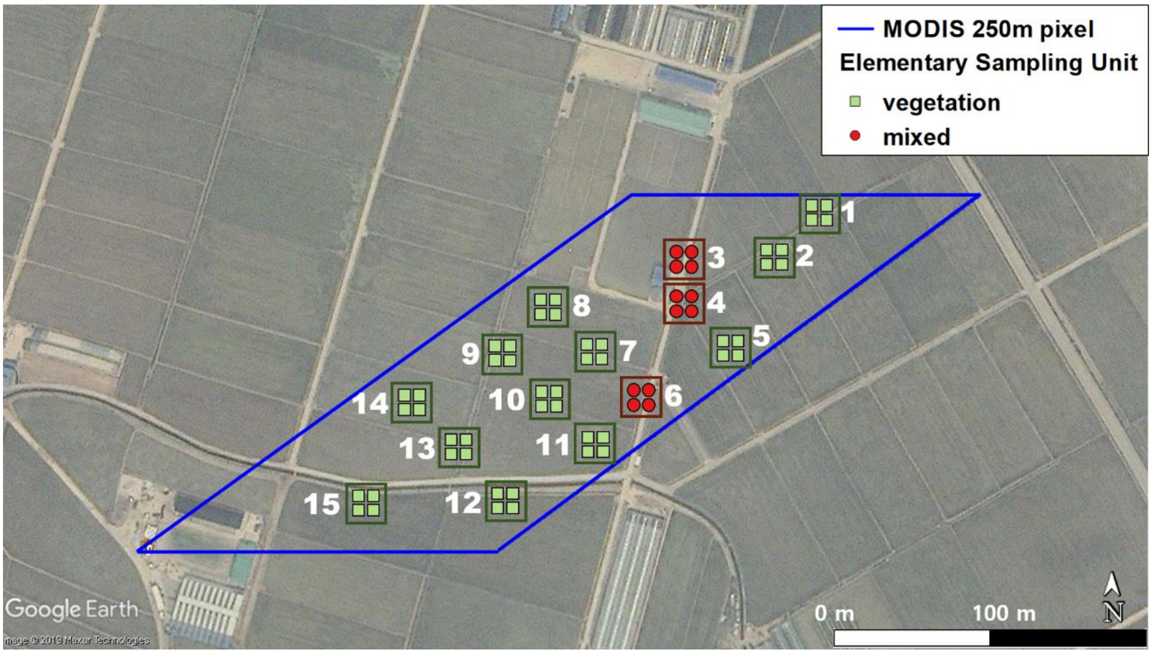


Figure 2 Schematic overview of the sampling design. The elementary sampling units (ESUs; 10 × 10 m) are represented by light green squares for vegetation land class and red circles for mixed land class. Correspondingly colored squares around each ESU indicate each plot (Background image source: Google).

We collected *in situ* measurements using a hyperspectral spectroradiometer (Jaz, Ocean Optics, Dunedin, FL, USA) equipped with fiber optics and a cosine corrector (CC-3-UV-T, Ocean Optics) to measure bi-hemispheric reflectance. We located the cosine corrector at 3.5 m above ground using a bar to cover 90% of the upwelling hemispherical irradiance in an ESU (Liu *et al*. 2017). To minimize the uncertainties of *in situ* measurements from changing sky conditions, we collected 6 to 10 spectra from the same position of each ESU, removed outliers that showed unstable values, and took the average of the remaining spectra. Considering the satellite overpass times of around 10 a.m. to 12:30 p.m. (Sentinel-2, Landsat 8, and MODIS in this study), we conducted all ground measurements between 9 a.m. and 1 p.m. local time (Coordinated Universal Time; UTC +9). To ensure the accurate and precise location of the samples, we recorded the GPS coordinates of each point using a commercial device that included a 0.5 m resolution base map (MONTANA 650TK, Garmin, Switzerland). Moreover, we cross-checked the spectral quality of the *in situ* Jaz spectrometer data with an ASD spectrometer (FieldSpec 4 Wide-ResField Spectroradiometer, ASD, Boulder, CO, USA) and found good agreement (Appendix 2). The dates on which we performed the *in situ* measurements are detailed in Figure 3.

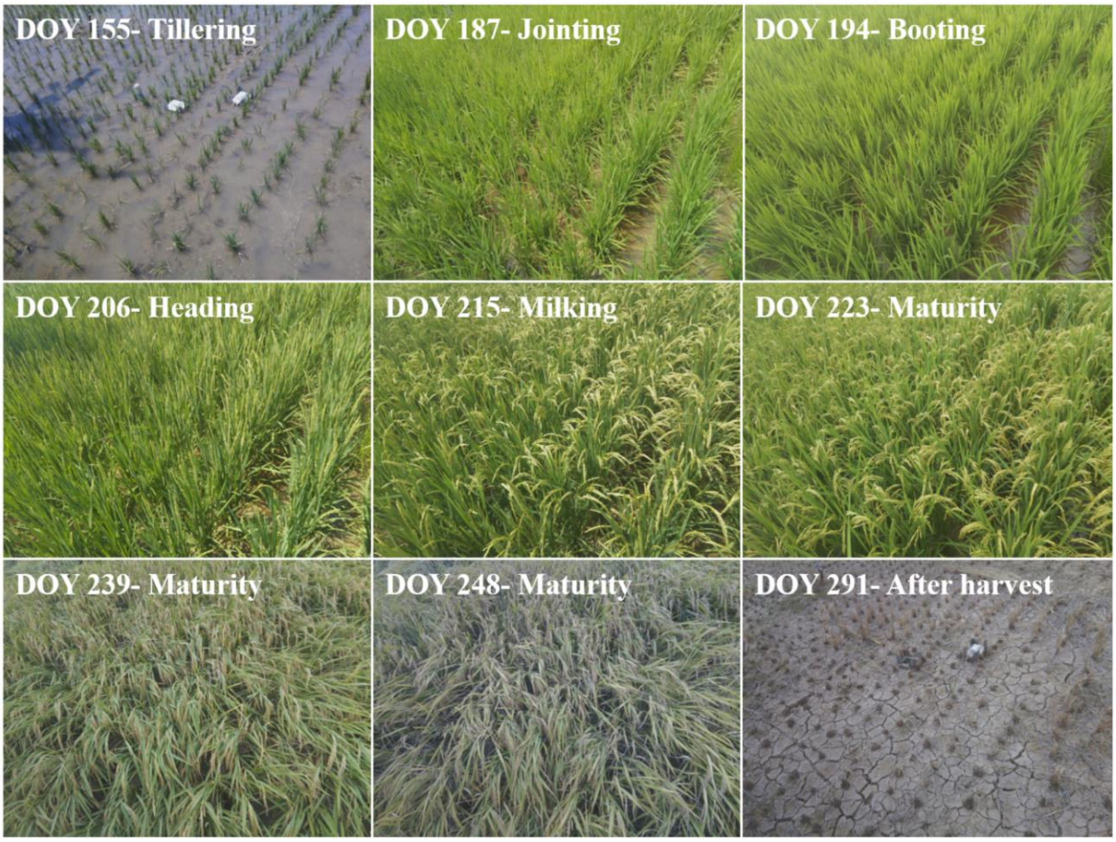


Figure 3 Images of the paddy in which the flux tower is located. Each image was labeled with day-of-year (DOY) *in situ* measurements and the phenological stage of the rice.

At the peak of the growing season, we classified the study landscape within a MODIS pixel into two land cover types: vegetation and non-vegetation (e.g., road, building, ditch) with 76% and 24% coverage, respectively. We did this by applying K-means classification to high-spatial-resolution CubeSat images (i.e., PlanetScope Ortho Tile product, 3 m spatial resolution). Of the 15 plots, Plots 3, 4, and 6 were located on mixed land cover that included vegetation, roads, buildings, and ditches (hereafter, “mixed” land class; Figure 2). Other plots were located on vegetation cover (hereafter, “vegetation” land class; Figure 2). In the vegetation land class, Plot 5 was harvested 2 weeks earlier than the other plots.

2.3 Image fusion products

We chose four state-of-the-art image fusion products that have been widely used or recently developed (Table 1; Figure 4; Supplementary 1). We defined spatial interpolation in this study as enhancing the spatial resolution of a coarse-resolution image (e.g., MODIS) on the predicted date using each algorithm with paired input data (Luo *et al*. 2018; Zhu *et al*. 2010; Zhu *et al*. 2016).

ESTARFM (Zhu *et al*. 2010) is a weight function–based method that requires two or more pairs of images. Each pair of images consists of fine-spatial-resolution images (e.g., Landsat) and coarse-spatial-resolution images (e.g., MODIS) acquired on the same date. Based on the pairs, ESTARFM predicts the fusion product using coarse-resolution images on the desired dates. It first searches for spectrally similar pixels from fine-resolution images and then weights them according to geographic distance from the central pixel in a moving window and the purity of their corresponding MODIS pixels. It then computes the coefficients of linear regression between the two pairs to convert coarse-resolution images to fine-resolution images. In a heterogeneous landscape, the reflectance changes in a coarse-resolution pixel across time do not equal the changes in the fine-resolution pixels within it. Thus, by introducing a conversion coefficient for each pixel, ESTARFM has advantages in heterogeneous areas (Zhu *et al*. 2010).

FSDAF (Zhu *et al*. 2016) is a hybrid method that requires minimum input data of one pair of fine- and coarse-spatial-resolution images (i.e., base images) on the same day and coarse-spatial-resolution images acquired on the prediction date. FSDAF uses unsupervised classification for the fine-resolution images and estimates the temporal change in each class at fine resolution from the coarse-resolution images. From the estimated temporal changes from the coarse images, FSDAF predicts the fine-resolution pixel by adding the temporal changes to the base fine-resolution pixel, i.e., temporal prediction. Then, FSDAF uses thin-plate spline interpolation to distribute residuals of temporal prediction for capturing abrupt land cover changes. An advantage of FSDAF is that it can capture both gradual and abrupt changes in land cover (Zhu *et al*. 2016).

STAIR (Luo *et al*. 2018) is a weight function–based method that uses all available Landsat and MODIS images. STAIR applies segmentation to homogeneous pixels to identify missing-value pixels, such as cloud pixels or data gaps caused by failures in the Landsat 7 Scan Line Corrector. The imputation of missing-value pixels in each segment is based on an adaptive-average correction, which assumes that changes in pixel values are approximately identical between neighborhood pixels within a short time frame (e.g., <2–3 weeks). Finally, STAIR fuses a daily MODIS time series with imputed Landsat-MODIS pairs to produce a daily time series of predicted surface reflectance images. A time-series based new cloud masking algorithm was applied for both MODIS and Landsat data to achieve higher performance in identifying cloud masked compared with the existing MODIS/Landsat cloud mask products. STAIR's strength is its ability to generate daily time series of fine-spatial-resolution products by systematically integrating Landsat-MODIS image pairs for missing-pixel imputation and automatically determining the input pair of fine and coarse images for the target date (Jiang *et al*. 2020; Kimm *et al*. 2020).

CESTEM (Houborg and McCabe 2018a) is a machine-learning method that leverages rigorously calibrated 'gold standard' satellites (e.g., MODIS, Landsat) in concert with lower quality but superior resolution CubeSats (e.g., PlanetScope products; Planet Labs, San Francisco, CA, USA) to produce a radiometrically harmonized and temporally consistent surface reflectance product. To ascertain associations between training data of input explanatory variables (CubeSat spectral data) and the target variable (Landsat 8 surface reflectance data), CESTEM uses a cubist rule-based regression technique that is a nonparametric machine-learning approach belonging to the family of regression tree methods. CESTEM harmonizes multi-sensor CubeSat images against Landsat and MODIS images for consistent radiometric correction. Its key strength is its ability to ingest data from multiple sensors with differing radiometric and spectral responses and generate images that are spectrally consistent with Landsat 8 surface reflectance data while inheriting the high spatial resolution of PlanetScope data (3 m). In this study, we used CESTEM version 1.0, which does not include a spatial and temporal gap-filling process.

Table 1 The image fusion products used in this study. '√' and 'N/A' indicate whether the specific process was included or not, respectively.

Fusion product	Type	Target resolution	Spatial interpolation	Temporal gap filling	Radiometric correction	Input data (revisit frequency)
ESTARFM	Weight function–based method	30 m	√	N/A	N/A	MODIS (MOD09GQ; daily), Landsat (Landsat 8 OLI C1 Level 2; 16 days)
FSDAF	Hybrid method (weight function-based and unmixing method)	30 m	√	N/A	N/A	MODIS (MOD09GQ; daily), Landsat (Landsat 8 OLI C1 Level 2; 16 days)
STAIR	Weight function–based method	30 m	√	√	N/A	MODIS (MCD43A4; daily), Landsat (Landsat 7 and 8 Level 2; 16 days)
CESTEM	Machine-learning method	3 m	N/A	N/A	√	MODIS (MCD43A4; daily), Landsat (Landsat 8 OLI products Level 1T-6SV; 16 days), Planet (Planet Scope Ortho Tile; daily)

For comparisons across products, all fusion products were resampled to 30 m spatial resolution with the map projection WGS 84, UTM zone 52N. To reduce the potential errors from resampling, we tested three resampling methods (i.e. nearest-neighbor, bilinear, bicubic) with and without antialiasing for CESTEM. Among the methods, we chose the nearest-neighbor interpolation method without antialiasing as it showed the best performance to in-situ measurements (Supplementary 2).

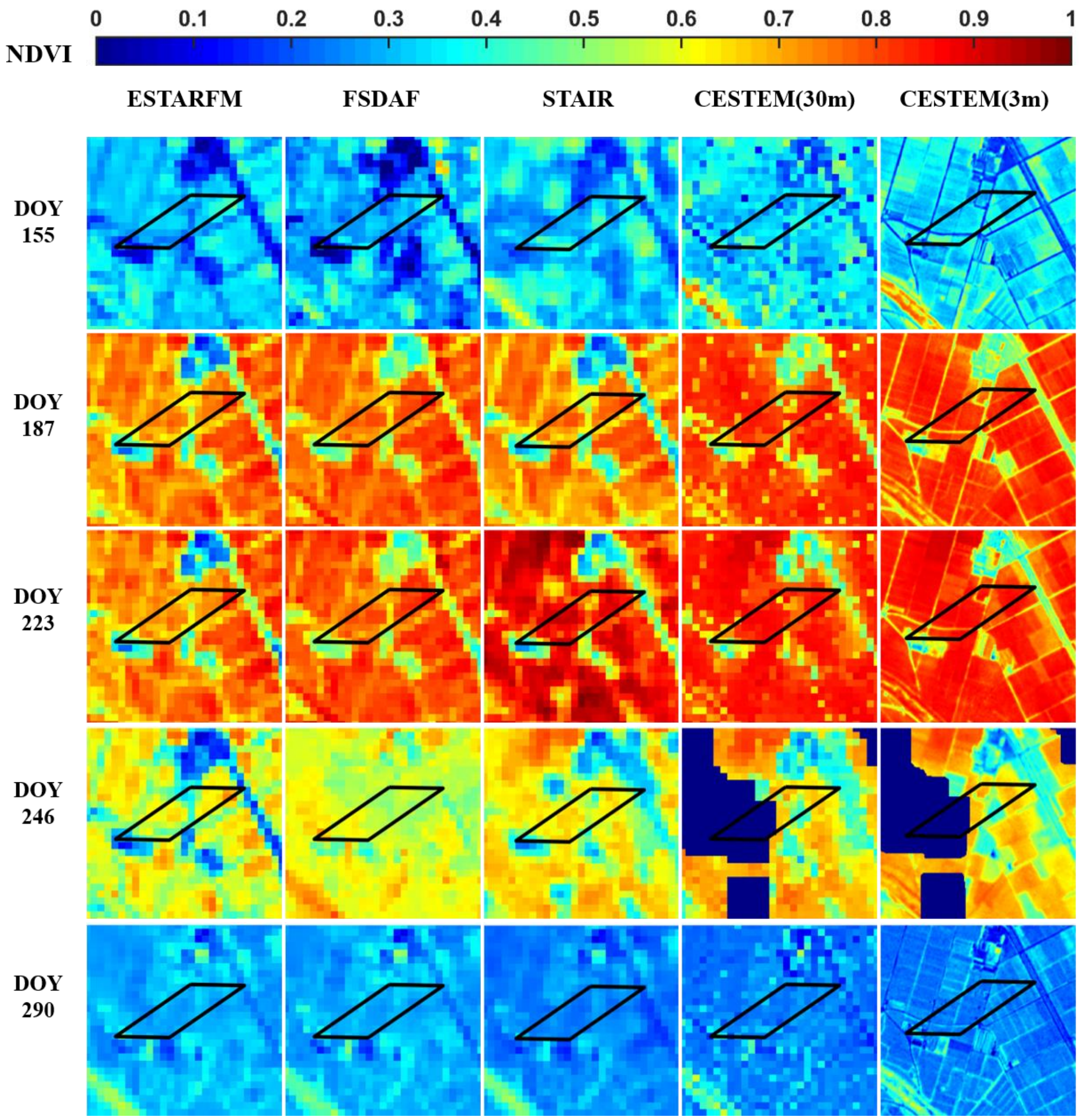


Figure 4 NDVI maps (870 × 870 m) of fusion products on *in situ* measurement dates, including when all fusion products were available on the same date. The study site is located within the black polygon that indicates a MODIS 250 m pixel. The dark navy color on CESTEM DOY 246 indicates masking due to clouds. All image fusion products were 30 m resolution except for CESTEM (3 m), which was aggregated to 30 m resolution for consistency.

2.4 Input data for image fusion products

Satellite data with different spatial and temporal resolutions were used to generate image fusion products (Table 1). The ESTARFM and FSDAF products used MODIS surface reflectance daily products (C6 MOD09GQ) and Landsat 8 Operational Land Imager (OLI; Landsat 8 OLI/TIRS C1 Level 2). The STAIR product used Landsat reflectance products from all platforms (Landsat 7 Enhanced Thematic Mapper and Landsat 8 OLI) and MODIS (C6 MCD43A4 product). The CESTEM product used the Planet Scope Ortho Tile product (Planet Team 2018), Landsat 8 OLI products Level 1T corrected to surface reflectance by 6SV, and nadir bidirectional reflectance distribution function (BRDF)-adjusted MODIS daily surface reflectance.

2.5 Evaluation

To evaluate fusion products with *in situ* measurements, we used NDVI (Rouse 1974; Tucker 1979), which is widely used as a measure of vegetation greenness as plants appear relatively dark in red but bright in the near-infrared (NIR). NDVI is a band ratio formulation (Eq. (1)) that is less sensitive to sensor-target-sun geometry than the surface reflectance of individual red and near-infrared channels (Feng *et al*. 2002; Ryu *et al*. 2010). Moreover, the spectral discrepancies between MODIS and Landsat did not cause significant differences in NDVI products (Zhou *et al.* 2021). Our *in situ* observation setting measured bi-hemispheric reflectance (i.e., blue-sky albedo), whereas satellite data provided the bidirectional reflectance factor (BRF). We confirmed that the discrepancy in view geometry between *in situ* and satellite images led to negligible differences in surface reflectance and NDVI (Appendix 3), which were within ranges reported by Czapla-Myers *et al*. (2015).

$$\boldsymbol{NDVI} = \frac{\boldsymbol{\rho NIR} - \boldsymbol{\rho Red}}{\boldsymbol{\rho NIR} + \boldsymbol{\rho Red}} \qquad \text{Eq. (1)}$$

where ρ indicates reflectance. Hence, ρNIR is reflectance measurements in near-infrared region, and ρRed is reflectance measurements red region.

To ensure spectral consistency between *in situ* measurements and fusion NDVI products that emulated Landsat-like images, we converted the in-situ measurements to Landsat 8 OLI band-like data (red band: 636–673 nm, NIR band: 851–879 nm) using the spectral response function of OLI (Barsi *et al*. 2014) and then computed NDVI values.

The coefficient of determination (R^2) of linear regression models, relative bias (*rbias*; Eq. (2)), and the relative root mean square error ((*rRMSE*; Eq. (3)) were used to evaluate the image fusion products, as in previous studies (e.g. Chen *et al*. 2015a; Zhu *et al*. 2018). To evaluate the linear correlation between *in situ* measurement and fusion products, R^2 was used, which equals squared Pearson correlation for linear regression models. From *rbias*, we can determine whether fusion products overestimate or underestimate compared to *in situ* measurements. *rRMSE* is widely used in quantitative assessments of image quality (Zhou and Bovik 2002).

$$\boldsymbol{rbias} = \frac{\boldsymbol{E}(\|\boldsymbol{A}-\boldsymbol{B}\|)}{\boldsymbol{E}(\boldsymbol{B})} \qquad \text{Eq. (2)}$$

$$\boldsymbol{rRMSE}(\boldsymbol{A}) = \frac{\sqrt{\boldsymbol{E}((\boldsymbol{B}-\boldsymbol{A})^{2})}}{\boldsymbol{E}(\boldsymbol{B})} \qquad \text{Eq. (3)}$$

where A is fusion NDVI or original satellite NDVI products, B is *in situ* NDVI, and E is the mean operator.

To evaluate the overall performance of fusion NDVI products, we compared *in situ* NDVI ($NDVI_{\textit{in situ}}$) data to the fusion NDVI products ($NDVI_{fusion}$; $NDVI_{ESTARFM}$, $NDVI_{FSDAF}$, $NDVI_{STAIR}$, $NDVI_{CESTEM}$) and original satellite NDVI products ($NDVI_{MODIS,}$ $NDVI_{Landsat}$). Regarding spatial variation, $NDVI_{fusion}$ were evaluated on each date when $NDVI_{\textit{in situ}}$ were measured and also evaluated separately for the two land cover classes (i.e., vegetation and mixed land cover). To characterize changes in NDVI within a MODIS pixel, we compared the $NDVI_{MODIS}$ value on each date to the average of the $NDVI_{\textit{in situ}}$ over each land cover class. Regarding temporal variation, $NDVI_{fusion}$ on each plot were evaluated against corresponding $NDVI_{\textit{in situ}}$ throughout the entire growing season. Moreover, to evaluate the time-lag effects on $NDVI_{fusion,}$especially in peak growing season, we used the relative values that are the differences between mean $NDVI_{\textit{in situ}}$ and mean $NDVI_{fusion}$ over the two land cover classes. We evaluated the time-lag effect on $NDVI_{fusion}$ from DOY 206 (available date for both MODIS and Landsat 8) to DOY 246 as $NDVI_{fusion}$ after DOY246 can be influenced by DOY 286 Landsat 8 data, which is near the harvest. For the purpose of comparison, we set all relative values to start from zero on DOY 206. Due to data-gap by the cloud contamination or no data, $NDVI_{MODIS,}$ $NDVI_{Landsat}$, and $NDVI_{fusion}$ were not available on several *in situ* data dates. In this case, we used satellite data within ± 3 days centered on *in situ* date (Table 2). The data dates used for the evaluation are listed in Supplementary 3.

Table 2 Days-of-year used to compare satellite data and fusion products against *in situ* NDVI. Underline: date within ±3 days of in-situ date due to clouds or no-data, *: no data or cloud contamination within ±3 days, **: only a fraction of data was available for the current date.)

Data	Day-of-year

In-situ	155	187	194	206	215	223	239	248	291
MODIS (MOD09GQ)	155	187	195	206	213	223	243	246	290
Landsat (Landsat8 OLI)	*	*	*	206	*	*	*	*	286
ESTARFM	155	187	195	*	213	223	243	246	290
FSDAF	155	187	195	206	213	223	243	246	290
STAIR	155	187	194	206	215	223	239	248	291
CESTEM	155	187	*	**	*	223	238	247	290

3. Results

3.1 Spatio-temporal evaluation of satellite and image fusion NDVI products against *in situ* NDVI

Regression analyses of satellite NDVI products against all available *in situ* measurements showed strong linear relationships and negative bias ($R^2 > 0.93$, *rbias* up to –7%). $NDVI_{Landsat}$ against $NDVI_{in\ situ}$ was strongly correlated over the growing season ($R^2 = 0.93$, *rbias* = –4%). $NDVI_{MODIS}$ also showed a strong correlation with the average $NDVI_{fusion}$ values that covered the 250 m MODIS pixel ($R^2 = 0.95$, *rbias* = –7%; Figure 5).

$NDVI_{fusion}$ products and all available $NDVI_{in\ situ}$ showed strong linear relationships ($R^2 = 0.73–0.93$) and negative biases (rbias up to –9%). Linear relationships were strong with small biases in CESTEM, STAIR, FSDAF, and ESTARFM, with R^2 (*rbias*) of 0.93 (–1%), 0.84 (–2%), 0.75 (–4%), and 0.73 (–9%), respectively (Figure 5).

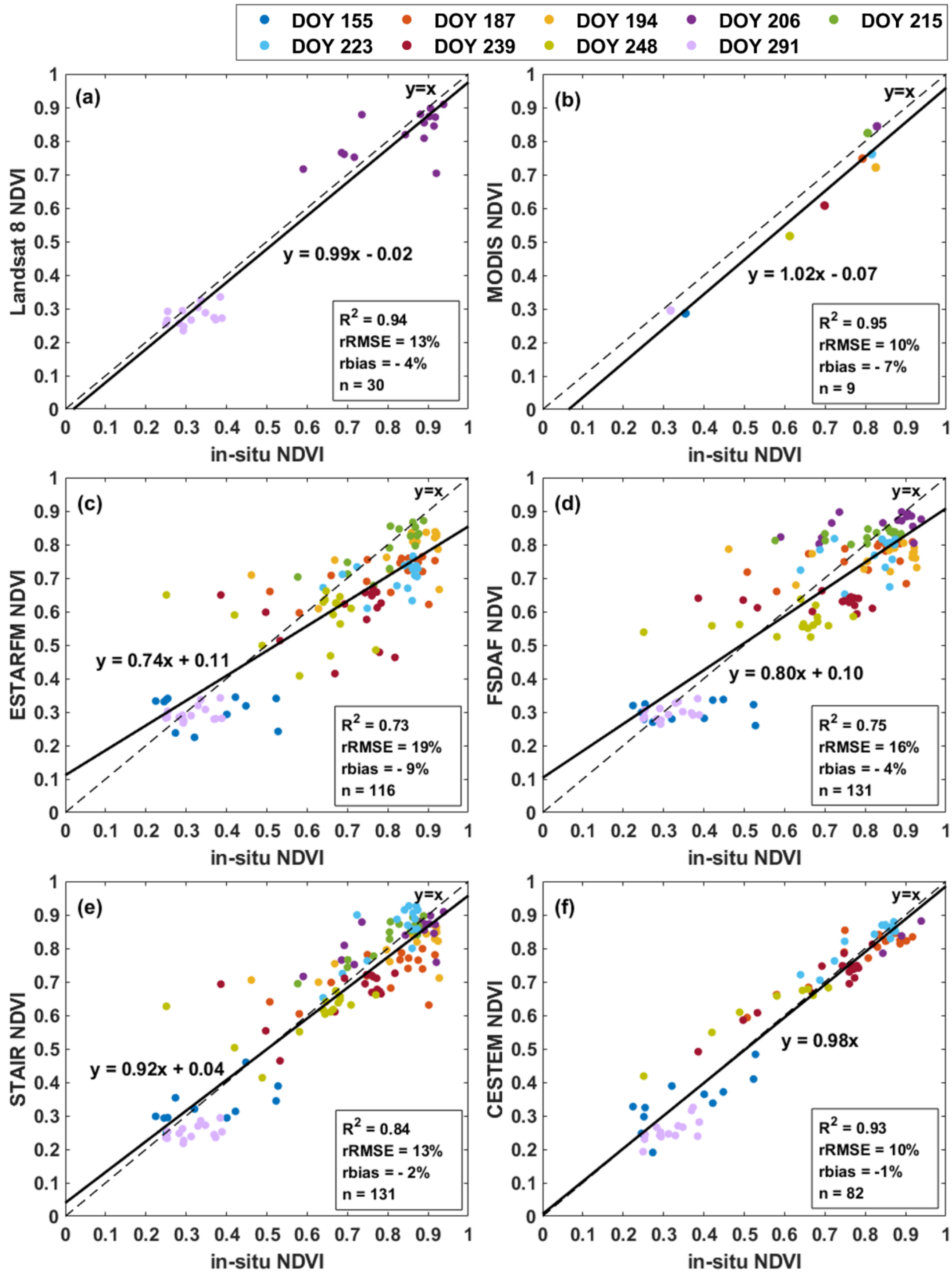


Figure 5 Evaluation of Landsat and MODIS NDVI products (a-b) and image fusion NDVI products (d–f) against *in situ* NDVI data sets across available dates. (a) Landsat 8, (b) MODIS, (c) ESTARFM, (d) FSDAF, (e) STAIR, (f) CESTEM. R^2, *rRMSE*, *rbias*, and n are the coefficient of determination, relative root mean square error, relative bias, and number of samples, respectively. Colored circles indicate NDVI values for each day. Dashed and thick lines indicate one-to-one lines (y = x) and trend lines, respectively.

The difference between $NDVI_{MODIS}$ values and $NDVI_{in\ situ}$ values showed different trends for plots of vegetation and mixed land classes. In the vegetation land class, $NDVI_{MODIS}$ values were negatively biased against 86% of $NDVI_{in\ situ}$ (Figure 6). In the mixed land class, however, $NDVI_{MODIS}$ were positively biased against 76% of $NDVI_{in\ situ}$ (Figure 6).

The bias distributions in $NDVI_{fusion}$ against $NDVI_{in\ situ}$ were similar to the biases in $NDVI_{MODIS}$. Histograms of the difference between $NDVI_{fusion}$ and $NDVI_{in\ situ}$ mainly showed negative bias for the vegetation land class (ESTARFM, 85%; FSDAF, 83%; STAIR, 74%; CESTEM, 69%; Figure 6) and positive bias for the mixed land class (ESTARFM, 71%; FSDAF, 78%; STAIR, 74%; CESTEM, 72%; Figure 6).

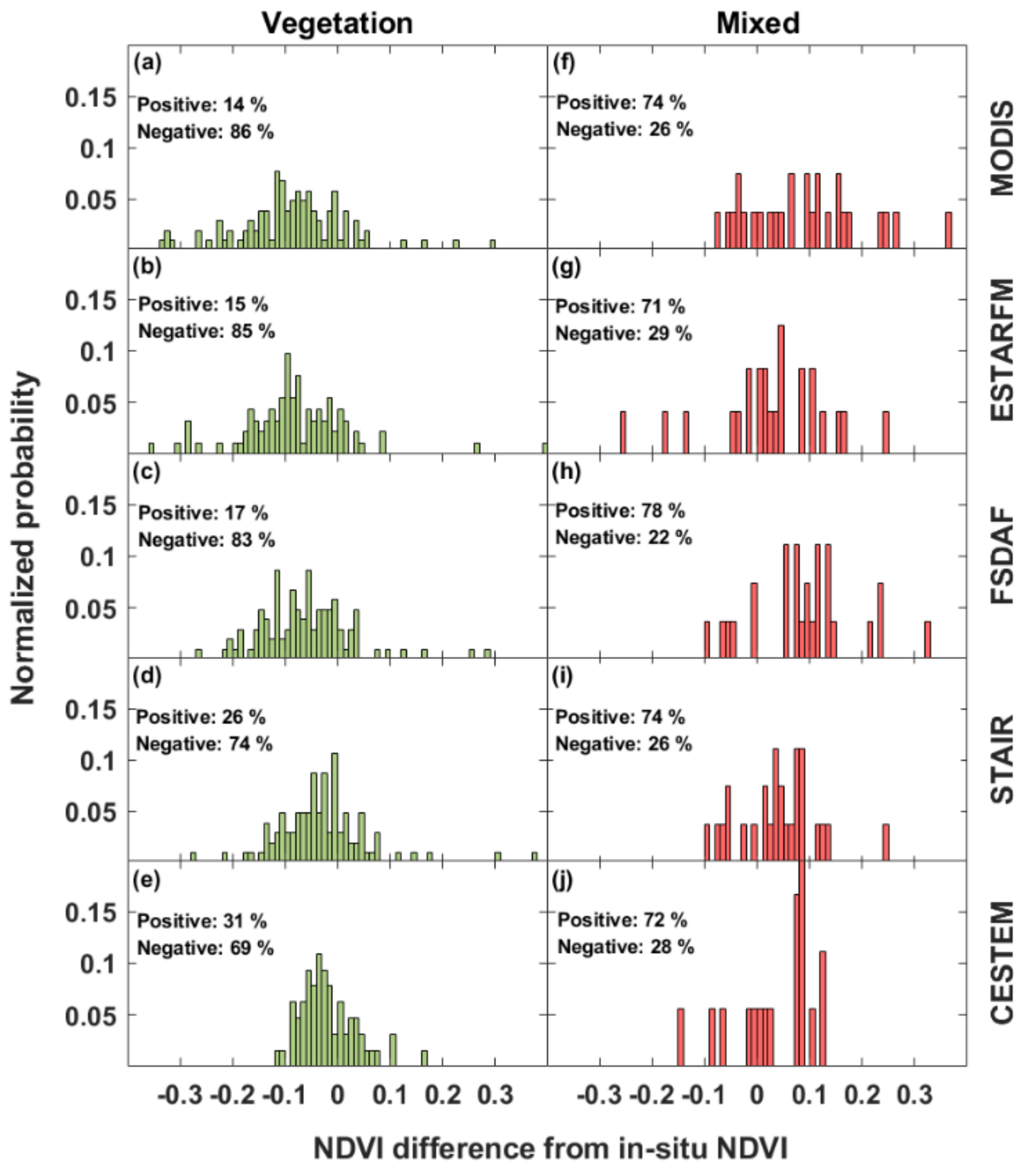


Figure 6 Histogram of difference between MODIS or fusion NDVI and *in situ* NDVI for vegetation (a–e) and mixed (f–j) land classes. Results shown here are based on data from all dates.

3.2 Spatial evaluation of fusion NDVI products against *in situ* NDVI on each date

Regression analyses of $NDVI_{Landsat}$ and $NDVI_{fusion}$ against $NDVI_{in\ situ}$ on each date showed wide ranges in R^2 and rbias (Figure 7; Appendix 4). Spatial relationships between $NDVI_{Landsat}$ and $NDVI_{in\ situ}$ on individual dates (Figure 7; R^2 = 0.17 and 0.22, *rbias* = –12.4% and – 1.0%, Appendix 4) were lower than the overall performance that included spatial and temporal variations (Figure 5a). Spatial relationships between $NDVI_{fusion}$ and $NDVI_{in\ situ}$ for each date were also considerably scattered (Figure 7; R^2 = 0–0.94, *rbias* –20.9% to 11.5%, Appendix 4).

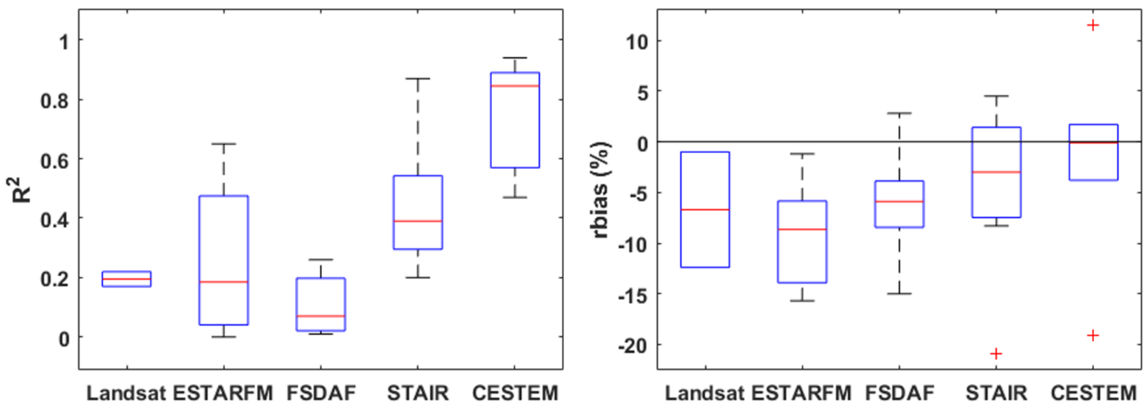


Figure 7 Evaluation of spatial performance of fusion methods and Landsat 8 against *in situ* NDVI on each date. Boxplots (the 25th to 75th percentiles) of R^2 and relative bias(*rbias*) are shown. On each box, the red line indicates the median, and the whiskers extend to both minimum and maximum values that do not consider outliers (+). A table with the corresponding detailed numbers is shown in Appendix 4

3.3 Temporal evaluation of fusion NDVI products against on each plot

Time series NDVI data sets across the 15 plots provided not only the seasonal change in NDVI but also the amplitude of the change in NDVI. Early in the growing season, $NDVI_{ESTARFM}$ and $NDVI_{FSDAF}$ were lower than $NDVI_{in\ situ}$, whereas $NDVI_{STAIR}$ and $NDVI_{CESTEM}$ did not show such discrepancies (Figure 8). All fusion products had similar values at the end of the growing season. At the peak of the growing season, $NDVI_{ESTARFM}$ and $NDVI_{FSDAF}$ showed inconsistent NDVI values. In particular, the NDVI of these two methods showed a clear drop near DOY 200 (Figure 8).

The $NDVI_{fusion}$ on each plot showed linear relationships to $NDVI_{in\ situ}$, but biases varied from negative to positive (Figure 8; Figure 9; Appendix 5).In the vegetation land class, $NDVI_{fusion}$ generally showed a negative bias against $NDVI_{in\ situ}$ (Appendix 5), in particular at the peak of the growing season (Figure 8). In the mixed land class, $NDVI_{fusion}$ generally showed a positive bias against $NDVI_{in\ situ}$ on plots (Figure 8; Appendix 5), in particular at the peak of the growing season (Figure 8).

Among the vegetation land class plots, $NDVI_{fusion}$ on Plot 5 and Plot 11 showed different patterns against $NDVI_{\textit{in situ}}$. $NDVI_{ESTARFM}$, $NDVI_{FSDAF}$, and $NDVI_{STAIR}$ did not catch a sharp decline in $NDVI_{\textit{in situ}}$ on Plot 5 (Figure 8), which led to a positive bias against $NDVI_{\textit{in situ}}$ (Appendix 5). At the peak of the growing season, the differences between $NDVI_{\textit{in situ}}$ and $NDVI_{ESTARFM}$, $NDVI_{FSDAF}$, and $NDVI_{STAIR}$ products were greater for Plot 11 than for the other vegetation plots (Figure 8), which resulted in a negative bias against $NDVI_{\textit{in situ}}$ up to –28.2% (Appendix 5).

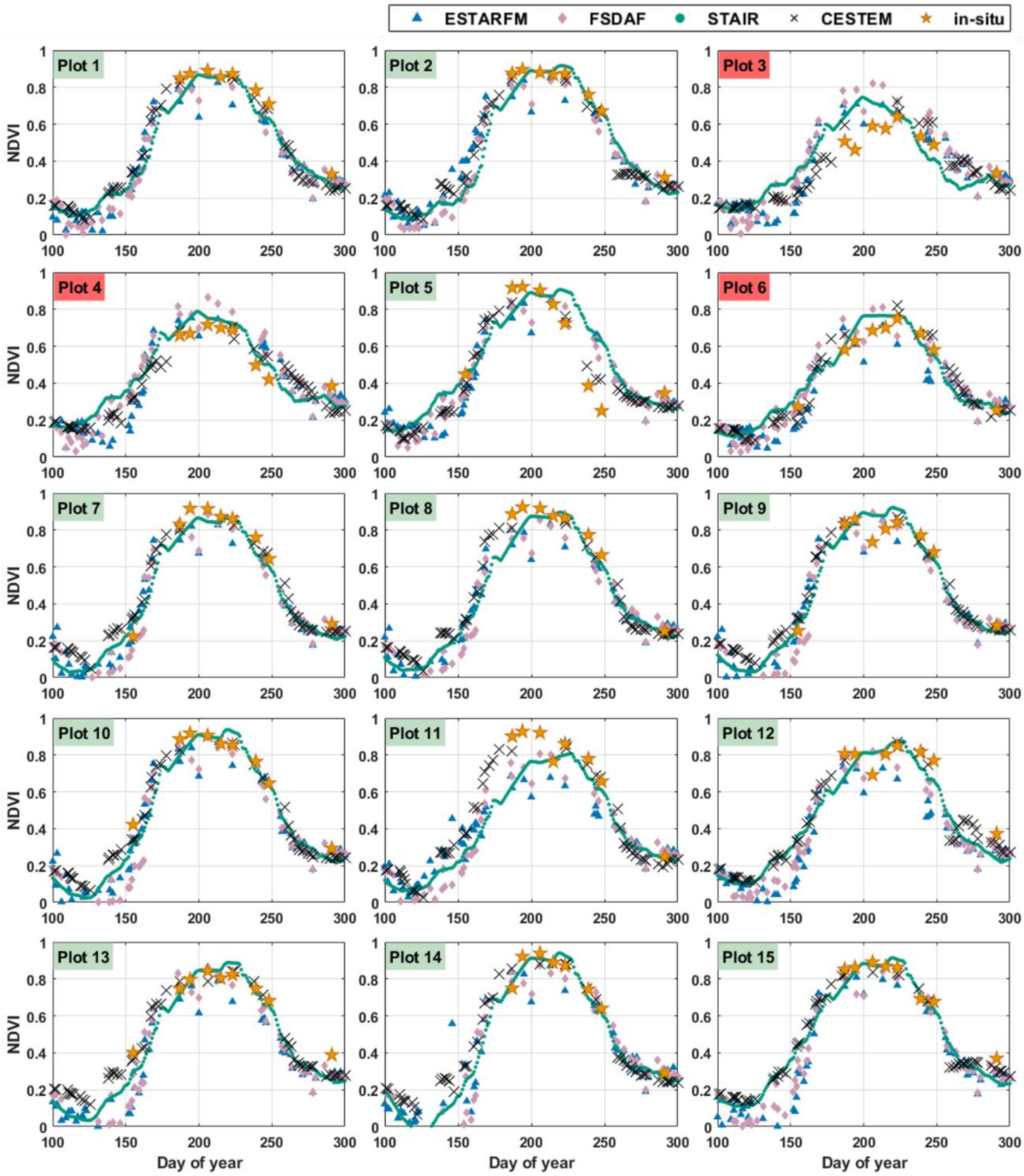


Figure 8 Seasonal variation in NDVI from image fusion products and *in situ* measurements. Background colors of plot numbers indicate vegetation (green) and mixed (red) land classes.

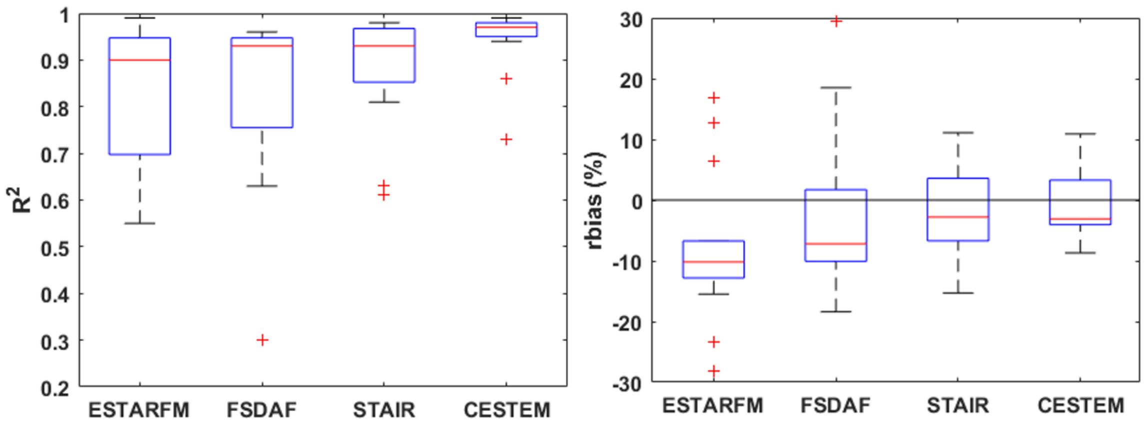


Figure 9 Evaluation of temporal performance of fusion methods against *in situ* NDVI across all 15 plots. Boxplots (the 25th to 75th percentiles) of R^2 and relative bias (*rbias*) are shown. On each box, the red line indicates the median, and the whiskers extend to minimum and maximum values that do not consider outliers (+). A table with the corresponding detailed numbers is shown in Appendix 5.

3.4 Evaluation of time-lag effects across land classes

Differences between $NDVI_{fusion}$ and $NDVI_{\mathit{in\ situ}}$ differed from differences between $NDVI_{Landsat}$ and $NDVI_{\mathit{in\ situ}}$ (blue line in Figure 10) over time. In the vegetation land class, $NDVI_{ESTARFM}$ and $NDVI_{FSDAF}$ showed clear time-lag effects as the relative NDVI difference became larger with time. $NDVI_{STAIR}$ and $NDVI_{CESTEM}$ differed from $NDVI_{Landsat}$ but became closer to $NDVI_{\mathit{in\ situ}}$ around DOY 240. In the mixed land class, $NDVI_{ESTARFM}$ and $NDVI_{STAIR}$ showed clear time-lag effects. $NDVI_{FSDAF}$ and $NDVI_{CESTEM}$ did not show clear time-lag effects.

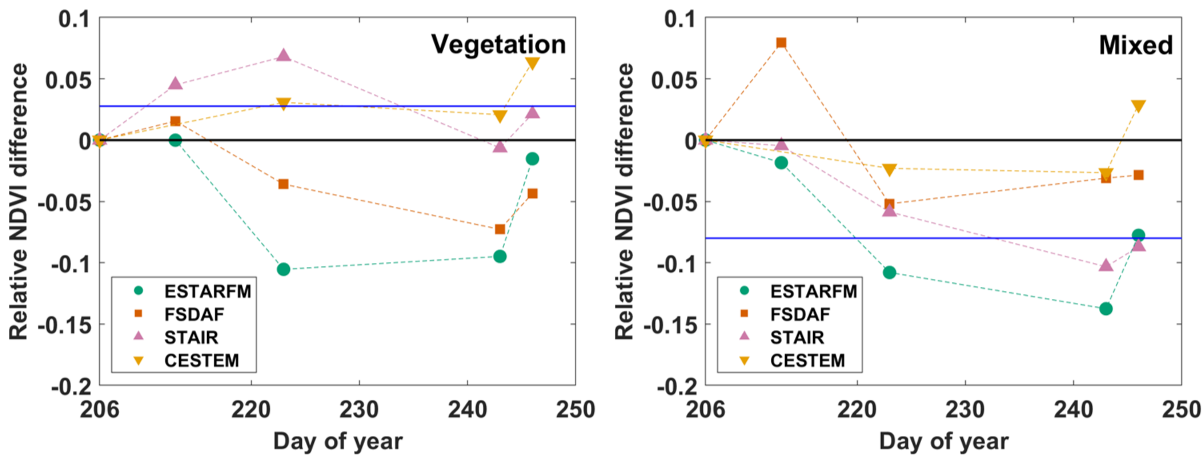


Figure 10 Evaluation of time-lag effects. Time-lag effects denote uncertainties caused by the temporal interpolation process due to the time lag between the dates of the input pair and the dates of predicted fusion products. Relative NDVI differences are the differences between mean

in situ **NDVI and mean fusion NDVI over the two land cover classes. Relative NDVI differences were set to start from zero on DOY 206, which was the date of input pair (Landsat and MODIS) for the fusion products. Blue horizontal lines indicate the difference between Landsat 8 NDVI and** ***in situ*** **NDVI on DOY 206, which is the target quality for fusion NDVI. MODIS was available on DOY206 and 256**

4. Discussion

4.1 Strengths of image fusion products

All four fusion products showed tendencies in performances that are consistent with the design features of their algorithms. This is discussed in more detail in the following for all four methods.

Our results support the ability of ESTARFM and FSDAF to capture spatial and temporal variation, respectively, in land surface properties. As $NDVI_{ESTARFM}$ and $NDVI_{FSDAF}$ products shared identical input data, the difference in output can be entirely attributed to the difference in algorithms. ESTARFM computes the coefficient of linear regression between two pairs (e.g., two Landsat-MODIS pairs) over each pixel because reflectance changes in coarse-spatial-resolution pixels across time do not equal those in fine-resolution pixels (Zhu *et al*. 2010). From the results, ESTARFM was consistent with design of the algorithm on heterogeneous landscape, as it was generally better than FSDAF at capturing spatial variation in NDVI on each date (Figure 7; Appendix 4). FSDAF was developed to capture both gradual and abrupt changes in land cover more effectively than ESTARFM by distributing residuals of temporal change using thin-plate spline interpolation (Zhu *et al*. 2016). This design of FSDAF was consistent with the results as $NDVI_{FSDAF}$ outperformed $NDVI_{ESTARFM}$ in explaining temporal variation for each plot (Figure 9; Appendix 5). FSDAF was also less sensitive than ESTARFM to time-lag effects (Figure 10).

STAIR confirmed the strengths of cloud-/gap-free products. The number of retrieved dates (n = 214) was highest for STAIR compared to the other products (n = 65, 66, and 60 for ESTARFM, FSDAF, CESTEM, respectively; Supplementary 3). The spatiotemporal variation in STAIR NDVI agreed better with $NDVI_{\textit{in situ}}$ values than that in $NDVI_{ESTARFM}$ and $NDVI_{FSDAF}$ (Figure 5; Figure 7; Figure 9). As time-lag effects were smaller in $NDVI_{STAIR}$ than in $NDVI_{ESTARFM}$ and $NDVI_{FSDAF}$ products (Figure 10). Moreover, the additional input pairs (e.g., Landsat 7) of $NDVI_{STAIR}$ were better than $NDVI_{ESTARFM}$ at capturing spatial variation in $NDVI_{\textit{in situ}}$ at the peak of the growing season (Appendix 4). In fact, STAIR even captured spatial NDVI patterns better than the original Landsat images (Figure 7; Appendix 4). The strong performance of STAIR is likely due to the use a temporal gap-filling method involving all available Landsat images (e.g., Landsat 7 and Landsat 8) to produce daily cloud-free products without missing pixels.

CESTEM fully utilized the merits of CubeSat constellation data (daily, 3 m resolution) in its harmonization process with Landsat and MODIS. The CESTEM product has already been proven a promising data set for land surface monitoring, but detailed evaluation of the product quality with pixel-scale *in situ* measurements has not been attempted before (Aragon *et al*. 2018; Houborg and McCabe 2018b). Regarding both spatial and temporal resolution, $NDVI_{CESTEM}$ showed the strongest linear relationship with the smallest bias to $NDVI_{in\ situ}$ values of all of the products (Figure 5; Figure 7; Figure 8). Also, spatial and temporal variation in $NDVI_{CESTEM}$ were in better agreement with $NDVI_{in\ situ}$ than that of $NDVI_{Landsat}$ against $NDVI_{in\ situ}$ despite aggregating the CESTEM product to the coarser spatial resolution of Landsat (Figure 5, Figure 7; Figure 9). We partly attribute this strong performance of CESTEM to lower uncertainties in input data and noise reduction via data aggregation (see section 4.2 below).

4.2 Impacts of input data and fusion algorithms on uncertainties in fusion NDVI

$NDVI_{Landsat}$ is the main driver that controls spatial variation in $NDVI_{fusion}$ products, so uncertainties in $NDVI_{Landsat}$ must be quantified systematically. Several studies have evaluated $NDVI_{Landsat}$ time series against ground-based spectral sensors in fixed positions (Ke *et al*. 2015; Kim *et al*. 2019; Ryu *et al*. 2014); however, a comprehensive evaluation of $NDVI_{Landsat}$ against *in situ* spectral sensors in terms of spatial variation is lacking. We found that $NDVI_{Landsat}$ explained only 17–22% of spatial variation over the landscape for each date (DOY 206, the peak of the growing season; DOY 286 after harvest; Figure 7; Appendix 4). Because we aligned the spectral response, acquisition time, and footprint between Landsat and *in situ* data, we assume that uncertainties in geolocation in Landsat 8 might have contributed to the scattered relationships. Earlier studies have reported that the absolute geolocation accuracy of a Landsat 8 Tier 2 product (i.e., surface reflectance) is greater than 12 m (Dwyer *et al*. 2018; Storey *et al*. 2014). Of the four image fusion products, ESTARFM and FSDAF are particularly dependent on Landsat 8 for generating sub-MODIS pixel spatial variability. Indeed, scattered relationships were also shown in the $NDVI_{fusion}$ on each date (Figure5; Figure 7; Appendix 4). Noteworthy is the fact that $NDVI_{CESTEM}$ had a higher R^2 than $NDVI_{Landsat}$ on DOY291 (Figure 7; Appendix 4). During CESTEM pre-processing, Landsat data is geometrically aligned to the CubeSat data to ensure a near-perfect alignment (Houborg and McCabe 2018a). Unlike the other products, CESTEM relies on CubeSats, which have a relatively high positional accuracy (<10 m) and therefore likely to be better correlated with the observed spatial NDVI variability (Planet Team 2018). In addition, we aggregated 3 m CESTEM pixels to match the Landsat pixel resolution (30 m), which will reduce random noise. This difference between Landsat and CESTEM might explain the performance of $NDVI_{CESTEM}$ in capturing spatial variation in $NDVI_{in\ situ}$ on each date (Figure 7; Appendix 4) which was superior to that of the other fusion NDVI products.

$NDVI_{MODIS}$ was negatively biased toward $NDVI_{in\ situ}$ (Figure 5). $NDVI_{MODIS}$ played a role in temporal patterns of $NDVI_{fusion}$. Ideally, $NDVI_{in\ situ}$ that covers 3 × 3 MODIS pixels would be measured to account for uncertainties in the MODIS geolocation product (Wolfe *et al*. 2002). However, this is extremely labor intensive and was beyond the scope of our experiment. We assumed that the negative bias in $NDVI_{MODIS}$ would result in the underestimation of $NDVI_{fusion}$ in the vegetation land class at the peak of the growing season (Figure 8). As the vegetation land class accounted for 76% of the MODIS 250 m pixel (section 2.2), the $NDVI_{MODIS}$ values mainly represented the dynamics of the vegetation land class. However, the $NDVI_{MODIS}$ values were higher than $NDVI_{in\ situ}$ in the mixed land class within a MODIS subpixel (Figure 6). The positive bias of $NDVI_{MODIS}$ against $NDVI_{in\ situ}$ on mixed land class might explain the overestimation of $NDVI_{fusion}$ values than $NDVI_{in\ situ}$ values in the mixed land class, in particular at the peak of the growing season (Figure 8). These findings suggest that biases in $NDVI_{MODIS}$ can lead to different signs in biases for dominant and minor land cover classes within the MODIS pixel.

The coarse spatial resolution of MODIS caused biases in the spatial interpolation process. In the vegetation land class, Plot 5 was harvested 2 weeks earlier than the other plots (section 2.2), so it had a lower $NDVI_{in\ situ}$ value than the other vegetation class plots around DOY 230 (Figure 8). The area harvested on DOY 230 was only 3.7% of the total study area over MODIS 250 m pixels. Therefore, the harvest event on Plot 5 was not reflected in the $NDVI_{MODIS}$ value, which explains the positively biased $NDVI_{fusion}$ values in this plot (Figure 8). $NDVI_{CESTEM}$ successfully detected the early harvest in Plot 5 (Figure 8) and showed a stronger linear relationship against $NDVI_{in\ situ}$ for Plot 5 compared to the other $NDVI_{fusion}$ (Appendix 5). This performance of $NDVI_{CESTEM}$ is due to the use of the CubeSat constellation that has near-daily temporal coverage at fine spatial resolution. (Houborg and McCabe 2018a).

Weight function–based fusion NDVI, which includes ESTARFM, FSDAF, and STAIR, revealed uncertainties across land cover transition zones. Such methods give greater weight to neighboring pixels for spatial interpolation (Zhu *et al*. 2018). Plot 11 was close to a road and ditch to the east and south (Figure 2); thus, five out of eight neighboring pixels were classified as belonging to the mixed class, which had lower NDVI than the vegetation class during the peak of the growing season. Indeed, compared to the other vegetation plots, $NDVI_{fusion}$ values for Plot 11 were underestimated during the peak growing season (Figure 8). Presumably, the mixed land class surrounding Plot 11 caused the negative bias through the weight function. Thus, the correlation between $NDVI_{fusion}$ values and $NDVI_{in\ situ}$ values on Plot 11 was relatively weak (Appendix 5). By contrast, $NDVI_{CESTEM}$ values use each CubeSat pixel without spatial interpolation. Therefore, $NDVI_{CESTEM}$ values on Plot 11 captured the dynamics of $NDVI_{in\ situ}$ (Figure 8).

Time-lag effects did not always degrade the performance of $NDVI_{fusion}$ when input data were biased against *in situ* data. Previous studies have reported that time-lag effects degrade the performance of fusion products against input data (Fu *et al*. 2015; Gao *et al*. 2006; Olexa and Lawrence 2014; Walker *et al*. 2012; Xie *et al*. 2018; Zhu *et al*. 2010). As in earlier studies, the relative difference in NDVI between fusion products and *in situ* data tended to increase over time from the paired dates, in particular for $NDVI_{ESTARFM}$ and $NDVI_{FSDAF}$ in the vegetation land class and $NDVI_{ESTARFM}$ and $NDVI_{STAIR}$ in the mixed land class (Figure 10). However, time-lag effects in $NDVI_{STAIR}$ and $NDVI_{CESTEM}$ in the vegetation land class were not apparent (Figure 10). In the mixed land class, $NDVI_{FSDAF}$ and $NDVI_{CESTEM}$, which did not show clear time-lag effects, did not perform better than $NDVI_{ESTARFM}$ and $NDVI_{STAIR}$, which showed clear time-lag effects, in predicting *in situ* NDVI (Figure 10). Thus, when the input pair data include biases against *in situ* data, time-lag effects do not necessarily correspond to a decrease in performance against *in situ* data.

4.3 Future works and perspectives

Rapid and substantial advancements in image fusion methods have emerged in recent years. Even during the preparation of this report, we found that each image fusion method used in this study had been improved. For example, highly scalable STARFM was recently implemented on Google Earth Engine (Gorelick *et al.* 2017) to produce daily gap-free reflectance products (Moreno-Martínez *et al*. 2020). FSDAF was improved by the incorporation of constrained least squares theory and the subpixel class fraction change, called IFSDAF (Liu *et al*. 2019) and SFSDAF (Li *et al*. 2020), respectively. In the present study, STAIR 1.0 integrated all available Landsat and MODIS data sets. Recently, STAIR also assimilated Sentinel-2 data (STAIR 2.0) (Luo *et al.* 2020), CubeSats, and the latter work allows the production of 3 m daily leaf area index (LAI) maps across the U.S. Corn Belt (Kimm *et al*. 2020). STAIR also develops the real-time production capability that enables the same-day 10m and 30m daily surface reflectance products for anywhere in the earth, through Aspiring Universe Corporation for the commercialization applications. CESTEM version 1.0, used in the present study, did not include a gap-filling algorithm, which led to substantial data gaps in parts of the images (e.g., DOY246 in Figure 4) or even entire scenes during the growing season (e.g., DOY 225 to DOY 237). Recently, a significantly updated implementation of CESTEM has become an integral component towards Planet's vision of producing a next generation, analysis ready, and harmonized product, which delivers clean (i.e., free from clouds and shadows), gap-filled (i.e., daily, 3 m), temporally consistent, and radiometrically robust surface reflectance integrating the best features from both public (e.g., Landsat, Sentinel, MODIS) and private missions (Houborg and Zuleta 2019). Evaluating these recently updated image fusion methods is beyond the scope of this work, but we anticipate that our systematic, in situ spectral data will be helpful in evaluating those methods.

Current study has limitations and suggests further improvements. We focused on a flat, rice paddy landscape to evaluate image fusion products. The data collection and evaluation scheme in this study might be applied to other ecosystems in short vegetation. For tall woody ecosystems, our manual data collection scheme will not work. In this case, drone could be used to generate reference maps in space and time (Candiago *et al.* 2015; Hashimoto *et al.* 2019; Zhang *et al.* 2020). Our results suggest each image fusion algorithm has pros and cons. Furthermore, each algorithm has been under rapid improvements as discussed above. Though one convergent lesson is clear. In heterogeneous landscapes not limited to rice paddy, image fusion algorithm based on spatial or temporal interpolation methods will involve inherent random errors and biases. Therefore, preparing more fine resolution images that allow more input pair dates and minimize interval of prediction dates is the key to reduce uncertainties in image fusion methods.

Our results highlight the importance of fine-spatial-resolution images with high revisit frequency. In the past, MODIS played a major role in capturing temporal variation, as satellite images that capture fine spatial scales, such as those captured by Landsat, have low revisit frequencies. With the emergence of new satellites and CubeSat constellations, it is possible to produce images that have both fine spatial and temporal resolution, such as Harmonized Landsat/Sentinel-2 products (Claverie et al. 2018) and PlanetScope products (Planet Team 2018). These harmonized data sets require careful, precise cross-calibration across satellites. Recent advancements in inexpensive spectral sensors (Kim et al. 2019; Ryu et al. 2010) offer new opportunities to continuously monitor land surface reflectance in situ, which could be deployed as a network across landscape. Incorporating geostationary satellites into the image fusion framework will help manage sustained data gaps due to clouds and decrease uncertainties in fusion products caused by temporal interpolation. We envision that the integration of ground-based spectral sensing networks and harmonized satellite images will lead to fundamental advancements in high-quality image fusion products.

5. Summary and Conclusions

In this study, we comprehensively evaluated four state-of-the-art image fusion products against *in situ* spectral measurements over a heterogeneous rice paddy landscape during the 2017 growing season. All four fusion products showed strong linear correlations to in-situ data when focusing on temporal observations or pooling spatial and temporal observations. However, we found large differences in performance when evaluating performance for spatial variations. Overall, CESTEM outperformed the other three products that relied on the spatial interpolation and STAIR showed better performance than ESTARFM and FSDAF. The detailed evaluation of fusion products against *in situ* measurements revealed the following insights on the causes underlying the limited performance of fusion methods. First, biases in forcing data (i.e., Landsat 8 and MODIS) propagated rather directly to image fusion products. Second, Landsat 8 NDVI only explained 17–22% of spatial variability against *in situ* data, which explains the poor performance of the Landsat-based fusion products to predict NDVI spatial patterns. Third, the coarse spatial resolution of MODIS led to NDVI biases in vegetation and mixed land cover classes. Fourth, image fusion products based on spatial interpolation showed large biases in vegetation plots surrounded by mixed land cover plots. Finally, time-lag effects did not always degrade the performance of fusion NDVI when input pair data were biased. Our results identify key sources of uncertainty, which will be important for improving image fusion products. Also, constellation-based products such as CESTEM are expected to be more widely used due to their robust performance and high spatio-temporal resolution without the need for spatial interpolation.

Acknowledgement:

This research was conducted with support of the Korea Environment Industry & Technology Institute (KEITI) through its Urban Ecological Health Promotion Technology Development Project funded by the Korea Ministry of Environment (MOE) (2019002760002). English proofreading was supported by the Research Institute of Agriculture and Life Sciences, Seoul National University. We have made our *in situ* data openly accessible via http://environment.snu.ac.kr/landscape_spectral/ and expect that they will be useful for evaluating future image fusion products.

6. References

Aragon, B., Houborg, R., Tu, K., Fisher, J.B., & McCabe, M. (2018). CubeSats enable high spatiotemporal retrievals of crop-water use for precision agriculture. Remote Sensing, 10, 1867

Barsi, J., Lee, K., Kvaran, G., Markham, B., & Pedelty, J. (2014). The spectral response of the Landsat-8 Operational Land Imager. Remote Sensing, 6, 10232

Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., & Gattelli, M. (2015). Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. Remote Sensing, 7, 4026-4047

Chen, B., Huang, B., & Xu, B. (2015a). Comparison of spatiotemporal fusion models: A review. Remote Sensing, 7, 1798–1835

Chen, B., Huang, B., & Xu, B. (2015b). Fine land cover classification using daily synthetic Landsat-like images at 15-m resolution. IEEE Geoscience and Remote Sensing Letters, 12, 2359–2363

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

Czapla-Myers, J., McCorkel, J., Anderson, N., Thome, K., Biggar, S., Helder, D., Aaron, D., Leigh, L., & Mishra, N. (2015). The ground-based absolute radiometric calibration of Landsat 8 OLI. Remote Sensing, 7, 600–626

Dechant, B., Ryu, Y., & Kang, M. (2019). Making full use of hyperspectral data for gross primary productivity estimation with multivariate regression: Mechanistic insights from observations and process-based simulations. Remote Sensing of Environment, 234, 111435

Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J.A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., Kang, M., Li, J., & Moya, I. (2020). Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops. Remote Sensing of Environment, 241, 111733

Ding, Y., Zhao, K., Zheng, X., & Jiang, T. (2014). Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI imagery. International Journal of Applied Earth Observation and Geoinformation, 30, 139–145

Dwyer, J.L., Roy, D.P., Sauer, B., Jenkerson, C.B., Zhang, H.K., & Lymburner, L. (2018). Analysis ready data: Enabling analysis of the Landsat archive. Remote Sensing, 10, 1363

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., & van Dijk, A.I.J.M. (2013). Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. Remote Sensing of Environment, 133, 193–209

Feng, G., Yufang, J., Schaaf, C.B., & Strahler, A.H. (2002). Bidirectional NDVI and atmospherically resistant BRDF inversion for vegetation canopy. IEEE Transactions on Geoscience and Remote Sensing, 40, 1269–1278

Fu, D.J., Zhang, L.F., Chen, H., Wang, J., Sun, X.J., & Wu, T.X. (2015). Assessing the effect of temporal interval length on the blending of Landsat-MODIS surface reflectance for different land cover types in southwestern continental United States. ISPRS International Journal of Geo-Information, 4, 2542–2560

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., & Prueger, J.H. (2017a). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sensing of Environment, 188, 9–25

Gao, F., Anderson, M.C., Zhang, X.Y., Yang, Z.W., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., & Prueger, J.H. (2017b). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sensing of Environment, 188, 9–25

Gao, F., He, T., Masek, J.G., Shuai, Y., Schaaf, C.B., & Wang, Z. (2014). Angular effects and correction for medium resolution sensors to support crop monitoring. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 4480–4489

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

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18-27

Hashimoto, N., Saito, Y., Maki, M., & Homma, K. (2019). Simulation of Reflectance and Vegetation Indices for Unmanned Aerial Vehicle (UAV) Monitoring of Paddy Fields. Remote Sensing, 11, 2119

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., & White, J.C. (2009a). 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

Hilker, T., Wulder, M.A., Coops, N.C., Seitz, N., White, J.C., Gao, F., Masek, J.G., & Stenhouse, G. (2009b). 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, 1988–1999

Houborg, R., & McCabe, M.F. (2018a). A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. Remote Sensing of Environment, 209, 211–226

Houborg, R., & McCabe, M.F. (2018b). Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. Remote Sensing, 10, 890

Houborg, R., & Zuleta, I. (2019). Very high resolution (daily, 3 m) data cubes of surface reflectance synergizing observations from CubeSats, Sentinel-2, Landsat 8, and MODIS. In, AGU Fall Meeting Abstracts (pp. B54D-01)

Huang, Y., Ryu, Y., Jiang, C., Kimm, H., Kim, S., Kang, M., & Shim, K. (2018). BESS-Rice: A remote sensing derived and biophysical process-based rice productivity simulation model. Agricultural and Forest Meteorology, 256–257, 253–269

Hwang, Y., Ryu, Y., Huang, Y., Kim, J., Iwata, H., & Kang, M. (2020). Comprehensive assessments of carbon dynamics in an intermittently-irrigated rice paddy. Agricultural and Forest Meteorology, 285–286, 107933

Jiang, C., Guan, K., Pan, M., Ryu, Y., Peng, B., & Wang, S. (2020). BESS-STAIR: A framework to estimate daily, 30 m, and all-weather crop evapotranspiration using multi-source satellite data for the US Corn Belt. Hydrol. Earth Syst. Sci., 24, 1251–1273

Ke, Y., Im, J., Lee, J., Gong, H., & Ryu, Y. (2015). Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations. Remote Sensing of Environment, 164, 298–313

Kim, J., Ryu, Y., Jiang, C., & Hwang, Y. (2019). Continuous observation of vegetation canopy dynamics using an integrated low-cost, near-surface remote sensing system. Agricultural and Forest Meteorology, 264, 164–177

Kimm, H., Guan, K., Jiang, C., Peng, B., Gentry, L.F., Wilkin, S.C., Wang, S., Cai, Y., Bernacchi, C.J., Peng, J., & Luo, Y. (2020). Deriving high-spatiotemporal-resolution leaf area index for agroecosystems in the U.S. corn belt using Planet Labs CubeSat and STAIR fusion data. Remote Sensing of Environment, 239, 111615

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

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., & Helmer, E.H. (2019). An Improved Flexible Spatiotemporal Data Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote Sensing of Environment, 227, 74–89

Liu, X., Liu, L., Hu, J., & Du, S. (2017). Modeling the footprint and equivalent radiance transfer path length for tower-based hemispherical observations of chlorophyll fluorescence. Sensors (Basel), 17

Liu, X.L., Bo, Y.C., Zhang, J., & He, Y.Q. (2015). Classification of C3 and C4 vegetation types using MODIS and ETM plus blended high spatio-temporal resolution data. Remote Sensing, 7, 15244–15268

Luo, Y., Guan, K., Peng, J., Wang, S., & Huang, Y. (2020). STAIR 2.0: A Generic and Automatic Algorithm to Fuse Modis, Landsat, and Sentinel-2 to Generate 10 m, Daily, and Cloud-/Gap-Free Surface Reflectance Product. Remote Sensing, 12, 3209

Luo, Y., Guan, K., & Peng, J. (2018). STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. Remote Sensing of Environment, 214, 87–99

Moreno-Martínez, Á., Izquierdo-Verdiguier, E., Maneta, M.P., Camps-Valls, G., Robinson, N., Muñoz-Marí, J., Sedano, F., Clinton, N., & Running, S.W. (2020). Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. Remote Sensing of Environment, 247, 111901

Morisette, J.T., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J.E., Garrigues, S., Shabanov, N.V., Weiss, M., Fernandes, R.A., Leblanc, S.G., Kalacska, M., Sanchez-Azofeifa, G.A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A.N., Lewis, T.E., Iiames, J.S., Colombo, R., Meroni, M., Busetto, L., Cohen, W.B., Turner, D.P., Warner, E.D., Petersen, G.W., Seufert, G., & Cook, R. (2006). Validation of global moderate-resolution LAI products: A framework proposed within the CEOS land product validation subgroup. IEEE Transactions on Geoscience and Remote Sensing, 44, 1804–1817

Olexa, E.M., & Lawrence, R.L. (2014). Performance and effects of land cover type on synthetic surface reflectance data and NDVI estimates for assessment and monitoring of semi-arid rangeland. International Journal of Applied Earth Observation and Geoinformation, 30, 30–41

Planet Team (2018). Planet satellite imagery products. In. [WWW Document]. https://www.planet.com/products/planet-imagery/

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, 309–317

Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C.B., 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

Ryu, Y., Baldocchi, D.D., Verfaillie, J., Ma, S., Falk, M., Ruiz-Mercado, I., Hehn, T., & Sonnentag, O. (2010). Testing the performance of a novel spectral reflectance sensor, built with light emitting diodes (LEDs), to monitor ecosystem metabolism, structure and function. Agricultural and Forest Meteorology, 150, 1597–1606

Ryu, Y., Lee, G., Jeon, S., Song, Y., & Kimm, H. (2014). Monitoring multi-layer canopy spring phenology of temperate deciduous and evergreen forests using low-cost spectral sensors. Remote Sensing of Environment, 149, 227–238

Schmidt, M., Lucas, R., Bunting, P., Verbesselt, J., & Armston, J. (2015). Multi-resolution time series imagery for forest disturbance and regrowth monitoring in Queensland, Australia. Remote Sensing of Environment, 158, 156–168

Senf, C., Leitao, P.J., Pflugmacher, D., van der Linden, S., & Hostert, P. (2015). Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sensing of Environment, 156, 527–536

Storey, J., Choate, M., & Lee, K. (2014). Landsat 8 Operational Land Imager on-orbit geometric calibration and performance. Remote Sensing, 6, 11127–11152

Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8, 127–150

Walker, J.J., de Beurs, K.M., & Wynne, R.H. (2014). Dryland vegetation phenology across an elevation gradient in Arizona, USA, investigated with fused MODIS and Landsat data. Remote Sensing of Environment, 144, 85–97

Walker, J.J., de Beurs, K.M., Wynne, R.H., & Gao, F. (2012). Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. Remote Sensing of Environment, 117, 381–393

Wang, P., Gao, F., & Masek, J.G. (2014). Operational data fusion framework for building frequent Landsat-like imagery. IEEE Transactions on Geoscience and Remote Sensing, 52, 7353–7365

Wolfe, R.E., Nishihama, M., Fleig, A.J., Kuyper, J.A., Roy, D.P., Storey, J.C., & Patt, F.S. (2002). Achieving sub-pixel geolocation accuracy in support of MODIS land science. Remote Sensing of Environment, 83, 31–49Xie, D., Gao, F., Sun, L., & Anderson, M. (2018). Improving spatial-temporal data fusion by choosing optimal input image pairs. Remote Sensing, 10, 1142

Yang, K., Ryu, Y., Dechant, B., Berry, J.A., Hwang, Y., Jiang, C., Kang, M., Kim, J., Kimm, H., Kornfeld, A., & Yang, X. (2018). Sun-induced chlorophyll fluorescence is more strongly related to absorbed light than to photosynthesis at half-hourly resolution in a rice paddy. Remote Sensing of Environment, 216, 658–673

Zhang, N., Su, X., Zhang, X., Yao, X., Cheng, T., Zhu, Y., Cao, W., & Tian, Y. (2020). Monitoring daily variation of leaf layer photosynthesis in rice using UAV-based multi-spectral imagery and a light response curve model. Agricultural and Forest Meteorology, 291, 108098

Zheng, Y., Wu, B.F., Zhang, M., & Zeng, H.W. (2016). Crop phenology detection using high spatio-temporal resolution data fused from SPOT5 and MODIS products. Sensors, 16, 21

Zhou, J., Chen, J., Chen, X., Zhu, X., Qiu, Y., Song, H., Rao, Y., Zhang, C., Cao, X., & Cui, X. (2021). Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction. *Remote Sensing of Environment, 252*, 112130

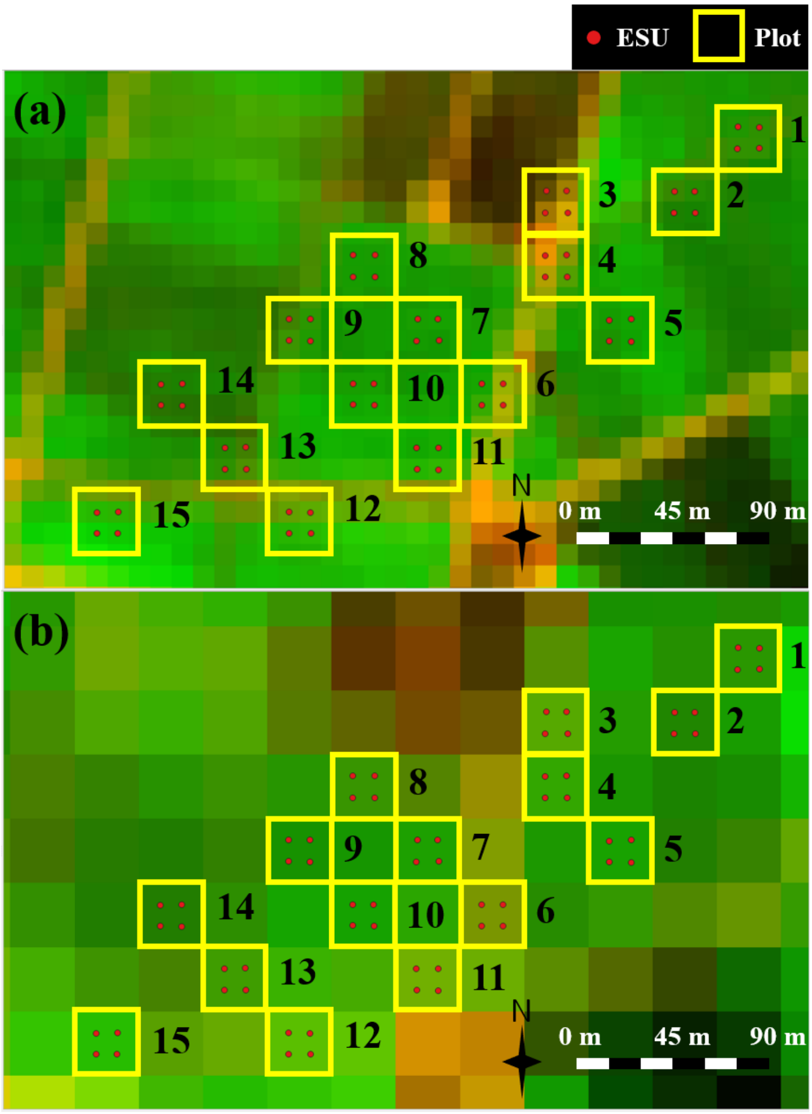
Zhou, W., & Bovik, A.C. (2002). A universal image quality index. IEEE Signal Processing Letters, 9, 81–84

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, 527

Zhu, X.L., Chen, J., Gao, F., Chen, X.H., & 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.L., Helmer, E.H., Gao, F., Liu, D.S., Chen, J., & Lefsky, M.A. (2016). A flexible spatiotemporal method for fusing satellite images with different resolutions. Remote Sensing of Environment, 172, 165–177

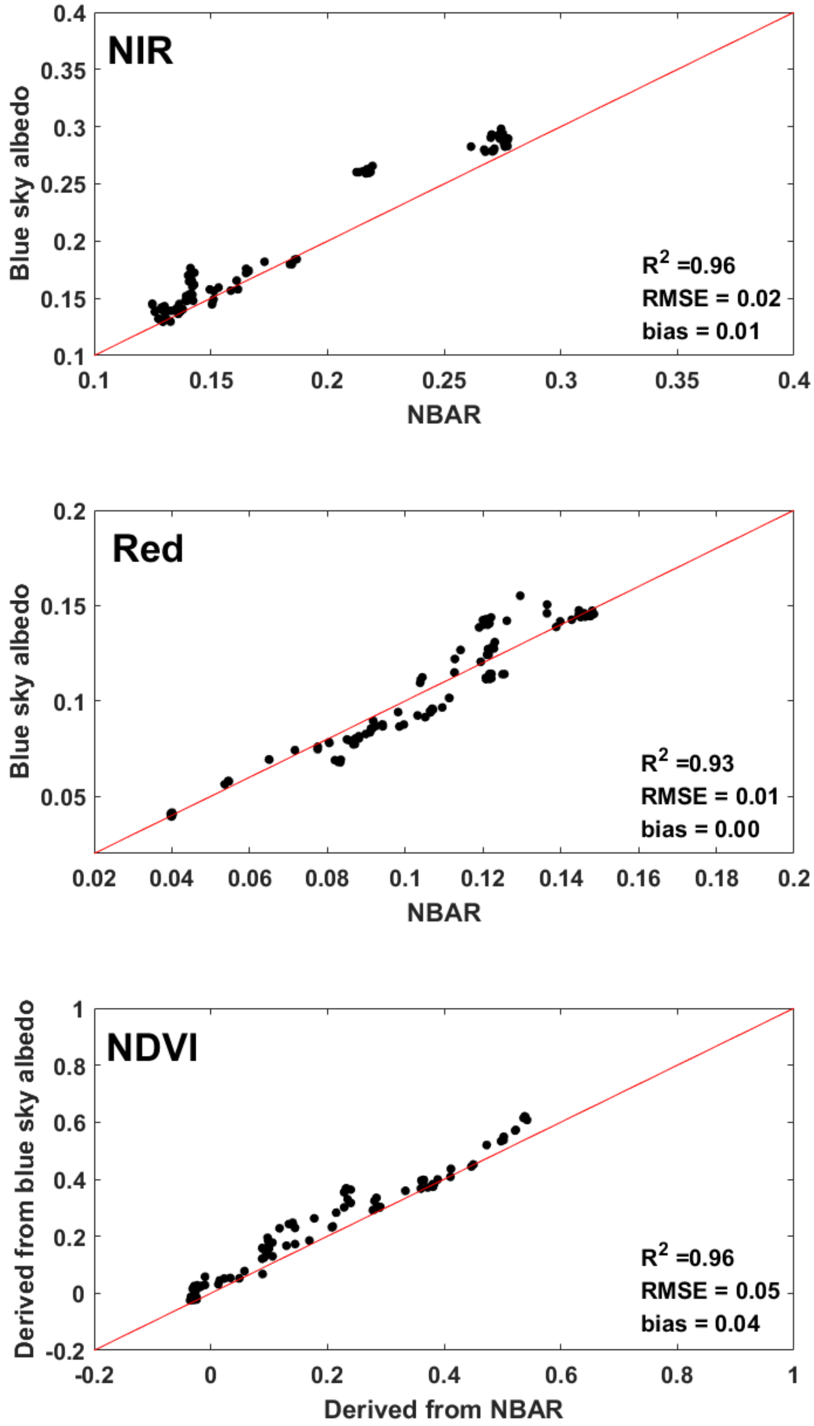
Appendix



Appendix 1 The location of ESUs over (a) Sentinel-2 images (DOY 173 since 2017) and (b) Landsat 8 images (DOY 174 since 2017; RGB rendering: red, NIR, none).

Appendix 2 Quality check of Jaz spectrometer data

To check the Jaz spectrometer data, we used an ASD spectrometer (FieldSpec 4 Wide-ResField Spectroradiometer, ASD, Boulder, CO, USA) on June 4 (DOY 155). Near the flux tower, we compared *in situ* Jaz spectral measurements with cosine-corrected fiber to 20 ASD measurements with bare fiber within the Jaz spectrometer footprint. The differences in band between ASD and Jaz in Landsat 8 band-like data were less than 0.05 (ASD red: 0.065 $\pm$0.007, NIR: 0.127 $\pm$0.006, NDVI = 0.327 $\pm$0.053 [95% confidence interval]; Jaz red: 0.054 $\pm$0.0001, NIR: 0.118 $\pm$0.004 NDVI = 0.376 $\pm$0.012 [95% confidence interval]).



Appendix 3 Evaluation of view geometry effects on *in situ* spectral measurements. Using MCD43A1 BRDF parameters, we converted *in situ* bi-hemispheric reflectance measured by Jaz spectrometer (Yang *et al*. 2018) into blue sky albedo and Nadir BRDF-adjusted reflectance (NBAR) at 10:30 a.m. (local time, UTC +9). *In situ* data between DOY 100 and DOY 300 (since 2017) were used, excluding cloud-contaminated data.

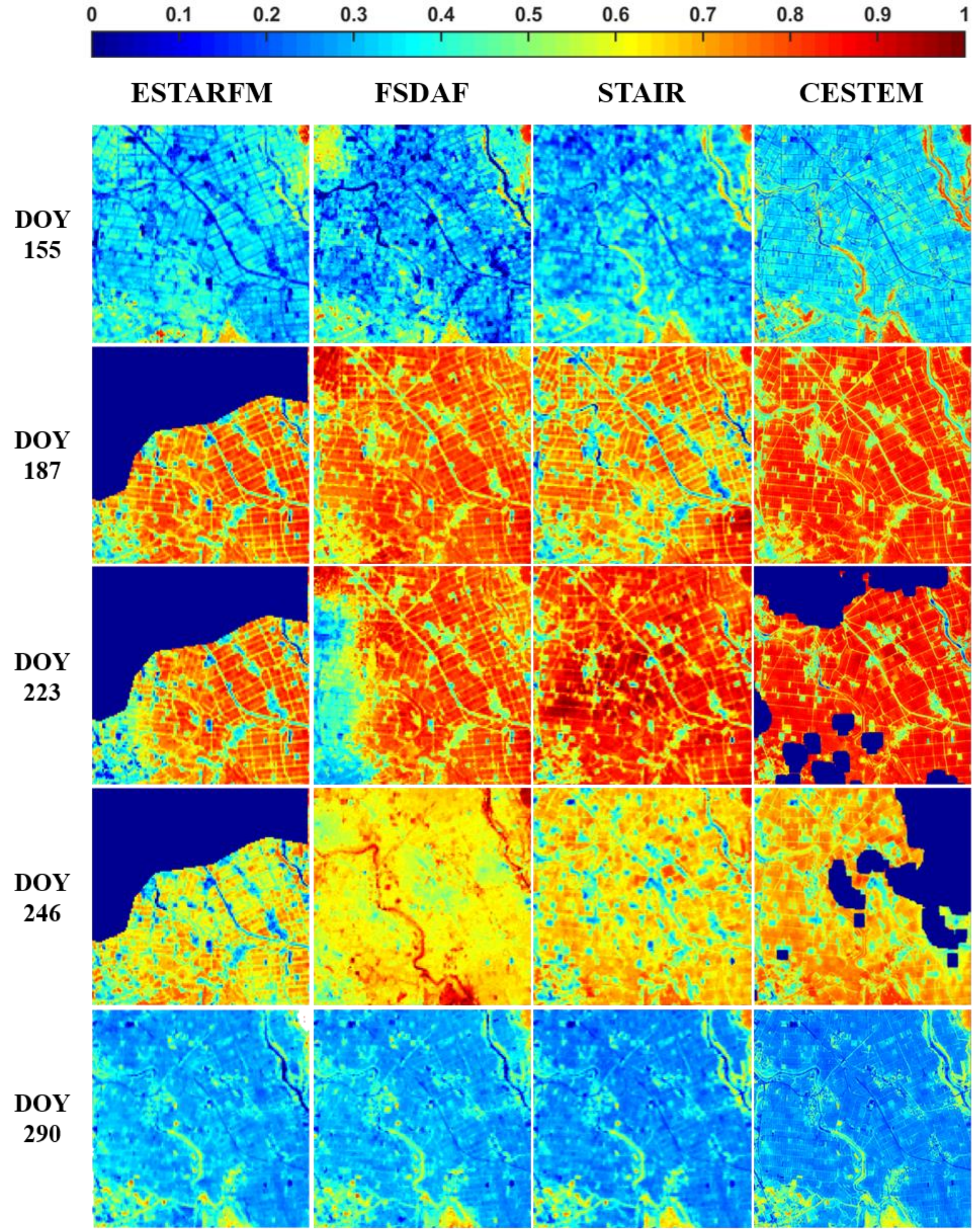
Appendix 4 Evaluation of Landsat 8 and fusion NDVI products against *in situ* NDVI on specific dates. CESTEM on DOY 206 was excluded because of cloud contamination (Table 2). *: no data or cloud contamination within ±3 days.

	DOY	155	187	194	206	215	223	239	248	291
Landsat	R^2	*	*	*	0.22	*	*	*	*	0.17
	rRMSE	*	*	*	11.6 %	*	*	*	*	18.4 %
	rbias	*	*	*	-1.0 %	*	*	*	*	-12.4 %
ESTARFM	R^2	0.65	0.27	0.36	*	0.59	0.08	0	0.01	0.10
	rRMSE	43.8 %	16.4 %	14.3 %	*	6.8 %	16.5 %	26.1 %	25.7 %	15.8 %
	rbias	-14.1 %	-10.3 %	-6.2 %	*	-1.2 %	-13.7 %	-15.7 %	-5.5 %	-7.0 %
FSDAF	R^2	0.01	0.26	0.04	0.24	0.25	0.07	0.01	0.11	0.18
	rRMSE	34.7 %	13.5 %	17.6 %	12.3 %	10.1 %	11.1 %	20.5 %	21.2 %	14.6 %
	rbias	-15 %	-3.8 %	-7.9 %	3.9 %	2.8 %	-5.6 %	-10.1 %	-7.6 %	-5.9 %
STAIR	R^2	0.38	0.39	0.52	0.49	0.87	0.61	0.31	0.25	0.20
	rRMSE	25.7 %	14.4 %	12.4 %	9.6 %	5.8 %	7.7 %	15.2 %	18.1 %	24.7 %
	rbias	-7.2 %	-8.3 %	-3.0 %	0.7 %	3.6 %	4.5 %	-4.6 %	-1.0 %	-20.9 %
CESTEM	R^2	0.57	0.82	*	*	*	0.87	0.89	0.94	0.47
	rRMSE	20.6 %	7.6 %	*	*	*	4.2 %	7.6 %	16.6 %	21.9 %
	rbias	-3.8 %	- 0.9 %	*	*	*	1.7 %	0.7 %	11.5 %	-19.1 %

Appendix 5 Statistical analyses of each plot during the entire growing season. *, mixed land cover plot

	ESTARFM			FSDAF			STAIR			CESTEM		
Plot	R^2	*rRMSE*	*rbias*	R^2	*rRMSE*	*rbias*	R^2	*rRMSE*	*rbias*	R^2	*rRMSE*	*rbias*
1	0.92	15.5%	-10.3%	0.94	11.6%	-8.2%	0.97	9.4%	-7.5%	0.98	7.5%	-4.1 %
2	0.95	12.8%	-7.9%	0.95	10.9%	-7.2%	0.96	7.5%	-2.8%	0.97	8.1%	-3.4 %
* 3	0.66	24.7%	16.8%	0.75	36.3%	29.4%	0.63	23.9%	11.1%	0.94	17.5%	10.9 %
* 4	0.87	17.0%	12.8%	0.93	21.0%	18.5%	0.92	11.7%	7.3%	0.73	20.4%	7.6 %
5	0.55	31.6%	6.4%	0.73	23.3%	5.3%	0.61	29.3%	10.5%	0.86	20.2%	7.3 %
* 6	0.68	21.9%	-12.3%	0.83	13.7%	3.6%	0.93	9.6%	4.8%	0.97	12.4%	4.3 %
7	0.96	13.0%	-11.4%	0.96	11.8%	-10.1%	0.97	8.4%	-7.0%	0.98	6.4%	-3.9 %
8	0.90	18.4%	-15.5%	0.30	17.4%	-14.9%	0.92	12.6%	-9.4%	0.98	10.2%	-8.7 %
9	0.94	10.3%	-6.9%	0.85	12.7%	-4.0%	0.92	10.9%	-0.1%	0.99	6.9%	-3.2 %
10	0.95	11.8%	-9.2%	0.96	11.9%	-10.0%	0.96	8.6%	-4.2%	0.99	6.5%	-4.9 %
11	0.75	32.2%	-28.2%	0.77	23.0%	-18.4%	0.83	19.5%	-15.3%	0.98	7.0%	-4.7 %
12	0.62	26.9%	-23.3%	0.63	16.9%	-10.2%	0.81	12.8%	-5.9%	0.99	4.5%	-2.4 %
13	0.88	15.1%	-13.0%	0.93	10.2%	-7.1%	0.98	8.0%	-2.0%	0.95	8.5%	-1.9 %
14	0.94	9.7%	-6.7%	0.93	10.2%	-6.5%	0.97	5.2%	-1.6%	0.95	8.0%	0.3 %
15	0.99	10.5%	-10.2%	0.95	9.0%	-7.2%	0.96	7.6%	-3.7%	0.96	6.3%	-3.1 %

Supplementary



Supplementary 1 NDVI maps (3. 84 × 3.84 km) of fusion products on *in situ* measurement dates. The study site is located in the center of each map. For ESTARFM and CESTEM, cloud masks are applied (dark navy).

Supplementary 2 Evaluation of resampling methods for aggregating CESTEM 3m products to CESTEM 30 m product.

Interpolation Method	R^2	bias	RMSE
Nearest-neighbor	0.929	0.063	-0.005
Bilinear	0.928	0.063	-0.004
Bicubic	0.928	0.063	-0.003
Nearest-neighbor with antialiasing	0.914	0.069	-0.010
Bilinear with antialiasing	0.907	0.073	-0.010
Bicubic with antialiasing	0.917	0.070	-0.011

Supplementary 3 Day-of-year (DOY) data for the NDVI time series.

Data (Number of Data)	**Day of Year since 2017**
In situ (9)	155 187 194 206 215 223 239 248 291
Landsat-8 OLI (6)	94 134 174 206 286 302
MODIS (73) (MOD09GQ)	92 93 97 100 102 103 106 109 111 113 116 117 120 121 122 123 126 127 131 134 139 140 141 146 147 152 153 154 155 156 159 161 162 163 166 168 169 173 186 187 195 200 206 213 223 243 244 245 246 256 257 259 263 264 268 269 271 272 275 278 280 286 287 290 293 294 296 297 298 300 301 303 304
ESTARFM (65)	92 97 100 102 103 106 109 111 113 116 117 120 121 122 123 127 131 134 139 140 141 146 152 153 154 155 156 159 161 162 163 166 168 169 186 187 195 200 213 223 243 244 245 246 256 257 259 263 264 268 269 271 272 275 278 280 290 293 294 296 297 298 300 303 304
FSDAF (66)	92 97 100 102 103 106 109 111 113 116 117 120 121 122 123 127 131 134 139 140 141 146 152 153 154 155 156 159 161 162 163 166 168 169 186 187 195 200 213 223 243 244 245 246 256 257 259 263 264 268 269 271 272 275 278 280 286 290 293 294 296 297 298 300 303 304
STAIR (214)	91 to 304 (Daily data, No data gap in STAIR)
CESTEM (60)	92 98 101 102 110 111 112 113 114 119 120 122 126 138 139 141 142 144 147 154 155 161 163 167 168 169 171 174 178 185 187 195 206 215 223 224 238 244 247 258 259 261 263 265 268 269 271 275 277 278 287 290 293 294 296 300 301 303 304