International Journal of Applied Earth Observation and Geoinformation 58 (2017) 1–11

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Improving the mapping of crop types in the Midwestern U.S. by fusing Landsat and MODIS satellite data

Likai Zhu[a,*], Volker C. Radeloff[a], Anthony R. Ives[b]

[a] SILVIS Lab, Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA
[b] Department of Zoology, University of Wisconsin-Madison, 250 N. Mills Street, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 6 June 2016
Received in revised form
16 December 2016
Accepted 23 January 2017
Available online 1 February 2017

Keywords:
Crop-type classification
Agricultural systems
STARFM
Landsat
MODIS

ABSTRACT

Mapping crop types is of great importance for assessing agricultural production, land-use patterns, and the environmental effects of agriculture. Indeed, both radiometric and spatial resolution of Landsat's sensors images are optimized for cropland monitoring. However, accurate mapping of crop types requires frequent cloud-free images during the growing season, which are often not available, and this raises the question of whether Landsat data can be combined with data from other satellites. Here, our goal is to evaluate to what degree fusing Landsat with MODIS Nadir Bidirectional Reflectance Distribution Function (BRDF)-Adjusted Reflectance (NBAR) data can improve crop-type classification. Choosing either one or two images from all cloud-free Landsat observations available for the Arlington Agricultural Research Station area in Wisconsin from 2010 to 2014, we generated 87 combinations of images, and used each combination as input into the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) algorithm to predict Landsat-like images at the nominal dates of each 8-day MODIS NBAR product. Both the original Landsat and STARFM-predicted images were then classified with a support vector machine (SVM), and we compared the classification errors of three scenarios: 1) classifying the one or two original Landsat images of each combination only, 2) classifying the one or two original Landsat images plus all STARFM-predicted images, and 3) classifying the one or two original Landsat images together with STARFM-predicted images for key dates. Our results indicated that using two Landsat images as the input of STARFM did not significantly improve the STARFM predictions compared to using only one, and predictions using Landsat images between July and August as input were most accurate. Including all STARFM-predicted images together with the Landsat images significantly increased average classification error by 4% points (from 21% to 25%) compared to using only Landsat images. However, incorporating only STARFM-predicted images for key dates decreased average classification error by 2% points (from 21% to 19%) compared to using only Landsat images. In particular, if only a single Landsat image was available, adding STARFM predictions for key dates significantly decreased the average classification error by 4 percentage points from 30% to 26% ($p < 0.05$). We conclude that adding STARFM-predicted images can be effective for improving crop-type classification when only limited Landsat observations are available, but carefully selecting images from a full set of STARFM predictions is crucial. We developed an approach to identify the optimal subsets of all STARFM predictions, which gives an alternative method of feature selection for future research.

1. Introduction

Land-use activities have transformed a large proportion of the Earth's land surface, changed the structure and functions of ecosystems, and in turn, caused far-reaching consequences for human well-being (Foley et al., 2005). This is why land-change science aims to understand the dynamics of land cover and land use as a coupled human-environment system, and is a key component of global environmental change and sustainability research (Turner et al., 2007). One important dimension of land-change science is to monitor and map land cover, which is a prerequisite to understanding complex global processes such as climate change, biogeochemical cycles, and human-environmental interactions (DeFries et al., 2002; Foley et al., 2003; Tilman et al., 2009; Turner et al., 2007).

Agricultural expansion is among the most significant human alternations of the global environment (Matson et al., 1997). Nearly 40% of the planet's ice-free land surface is being used for agriculture (Ramankutty et al., 2008). Although agricultural expansion

* Corresponding author.
E-mail address: lzhu68@wisc.edu (L. Zhu).

http://dx.doi.org/10.1016/j.jag.2017.01.012
0303-2434/

and intensification have provided crucial service to humanity by meeting food demands, the replacement of natural lands with croplands and pastures undermines ecosystem services such as carbon sequestration, water quality and flow regulation, regional climate and air regulation, and the maintenance of habitats and biodiversity (Defries et al., 2004; Foley et al., 2005; Ramankutty et al., 2008). Crop-type composition and configuration affect biodiversity and the resilience of food webs to environmental change in agricultural systems (Fahrig et al., 2011). Therefore, it is important to map crop types in order to assess the social, economic and environmental effects of agricultural landscape changes and foster sustainable agriculture (Ramankutty et al., 2008).

With the development of air- and space-borne sensors over the past four decades, our ability to observe, monitor, and map croplands at regional and continental scales has dramatically improved (Friedl and Brodley, 1997; Roy et al., 2014). For example, the United States Department of Agriculture (USDA) National Agricultural Statistics Services (NASS) compiles annual Crop Data Layers (CDL) for the conterminous US (Boryan et al., 2011; Han et al., 2012), which have been widely used in a variety of management and research applications, especially for crop-type classification training and validation (Boryan et al., 2011; Yan and Roy, 2014; Zhong et al., 2014). Landsat satellite images are ideal for mapping and monitoring land cover because they have a relatively high spatial resolution of 30 m and the longest temporal record of space-based surface observations at over 40 years (Roy et al., 2014). However, Landsat satellites have only a 16-day repeat cycle, resulting often in a lack of data for applications when at least two cloud-free observations are required within a year (Ju and Roy, 2008). In particular, for crop-type classification, multi-seasonal images are important to discriminate crop types (Wardlow and Egbert, 2008; Zheng et al., 2015). Therefore, when Landsat images are lacking for specific regions, satellite images from different sensors such as the Advanced Wide Field Sensor (AWiFS) are used to generate CDLs, which results in different spatial resolutions for some years (56 m vs. 30 m for Landsat) (Han et al., 2012). This situation is even worse in tropical areas when mapping crop types at broad scales. Therefore, there is a need for an effective approach to generate enough images for crop-type classifications with high temporal resolution and adequate spatial resolution.

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor provides an alternative data source for land-cover classification at broad scales (Friedl et al., 2002, 2010). MODIS data have sufficient temporal frequency to detect unique signatures of crop types and other land-use practices (Wardlow et al., 2007; Wardlow and Egbert, 2008) due to frequent repeat cycles (up to twice daily). However, MODIS's coarse 250–1000 m spatial resolution limits its suitability for monitoring agricultural systems, which are often characterized by small cropland patches and high spatial complexity (Ozdogan and Woodcock, 2006). The ideal dataset would have the spatial resolution of Landsat data and the temporal resolution of MODIS data, raising the question of how to fuse the two datasets.

The Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) is an effective technique to fuse Landsat and MODIS surface reflectance data and to predict surface reflectance at Landsat's spatial resolution and MODIS temporal frequency (Gao et al., 2006). Implementing STARFM requires a "base pair", that is, a pair of Landsat and MODIS images acquired on the same day, and the STARFM algorithm allows for the input of one or two base pairs. STARFM can then predict surface reflectance for each date for which the MODIS NBAR product is available, i.e., every 8 days. STARFM has been successfully applied in various studies including land-cover classification, and the monitoring of forest disturbances and phenology (Hilker et al., 2009b; Walker et al., 2012, 2014). STARFM works particularly well in relatively homogenous landscapes which have spectrally "pure" MODIS pixels, while its performance is lower when predicting surface reflectance in heterogeneous, fine-grained landscapes (Gao et al., 2006). Moreover, the selection of MODIS products, vegetation classes, temporal variance, and the dates of the input base pairs can influence the accuracy of predictions (Emelyanova et al., 2013; Senf et al., 2015; Walker et al., 2012, 2014; Watts et al., 2011). When used for land-cover classifications, STARFM-predicted images can improve accuracy, but the degree of improvement depends on the number of Landsat observations used as input and the image dates of MODIS/Landsat base pairs (Jia et al., 2014a,b,c; Senf et al., 2015; Watts et al., 2011). Using image segmentation can reduce misclassifications due to in-field heterogeneity and spectral variation, and increase the overall classification accuracy of cropland mapping (Li et al., 2015). However, it is still unknown to what degree the inclusion of STARFM-predicted images can improve crop-type classification.

Here, our overarching goal was to evaluate to what degree crop-type classification can be improved by fusing Landsat and MODIS imagery. To achieve this goal, we used the fields from the Arlington Agricultural Research Station, Wisconsin, USA, from 2010 to 2014, for which we had excellent crop maps, as a test case. We first chose either one or two images from all available cloud-free Landsat images from each year, which resulted in 87 combinations. Second, we used each combination of either one or two original Landsat images as the input of STARFM to predict Landsat-like images at all nominal dates of the 8-day MODIS NBAR data. We evaluated STARFM predictions by comparing the STARFM-predicted image with the original Landsat image at the same date, and summarized the accuracies of 87 combinations by band wavelengths, and by the number and dates of base pairs. Third, we classified the original Landsat and/or STARFM-predicted images for each combination under three scenarios: 1) using only the one or two Landsat images of each combination, 2) using the one or two Landsat images plus all STARFM-predicted images, and 3) using the one or two Landsat images plus STARFM-predicted images for key dates. To quantify the differences, we summarized classification errors of all 87 combinations by the number and dates of base pairs.

2. Data and methods

2.1. Study site

Our study site is centered on the Arlington Agricultural Research Station and its surrounding area in southern Wisconsin (Fig. 1), and covers approximately 225 km^2. The climate is continental with some influence from Lakes Michigan and Superior. Mean annual temperatures are about 7.4 °C (2000–2014) with a January minimum of −8.4 °C and a July maximum of 21.2 °C. Mean annual precipitation is about 980 mm (2000–2014) and approximately 65% of annual precipitation occurs between April and September. The soils are fine-silty, mixed, and mesic Typic Argiudolls. The environmental conditions are favorable for agriculture, and cropland is the dominant land use class in the study site.

The Arlington Research Station, owned and managed by the College of Agricultural and Life Sciences of the University of Wisconsin-Madison, covers 8.2 km^2 with 6.9 km^2 of croplands (Fig. 1). Detailed information for each field has been recorded since 2000, including land-use type, plants, planting date, fertilizer use, harvesting date, and crop production. We employed the field data as the main independent data source for the classification accuracy assessment.

2.2. Data

We acquired 27 cloud-free images from Landsat (5, 7, and 8) for five years (2010–2014), all of which were free of clouds. All Land-

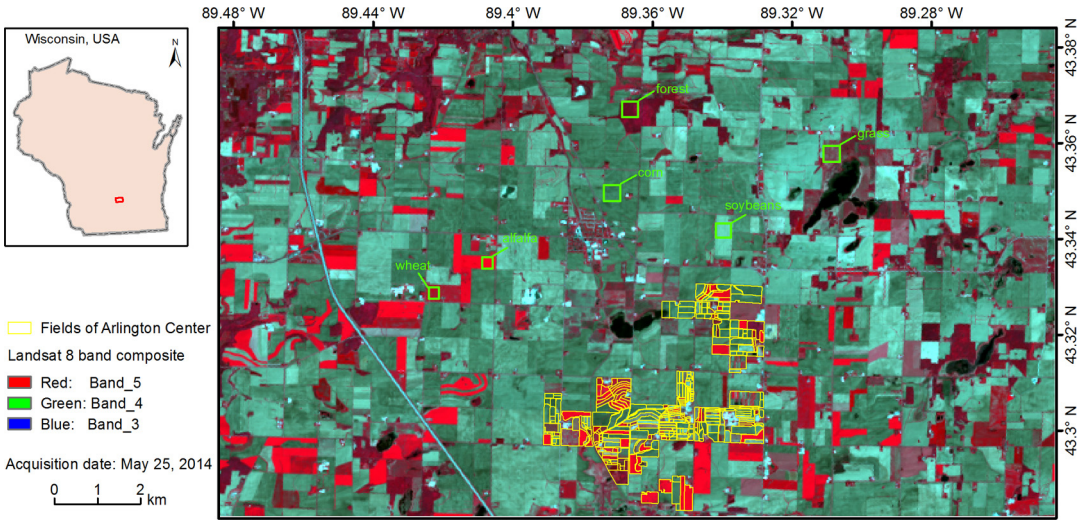

Fig. 1. Location of the study area in southern Wisconsin where land use is dominated by cropland. The false color composite is a Landsat OLI image acquired on May 25, 2014, when the composite color shades are ideal for discriminating land-use classes. The areas highlighted with a yellow boundary indicate the crop fields of the Arlington Agricultural Research Station. Detailed crop information was available for these fields, and used as the independent data source for the accuracy assessment of our crop-type classification. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sat bands were atmospherically corrected to surface reflectance using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Masek et al., 2006). Landsat ETM+ images acquired after 2003 have striping lines due to the SLC failure, but fortunately our study site is relatively unaffected by this because it is in the center of a Landsat path, and striping affects only 1% of the study area. We applied the simple and effective Neighborhood Similar Pixel Interpolator (NSPI) to interpolate the values of pixels for these gaps (Chen et al., 2011). For both the STARFM predictions and the crop type classifications, we used all spectral bands, except the thermal bands of all Landsat sensors and band 1 of Landsat 8.

We acquired the 8-day MODIS NBAR product (MCD43A4) from 2010 to 2014 because this dataset was found previously to be the optimal for STARFM prediction (Walker et al., 2012). The product provides 500-m reflectance data adjusted using BRDF to model the values as if they were taken from nadir view. The pixels contaminated by cloud were masked based on the MODIS quality flags. Following the STARFM input format protocol, all MODIS images were re-projected to the same projection as the Landsat images (UTM Zone 16 N/WGS84) and resampled to 30-m resolution using the nearest-neighbor interpolation.

2.3. Methods

2.3.1. STARFM prediction

STARFM predicts a synthetic image at day t_k with Landsat's 30-m spatial resolution from one or two MODIS/Landsat base pairs acquired on the same day (t_0) and a MODIS image for the prediction day (t_k) (Gao et al., 2006; Hilker et al., 2009b). A moving window is used to minimize the effect of outliers and predict changes of the central pixel using the spatially, spectrally, and temporally weighted mean difference of spectrally similar pixels within the window (Hilker et al., 2009b). Spectrally similar pixels are defined based on the standard deviation of the fine-resolution (Landsat) images and the number of classes specified in the input file. Using a larger number of classes represents a stricter condition for selecting candidate neighbor pixels from fine-resolution images (Gao et al., 2006). A detailed description of the algorithm can be found in Gao et al. (2006) and Hilker et al. (2009b).

We used the combination-based formula, $C(n, r) = n!/(r!\,(n-r)!)$, to calculate the number of combinations that resulted from choosing a sample of r images from a set of n cloud-free Landsat images of each year where order did not matter and replacements were not allowed. Here, r was set as one or two, because STARFM allows only one or two Landsat images as input. For example, five images acquired in 2014 at day-of-year (DOY) 145, 241, 249, 265, and 281 resulted in 15 combinations: (145), (241), (249), (265), (281), (145, 241), (145, 249), (145, 265), (145, 281), (241, 249), (241, 265), (241, 281), (249, 265), (249, 281), and (265, 281) (Fig. 2). In total, we generated 87 combinations with either one or two Landsat images for the five years from 2010 to 2014. We used the individual or paired Landsat images of each combination plus the corresponding 8-day MODIS NBAR images as the input base pairs of STARFM to predict Landsat-like images at the nominal dates of MODIS NBAR images for the time period from DOY 150 (May 29) to 300 (October 26). As an example, we demonstrated the full STARFM predictions using the single original Landsat image at DOY 249 as input base pair (Fig. 2). STARFM images were not produced for dates that coincided with those of the input base pairs, which meant that in our example, the image for DOY 249 was not predicted (Fig. 2). We used a window size of 1500 × 1500 m, consistent with prior studies (Gao et al., 2006; Hilker et al., 2009a,b; Senf et al., 2015; Walker et al., 2012).

2.3.2. Accuracy assessment of the STARFM prediction

We evaluated the STARFM predictions of each combination by comparing the predicted reflectance with the corresponding Landsat reflectance of 2000 random samples on a band-by-band basis. For example, we compared the Landsat-like Band 1 reflectance at DOY 265 predicted by the Landsat image at DOY 249 with the original Band 1 reflectance of Landsat image at the same date (DOY 265). We calculated the prediction accuracy as the relative root mean squared error (RRMSE), defined as the root mean squared difference between predicted reflectance and Landsat reflectance divided by the mean Landsat reflectance of random samples multiplied by 100. We summarized the RRMSEs from all 87 combinations by Landsat bands, and by the number and dates of input base pairs.

2.3.3. Crop-type classifications and accuracy assessment

We classified four crop types and two non-crop types: alfalfa, corn, soybeans, winter wheat, grass, and forest. According to NASS Usual Planting and Harvesting Dates for U.S. Field Crops, alfalfa in our study site is usually harvested at least once per year, and often twice or three times, between May to middle September. The char-

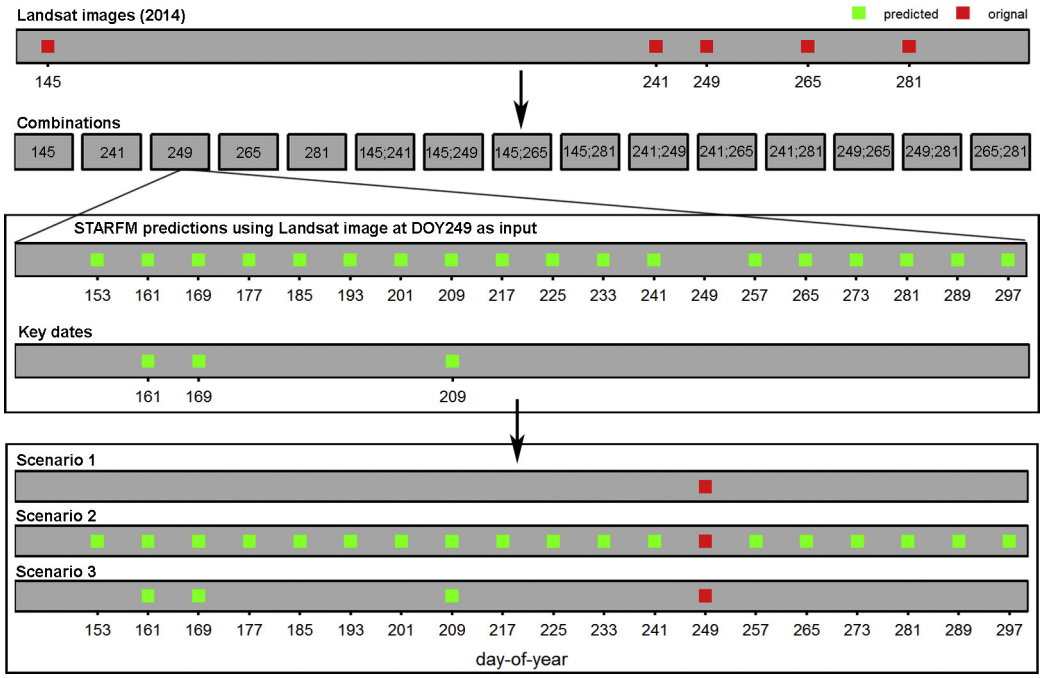

Fig. 2. An example of one combination containing one original Landsat image at DOY 249, 2014 to illustrate our major data processing procedure. Choosing either one or two images from all five available Landsat images in 2014 resulted in 15 combinations. We show here the combination containing the single Landsat image at DOY 249 as an example, which we used as the input of STARFM to predict Landsat-like images at the nominal dates of the 8-day MODIS NBAR product from DOY 150 (May 29) to 300 (October 26). No STARFM image was predicted for DOY249 because that date coincided with the date of input base pair. We classified the original Landsat and STARFM-predicted images under three scenarios: 1) only the original Landsat image at DOY 249, 2) the original Landsat image at DOY 249 plus all eighteen STARFM predictions using the Landsat image at DOY 249 as input base pair, and 3) the original Landsat image at DOY 249 together with STARFM predictions for key dates. In this example, DOY161, 169, and 209 were identified as key dates.

Table 1
Number of training samples used for our crop-type classifications from 2010 to 2014.

	Alfalfa	Corn	Soybeans	Wheat	Grass	Forest	Total
2010	234	258	326	185	256	246	1505
2011	195	293	258	213	264	261	1484
2012	231	232	203	217	170	306	1359
2013	239	239	211	222	219	315	1445
2014	143	204	200	142	196	229	1114

acteristic "growth and cut" cycles of alfalfa are captured by MODIS NDVI time series (Wardlow et al., 2007). Corn and soybeans are planted in May and harvested in October, and share a common phenological cycle, but corn has earlier spring greenup and earlier peak greenness than soybeans (Wardlow et al., 2007). Wheat is planted in September and harvested in July or August of the following year making the phenological cycle of wheat very distinctive from that of other vegetation types. Grass and forest have earlier greenup than corn and soybeans, and unlike alfalfa they often do not show "growth and cut" cycles. Given the differences in phenological cycles, we collected training samples based on color composite Landsat images of different periods and NDVI profiles generated from available Landsat images within a year. Other reference data sources included Google Earth imagery (e.g., GeoEye) and the USDA NASS CDLs. In total, we selected more than 1000 training samples for each year, and ensured that they were spatially well-distributed across the study area but excluded the fields of the Arlington Agricultural Research Station (Fig. 1), because we used these for the accuracy assessment (Table 1).

We used a support vector machine (SVM) algorithm for classification. SVM is superior to parametric classifiers (e.g., maximum likelihood classifier) because it can be used with small training datasets and high-dimensional data, and because it results typically in higher classification accuracies (Huang et al., 2015; Pal and Mather, 2005). We applied the R package 'e1071' (website: https://cran.r-project.org/web/packages/e1071/index.html) to implement the SVM classifier. We chose the Gaussian kernel function (or radial basis function, RBF) to map data into a high-dimensional feature space, and a 3-fold cross validation to select the RBF kernel width parameter γ and the regularization parameter C after comparing the accuracies of different parameter combinations. Given that classification error is often affected by the selection of classification methods, we tested whether using alternative methods (random forests and decision tree) altered the conclusions from those that were based on SVM.

For each combination, we applied SVM to classify the original Landsat and STARFM-predicted images for three scenarios: using only the one or two original Landsat images of one combination (scenario 1), using the one or two Landsat images plus all STARFM-predicted images (scenario 2), and using the one or two original Landsat images plus STARFM-predicted images for key dates (scenario 3). As an example, we show the original and/or predicted images that were used for the classifications under the three scenarios for the single Landsat image at DOY249 (Fig. 2). Regarding accuracy assessment, we used the detailed field data (2010–2014) from the Arlington Agricultural Research Station (Fig. 1), which were independent of the training samples. Randomly distributed points were generated within Arlington fields, and assigned to crop types according to the field information. The number of assessment points for each type and each year was more than 100. We summarized the classification errors of 87 combinations by the number and dates of base pairs, and compared the average classification error among the three scenarios with statistical tests.

2.3.4. Selecting the key dates

We defined the key dates for scenario 3 as the optimal date subset of the full STARFM predictions that resulted in the lowest classification error. Phenological metrics such as the start, end, and

peak of the growing season, which are often retrieved from dense MODIS NDVI/EVI time series, provide a good reference to select Landsat images for land-cover classification (Estel et al., 2015). However, crop phenology is often not synchronized, such as for winter wheat and summer crops. Feature selection and extraction methods, such as the divergence method or principle component analysis, are designed to select the best subset of bands for classification rather than the optimum subset of dates (Jensen, 2005). Here, we used an approach which better fitted our objective of determining the key dates by comparing classification errors of all date subsets, and identifying the best date subset as that with the lowest classification error. For example, we had 18 Landsat-like images at the nominal dates of 8-day MODIS data between DOY 150 and 300 which were predicted by STARFM using the single original Landsat image at DOY249 in 2014 as input (Fig. 2). For the first round, we classified each of the 18 predicted images respectively, and DOY 169 was the one with the lowest classification error (34%). For the second round, we classified the image at DOY 169 together with each of the remaining 17 predicted images, and identified a second image at DOY 161which together with the first image at DOY169 achieved the lowest classification error (32%). For the third round, we classified the images at DOY 169 and 161 together with each of the remaining 16 predicted images, and the third image at DOY 209 was identified which together with another two identified images (DOY169 and 161) reached the lowest classification error (23%). We continued the similar procedure until all 18 predicted images were included in the 18th round. We compared the classification errors of the18 rounds, and selected the subset of dates corresponding to the lowest classification error as the key dates. In our example, there were three key dates: DOY 169, 161, and 209. Accordingly, scenario 3 was defined as the classification of the original Landsat image at DOY 249 plus STARFM-predicted images for the key dates: DOY169, 161, and 209 (Fig. 2). For the other 86 combinations, we followed the same steps to determine the key dates for scenario 3.

3. Results

3.1. Accuracy of STARFM predictions

The predictive performance of STARFM varied considerably among Landsat band wavelengths (Fig. 3). The STARFM predictions for near infrared (NIR) (median RRMSE: 24%) and shortwave infrared 1 (SWIR1) band (26%) performed best, followed by green (30%) and blue (34%) band. The STARFM predictions for shortwave infrared 2 (SWIR2) (37%) and red (40%) band performed worst. Interestingly, whether using one or two base pairs did not substantially affect the accuracy of STARFM prediction. Compared to using one base pair, the average RRMSE of green, red, NIR, SWIR1, and SWIR2 band using two base pairs decreased by 2, 1, 1, 1, and 1% point, respectively, while that of blue band increased by 1% point. Furthermore, the differences in the average RRMSE for all band wavelengths were all not statistically significant.

The image dates of the input base pairs of STARFM greatly influenced the prediction accuracy (Fig. 4). Generally, STARFM predictions based on Landsat images in July and August performed best, followed by using Landsat images after August. STARFM predictions based on Landsat images before July performed worst. Even when using two base pairs from different seasons, incorporating Landsat image before July decreased prediction accuracy. Moreover, the accuracy of STARFM predictions differed with the dates being predicted (Fig. 4). Except for red and SWIR2 band, STARFM performed best when predicting the images for July and August, followed by the predictions after August. STARFM predictions performed worst when predicting the images before July. However, the predictions for red and SWIR2 bands in July and August had the lowest accuracy.

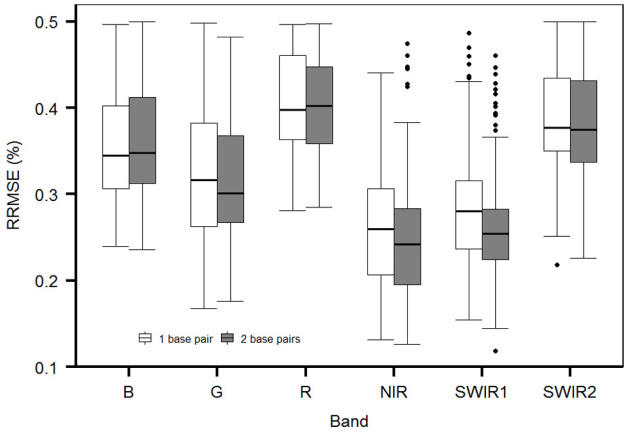

Fig. 3. Relative root mean squared errors (RRMSE) using one or two base pairs summarized for each Landsat band: blue (B), green (G), red (R), near infrared (NIR), shortwave infrared 1 (SWIR1), and shortwave infrared 2 (SWIR2). RRMSE was defined as the root-mean-squared difference between predicted reflectance and original Landsat reflectance at the corresponding date divided by the mean Landsat reflectance of 2000 random samples. The line inside each box denotes the median, and the upper and lower edges of the box indicate 75% and 25% percentiles. The lower whisker was equal to 25% percentile − 1.5 * IQR (inter-quartile range), and the upper whisker was equal to 75% percentile + 1.5 * IQR. For all bands, *t*-tests indicated that there was no statistically significant difference in average RRMSEs when using one base pair versus two.

3.2. Key dates for crop-type classification

We calculated how classification errors changed in response to the number of STARFM-predicted images that were included into classifications for all 87 combinations (Fig. 5). Classification errors initially decreased when more STARFM-predicted images were added to the classification, but reached a minimum fairly early, and then increased gradually. We also calculated the frequency of the key dates for all 87 combinations (Fig. 6). A high frequency indicated that the STARFM-predicted image at this date was identified as key-date image for more times. The STARFM-predicted image at DOY 217 had the highest frequency, followed by the images at DOY 161 and 289. This implied that images at the start, middle and end of the growing season were all essential for classification. Overall, the average frequency of image dates between July and August (26) was higher than that before July (25) and after August (20).

3.3. Classification error comparisons

When classifying Landsat images only, the average classification error when using two Landsat images was significantly lower than that using only a single Landsat image ($p < 0.01$) (Fig. 7). Overall, adding all STARFM-predicted images (scenario 2) increased the average classification error compared to using only the one or two original Landsat images (scenario 1) (Fig. 7). A *t*-test indicated that the average classification error of scenario 2 (25%) was significantly higher than that of scenario 1 (21%) ($p = 0.01$). However, adding the STARFM-predicted images for key dates (scenario 3) did decrease the average classification error compared to using one or two Landsat images only (scenario 1) (Fig. 7). The average classification error of scenario 3 (19%) was lower than that of scenario 1 (21%), but the difference was not statistically significant ($p = 0.15$). The number of Landsat images influenced the degree of classification improvement when adding STARFM-predicted images for key dates. When one Landsat image was available, adding STARFM-predicted images

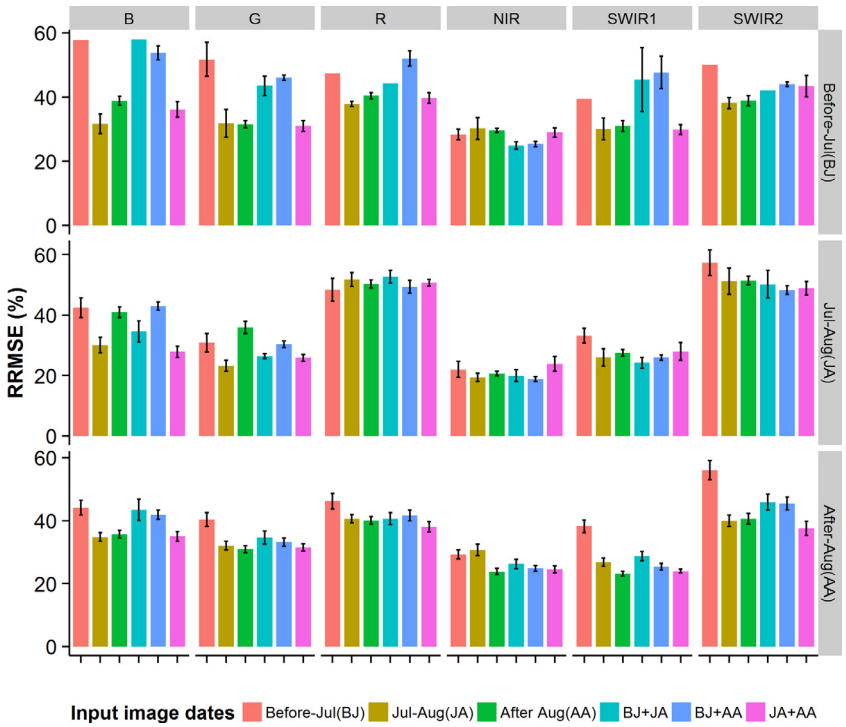

Fig. 4. Relative root mean squared errors of different bands summarized by the dates of the input base pairs and the dates for which images were predicted. Blue: B, green: G, red: R, near infrared: NIR, shortwave infrared 1: SWIR1, and shortwave infrared 2: SWIR2. The six columns of histograms corresponded to six bands, and the three rows correspond to different dates at which images were predicted. The different colors of the bars indicate different dates of input base pairs. Generally, STARFM predictions using base pairs in July and August performed best, while those using base pairs before July had the lowest accuracy. The STARFM-predicted images in July and August often had higher accuracy, while the performance of STARFM to predict images before July was poorer. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

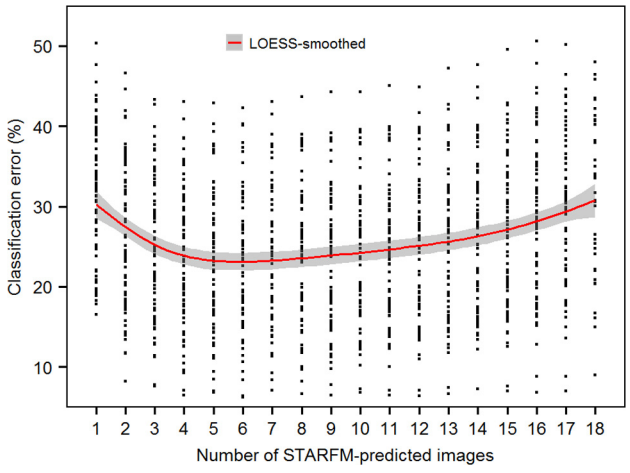

Fig. 5. Classification errors as a function of the number of STARFM-predicted images that were used for the classifications. We used a LOESS curve (local polynomial regression) to fit a line to these points, using an overall degree of the locally-fitted polynomial of two.

for key dates did significantly decrease the average classification error by 4 percentage points from 30% to 26% ($p < 0.05$) compared to using only the single available Landsat image. When two Landsat images were available, adding STARFM-predicted images for key dates decreased the average classification error by only 1% point from 17% to 16%, which was not statistically significant ($p = 0.25$) compared to using the two available Landsat images only.

To test whether classification methods influenced our general conclusions, we also tested forests and decision tree classifiers, conducted the same classifications of the same training data for three scenarios as we did using SVM, and summarized classification errors by the number of base pairs. There was no statistically significant difference between the average classification errors based on SVM and random forests for the three scenarios (scenario 1: $p = 0.53$; scenario 2: $p = 0.33$; scenario 3: $p = 0.77$). However, the average classification error based on SVM was significantly lower than that based on decision trees (scenario 1: $p < 0.01$; scenario 2: $p < 0.01$; scenario 3: $p < 0.01$). For both alternative classification methods, the differences in classification errors among the three scenarios were consistent with those based on SVM (Figs. 7, S1, S2).

Classification errors based on original Landsat images varied among years (2010–2014). The average classification error for 2011 was the lowest (17%), followed by average classification error for 2014 (18%), 2010 (21%), 2012 (22%), and 2013 (26%) (Fig. 8). Interestingly, we found the same order of average classification errors by years when adding all STARFM predictions (scenario 2) or incorporating STARFM predictions for key dates (scenario 3) (Fig. 8). For each year, adding all STARFM predictions increased the average classification error compared to using one or two original Landsat images only. The increase in the average classification error was 4 percentage points in 2010 ($p = 0.17$), 4 percentage points in 2011 ($p = 0.24$), 5 percentage points in 2012 ($p = 0.23$), 3 percentage points in 2013 ($p = 0.28$), and 5 percentage points in 2014 ($p = 0.32$). For all years, adding STARFM predictions for key dates generally reduced the average classification error compared to using one or two Landsat images only. Except for equal values in 2012 (22%), the decrease in the average classification error was 3 percentage points in 2010 ($p = 0.29$), 1% point in 2011 ($p = 0.80$), 3 percentage points in 2013 ($p = 0.23$), and 1% point in 2014 ($p = 0.77$).

When classifying only Landsat images, the acquisition dates of images affected classification errors (Fig. 9). Classifications using Landsat images in July and August achieved the lowest average classification error (22%), followed by classifications using Land-

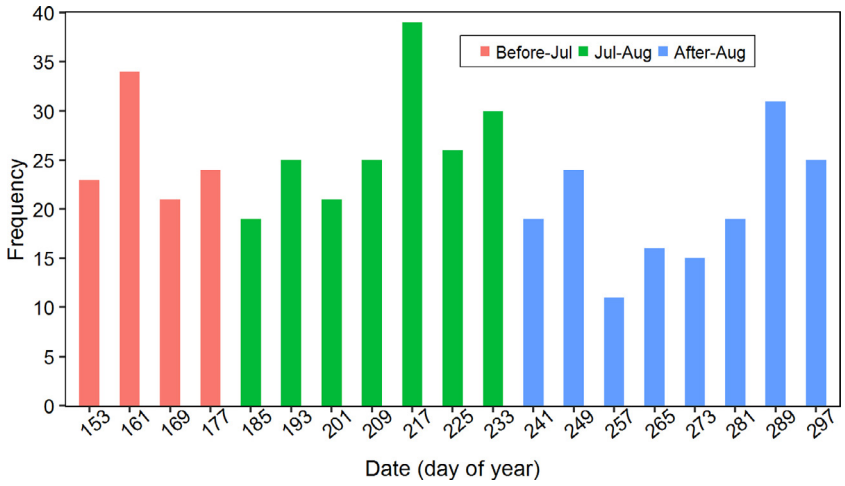

Fig. 6. Frequency at which a given date was selected as of a key date, i. e., the number of times the STARFM-predicted image for this date was identified as a key-date image for crop-type classification among our 87 combinations. The dates with the three highest frequencies were DOY 217, 161, and 289.

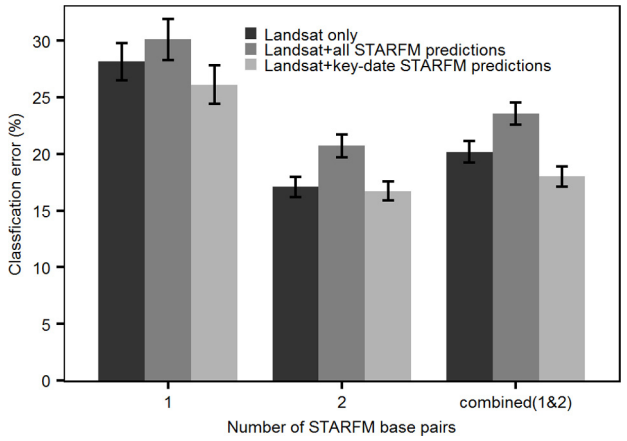

Fig. 7. Comparisons of classification error among three scenarios summarized by the number of STARFM base pairs. Scenario 1: using the one or two original Landsat images only, scenario 2: using the one or two Landsat images and all STARFM predictions with the one or two Landsat images as input base pairs, and scenario 3: using the one or two Landsat images and STARFM predictions for key dates. The group "combined (1 & 2)" indicates the data statistics of group "1" and "2" together. Classifying all STARFM-predicted images and the original Landsat images increased the average classification error compared to classifying the only original Landsat images, whereas adding the STARFM-predicted images for key dates decreased the average classification error. In particular, when only one Landsat image was available, adding STARFM-predicted images for key dates significantly decreased the average classification error by 4 percentage points compared to classifying a single Landsat image only ($p < 0.05$).

sat images after August (25%) and before July (29%). Even when Landsat images were acquired from two different seasons, classifications including images in July or August achieved lower classification errors. The acquisition dates of Landsat images also influenced the degree of classification improvement when adding STARFM-predicted images for key dates (Fig. 9). At any time, adding all STARFM-predicted images increased the average classification error compared to using the one or two original Landsat images only. However, when adding predicted images for key dates, we were able to decrease the average classification error in comparison to using the one or two original Landsat images only between July and August, before July, and after August by 2, 4 and 2% points, respectively. Nonetheless, once two original Landsat images were included, the decrease in classification error by adding further STARFM-predicted images was relatively small (<1% point).

Table 2
Average Kappa index of all 87 combinations for different crop types and scenarios. Scenario 1: using the one or two Landsat images only of each combination, scenario 2: using the one or two Landsat images plus all STARFM predictions using the available one or two Landsat image as base pairs, and scenario 3: using the one or two Landsat images plus STARFM predictions for key dates. For all crop types, adding all STARFM predictions decreased the average Kappa index compared to using only one or two Landsat images, whereas adding STARFM predictions for key dates generally increased the average Kappa index.

	Alfalfa	Corn	Soybeans	Winter wheat	Grassland	Forest
Scenario 1	70	77	76	74	65	85
Scenario 2	65	74	71	73	58	81
Scenario 3	73	79	78	78	66	87

We calculated the average Kappa index of all 87 combinations for different crop types and scenarios (Table 2). For each scenario, the Kappa index for forest was highest, and that for grass was lowest. Our findings for overall classification error were similar to those for crop-specific Kappa indices. For each crop type, adding all STARFM-predicted images always decreased the Kappa index compared to using only original Landsat images. However, adding only STARFM-predicted images for key dates did increase the Kappa index compared to using original Landsat images only. To test the influence of different classification methods on our comparisons, we also applied random forests and decision trees to calculate the average Kappa index of 87 combinations for different crop types and scenarios. The differences in the Kappa indices among the three scenarios based on random forests and decision tree methods were consistent with those based on SVM (Tables S1 and S3). There was no statistically significant difference between the average Kappa index of each crop type based on random forests and SVM (Table S2), whereas the average Kappa index values of most crops based on SVM were significantly higher than those based on decision trees (Table S4).

3.4. Comparing crop-type maps for different scenarios

For visualization purposes, we compared our STARFM-based classifications with the crop-type maps in 2014, which we digitalized from the paper map of the Arlington Agricultural Research Station (Fig. 10 a), and with the reference map with a low classification error of 5% for comparison, which was derived by classifying all five original Landsat images in 2014 (Fig. 10b). The crop-type pattern derived from five Landsat images (Fig. 10b) was highly consistent with the field records from the Arlington Research Station

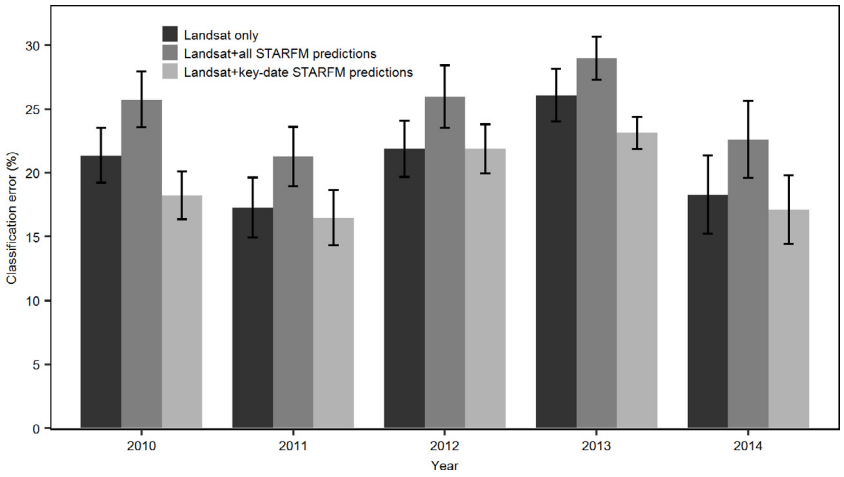

Fig. 8. Comparisons of classification error among three scenarios summarized by years. Scenario 1: using one or two Landsat images only, scenario 2: using the one or two Landsat images plus all STARFM predictions using the one or two Landsat images as input base pairs, and scenario 3: using the one or two Landsat images plus STARFM predictions for key dates.

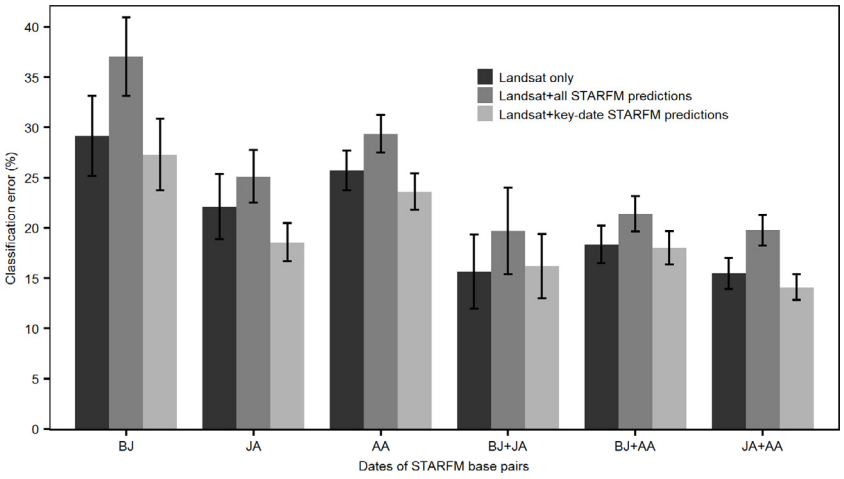

Fig. 9. Comparisons of classification error among three scenarios summarized by the dates of Landsat images which were included as base pairs. The *x* axis indicates the dates for STARFM base pairs. BJ: before July; JA: between July and August; AA: after August. For example, "BJ" denotes that the dates for the one or two base pair inputs of STARFM were before July. Classification errors based on original Landsat images were influenced by their acquisition dates (the first bar of each group). Classifications including Landsat images during July–August typically had lower average classification errors. Classification error based on STARFM predictions were influenced by the dates of input base pairs.

(Fig. 10a). However, when we compared crop-type patterns of the three scenarios with that derived from five Landsat images, there were notable differences. Crop-type classifications resulting from using only one Landsat image (Fig. 10c), using one Landsat image and all STARFM predictions (Fig. 10d), and using one Landsat image plus STARFM predictions for key dates (Fig. 10e) were inconsistent with the 5-image classification in 31%, 32%, and 29% of all pixels. When using Landsat images only, the inconsistent areas usually corresponded to single fields (Fig. 10c), suggesting that there was not enough spectral information in the single image to determine the right crop type. In contrast, the inconsistent areas for the STARFM prediction-based maps occurred for regions spanning several adjacent fields (Fig. 10d, e), most likely because different MODIS pixels had the reflectance of mixed crops, which resulted in the aggregation of surface reflectance of STARFM predictions.

4. Discussion

We found that fusing Landsat with MODIS data was effective for improving crop-type classifications when limited high-resolution Landsat images were available. Classifying Landsat images and STARFM-predicted images for key dates decreased classification error on average compared to classifying either one or two Landsat images only. However, classifying Landsat images and all STARFM predictions increased classification error on average compared to classifying Landsat images only. Our findings suggested that classification error was lowest with only 4–6 STARFM-predicted images for key dates above which classification error increased gradually. Therefore, choosing the optimal subset of STARFM predictions was crucial to avoid data redundancy, and we developed a stepwise approach to select the optimal the incorporation of STARFM predictions.

4.1. Improving STARFM prediction

The STARFM prediction performance showed strong differences among Landsat bands. The STARFM predictions of the NIR and SWIR1 bands performed best, followed by green, blue, and red bands, a result that is consistent with previous findings (Hilker et al., 2009a,b; Watts et al., 2011). Relatively lower prediction accuracy was likely due to higher levels of atmospheric scattering at shorter wavelengths (Roy et al., 2008). However, we also found poor predic-

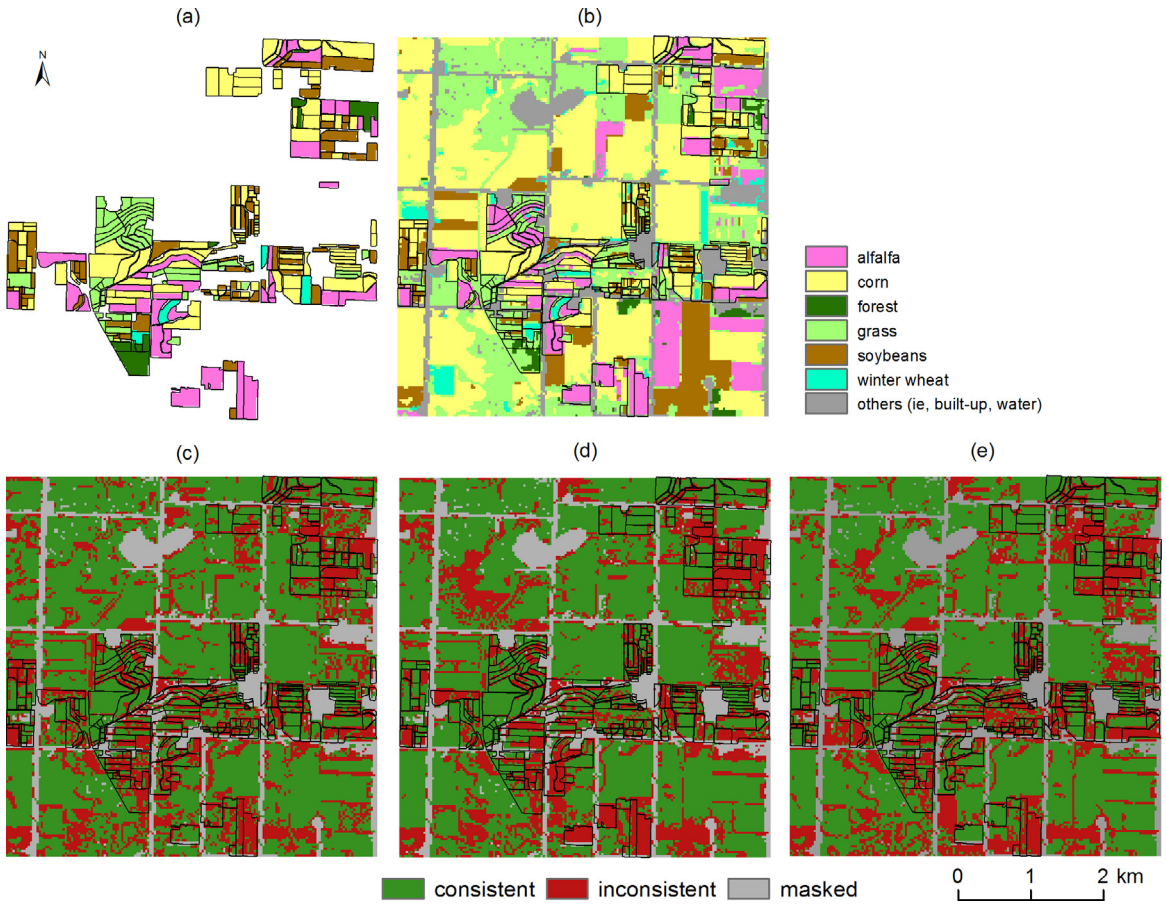

Fig. 10. (a) Crop maps from the Arlington Agricultural Research Station, (b) the crop-type classification based on all five original Landsat images from 2014, and (c–e) maps depicting the consistent and inconsistent areas relative to the crop-type map in (b) with crop-type classifications for (c) the single-date Landsat image for DOY 249 from 2014, (d) Landsat image at DOY 249 plus all STARFM-predicted images, and (e) Landsat image at DOY 249 plus STARFM-predicted images for key dates.

tions for the SWIR2 band. Given that SWIR2 is considered one of the best wavelengths to estimate the fractional cover of vegetation and bare soils (Asner and Heidebrecht, 2002), the poor predictions are most likely due to its high sensitivity to the fraction of vegetation and bare soils. Therefore, we suggest considering both the effects of atmosphere contamination on different bands and the sensitivity of different bands to land surface properties when making STARFM predictions.

Our results also highlighted that having two base pairs instead of one did not significantly improve STARFM predictions, and even increased the prediction errors for blue and red bands. This suggests that a single Landsat image can achieve satisfactory prediction, although the STARFM algorithm allows for the input of two base pairs. The incorporation of a second base pair possibly increases the time expense for processing, and decreases the spatial and temporal weights of more 'suitable pixels' (Gao et al., 2006; Walker et al., 2012). Previous studies also demonstrated a preference for using one base pair to make prediction (Hilker et al., 2009a,b; Senf et al., 2015; Walker et al., 2012; Watts et al., 2011).

Finally, we found that STARFM predictions based on base pairs from the middle of the growing season performed best, and predictions based on base pairs from the start or end of the growing season performed worse. This is because both the spatial and temporal variances are relatively smaller in the middle of growing season than during vegetation greenup or senescence (Emelyanova et al., 2013; Senf et al., 2015; Walker et al., 2012).

4.2. Improving crop-type classification

We compared classification error with statistical tests under different scenarios, and this provides a benchmark when using STARFM to fuse Landsat and MODIS data in order to improve crop-type classifications. Our results suggest that using two Landsat images significantly improved crop-type classification compared to using a single-date image, and images from July and August were most important, similar to prior research in Australia (Van Niel and Mcvicar, 2004) where February and March (the summer months in the southern hemisphere) were identified as the best time of images for crop-type discrimination.

Another key finding of our study was that adding all STARFM-predicted images caused an increase in classification error, while adding STARFM-predicted images for key dates did improve crop-type classification. An increase in classification error when using all STARFM-predicted images was also found by Senf et al. (2015) in complex Mediterranean landscapes. Redundant and temporal auto-correlated information in the continuous spectral time series appears to cause model overfitting and lower generalization power (Pal and Foody, 2010; Senf et al., 2015). Moreover, there are always some errors in STARFM-predicted images (Fig. 3), and including more predicted images tended to increase the chance to incorporate more errors (Gao et al., 2006). Our research clearly demonstrated the relationship between classification error and predicted image number, showing that classification errors did not decrease when more STARFM-predicted images were added, but rather increased after an initially decline (Fig. 5). Therefore,

it is necessary to carefully select predicted images for key dates, rather than use all STARFM-predicted images, to avoid data redundancy and error accumulation. Instead of using statistical metrics of STARFM-predicted images (e.g., maximum, minimum, or standard deviation) (Jia et al., 2014a,b,c; Senf et al., 2015), our study developed a stepwise approach to identify the optimal subset of all STARFM-predicted images that achieved the lowest classification error.

Our findings for the overall classification error were also supported by the crop-specific Kappa indices. Adding all STARFM predictions decreased the average Kappa index of each crop type compared to classifying only one or two Landsat images, whereas adding STARFM predictions for key dates increased the average Kappa index of each crop type. We also found that the average Kappa index of grasslands and alfalfa was lower than that of other crop types. Discriminating between grassland and alfalfa is challenging, because some types of human-managed grasslands, such as pasture, have a similar "growth and cut" cycle as alfalfa. Very frequent satellite observations are required to capture the phenology of within-season harvesting. Including STARFM predictions into classifications could provide more temporal information, but also create more noise due to prediction errors. This is why we did not see significant improvement in the average Kappa index of these two crop types when adding STARFM predictions.

4.3. Implications, limitations and future opportunities

We generated 87 combinations by choosing either one or two images from the 27 available cloud-free Landsat images for our study period. Aggregating results from 87 samples allowed us to make statistical comparisons among conclusions. First, our research tested how reflectance band, the number of base pairs, and the dates of base-pair images influenced the accuracy of STARFM prediction. Our results provide important information which can be used when choosing the best input Landsat images for STARFM in agricultural systems. Second, our research showed how classification errors varied with the number of STARFM-predicted images included in classifications, and we found that classifications had the lowest classification errors when there were 4–6 predicted images. This illustrates how data redundancy can affect classifications, and suggests that selecting images from a full set of STARFM predictions is important. Third, our research tested how fusing Landsat and MODIS with STARFM improved crop-type classification and clarified the influence of the number and dates of base pairs. This provides a benchmark and guidance under which conditions to apply STARFM, and how to use STARFM predictions to improve crop-type classifications.

Our approach can benefit crop-type classification at broader scales, especially in areas where severe cloud contamination limits data availability. Even though more Landsat-like images are available now with new sensors (e.g., Sentinel 2a, Malenovský et al., 2012), there are still data gaps for broad-scale crop-type mapping. Here, we show that fusing limited high-resolution images with MODIS imagery is an effective solution to obtain additional data where Landsat image availability is limited, such as tropical areas which are often subject to severe cloud contamination (Ju and Roy, 2008). Where only one or two Landsat images are available in a growing season, crop-type classification can be improved by fusing Landsat with MODIS data.

The errors underlying STARFM predictions affected classification error when they were added into crop-type classifications. In general, the Index-then-Blend approach (IB, where the indices are calculated first and then these indices become input for the fusing algorithms to predict indices) produces higher accuracies than the Blend-then-Index approach (BI, where reflectance bands are input into the fusing algorithm to predict reflectance bands at other dates, which are then used to calculate indices then) (Jarihani et al., 2014). This suggests that adding STARFM-predicted NDVI time series based on the IB approach (e.g., Landsat reflectance bands plus STARFM-predicted NDVI time series) may improve crop-type classifications, and should be tested in future studies. Similarly, studies have highlighted the advantage of object-based image segmentation in order to reduce noise from STARFM-predicted images (Li et al., 2015). Including the metrics from STARFM-predicted images (i.e., smoothed STARFM-predicted images, key phenological metrics, and band ratios) might reduce prediction error and improve crop-type classification (Alcantara et al., 2012; Estel et al., 2015; Prishchepov et al., 2012). Our research suggests that images from different seasons were all important for crop-type classification. It is worth determining the image dates that are optimal for stabilizing the classification accuracies, and a possible way of doing so would be evaluate how randomly choosing images from different seasons affects classification accuracies.

Although most studies have found that STARFM can fuse MODIS and Landsat data so that the predicted images are in good agreement with actual observations (Hilker et al., 2009b; Senf et al., 2015; Walker et al., 2012, 2014; Watts et al., 2011), some studies found limitations of STARFM (Hilker et al., 2009a; Zhang et al., 2013; Zhu et al., 2016), and others have suggested improvement for the STARFM algorithm (Gevaert and García-haro, 2015; Zhang et al., 2013; Zhu et al., 2016). However, these improvements typically require either two cloud-free base pairs or other auxiliary data, and do not always produce lower error than STARFM (Emelyanova et al., 2013). Moreover, if two or more original Landsat images are available, fusing MODIS/Landsat data to improve crop-type classifications might not be necessary, because, as we showed here, the decrease in classification errors is not significant (Senf et al., 2015).

5. Conclusions

By comparing classification errors of 87 combinations among three scenarios with statistical tests, we evaluated to what degree STARFM-predicted images can improve crop-type classification in agricultural systems. Our study supports prior results and gives new insights into STARFM predictions and their use in crop-type classifications. Our key findings included: 1) regarding STARFM predictions, there was no statistically significant improvement whether one or two base pairs were used, and the predictions using Landsat images in July and August as base pairs performed best; 2) if only Landsat images were used for classifications, both the acquisition dates and number of images were important; 3) adding all STARFM-predicted images decreased the average classification error compared to using only one or two Landsat images; 4) adding STARFM-predicted images for key dates only improved classification accuracy in comparison to using one or two original Landsat images, and in particular, the improvement was only statistically significant compared to classifying a single Landsat image. Given that the mapping of crop types requires frequent cloud-free images during the growing season, which are often not available, we conclude that adding STARFM-predicted images can be effective for crop-type classifications.

Acknowledgements

We gratefully acknowledge support for this work by NSF's Dimensions of Biodiversity program (No. 1240804), and by NASA's Biodiversity and Ecological Forecasting program (No. NNX14AP07G). Two anonymous reviewers provided constructive feedback that greatly improved the manuscript, and we thank them for their suggestions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jag.2017.01.012.

References

Alcantara, C., Kuemmerle, T., Prishchepov, A.V., Radeloff, V.C., 2012. Mapping abandoned agriculture with multi-temporal MODIS satellite data. Remote Sens. Environ. 124, 334–347.

Asner, G.P., Heidebrecht, K., 2002. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: comparing multispectral and hyperspectral observations. Int. J. Remote Sens. 23, 3939–3958.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto Int., 1–18.

Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 115, 1053–1064.

DeFries, R.S., Houghton, R.A., Hansen, M.C., Field, C.B., Skole, D., Townshend, J., 2002. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980 and 1990. Proc. Natl. Acad. Sci. U. S. A. 99, 14256–14261.

Defries, R.S., Foley, J.A., Asner, G.P., 2004. Land-use choices: balancing human needs and ecosystem function. Front. Ecol. Environ. 2, 249–257.

Emelyanova, I.V., Mcvicar, T.R., Van Niel, T.G., Tao, L., 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 Sens. Environ. 133, 193–209.

Estel, S., Kuemmerle, T., Alcántara, C., Levers, C., Prishchepov, A., Hostert, P., 2015. Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series. Remote Sens. Environ. 163, 312–325.

Fahrig, L., Baudry, J., Brotons, L., Burel, F.G., Crist, T.O., Fuller, R.J., Sirami, C., Siriwardena, G.M., Martin, J.L., 2011. Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. Ecol. Lett. 14, 101–112.

Foley, J.A., Costa, M.H., Delire, C., Ramankutty, N., Snyder, P., 2003. Green surprise? How terrestrial ecosystems could affect earth's climate. Front. Ecol. Environ. 1, 38–44.

Foley, J.A., Foley, J.A., Defries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use. Science 309, 570–574.

Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. 61, 399–409.

Friedl, M., McIver, D., Hodges, J.C., Zhang, X., Muchoney, D., Strahler, A., Woodcock, C., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C., 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sens. Environ. 83, 287–302.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.

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 Trans. Geosci. Remote Sens. 44, 2207–2218.

Gevaert, C.M., García-haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sens. Environ. 156, 34–44.

Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: a Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Comput. Electron. Agric. 84, 111–123.

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 Sens. Environ. 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 Sens. Environ. 113, 1988–1999.

Huang, C., Davis, L., Townshend, J.R.G., 2015. An assessment of support vector machines for land cover classification. Int. J. Remote Sens. 23, 725–749.

Jarihani, A.A., Mcvicar, T.R., Niel, T.G., Van Emelyanova, I.V., Callow, J.N., Johansen, K., 2014. Blending Landsat and MODIS data to generate multispectral indices: a comparison of index-then-blend and blend-then-index approaches. Remote Sens. 6, 9213–9238.

Jensen, J.R., 2005. Introductory Digital Image Processing: A Remote Sensing Perspective, 3rd edition. Prentice Hall, Inc., Old Tappan, NJ.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., Wang, X., 2014a. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens. 6, 11518–11532.

Jia, K., Liang, S., Zhang, L., Wei, X., Yao, Y., Xie, X., 2014b. Forest cover classification using Landsat ETM+ data and time series MODIS NDVI data. Int. J. Appl. Earth Obs. Geoinf. 33, 32–38.

Jia, K., Liang, S., Zhang, N., Wei, X., Gu, X., Zhao, X., Yao, Y., 2014c. Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data. ISPRS J. Photogramm. Remote Sens. 93, 49–55.

Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. Remote Sens. Environ. 112, 1196–1211.

Li, Q., Wang, C., Zhang, B., Lu, L., 2015. Object-based crop classification with landsat-MODIS enhanced time-series data. Remote Sens. 7, 16091–16107.

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., Berger, M., 2012. Sentinels for science: potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. Remote Sens. Environ. 120, 91–101.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.K., 2006. A landsat surface reflectance dataset for North America, 1990–2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72.

Matson, P.A., Parton, W.J., Power, A.G., Swift, M.J., 1997. Agricultural intensification and ecosystem properties. Science 277, 504–509.

Ozdogan, M., Woodcock, C.E., 2006. Resolution dependent errors in remote sensing of cultivated areas. Remote Sens. Environ. 103, 203–217.

Pal, M., Foody, G., 2010. Feature selection for classification of hyperspectral data by SVM. IEEE Trans. Geosci. Remote Sens. 53, 1689–1699.

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 26, 1007–1011.

Prishchepov, A.V., Radeloff, V.C., Dubinin, M., Alcantara, C., 2012. The effect of Landsat ETM/ETM+ image acquisition dates on the detection of agricultural land abandonment in Eastern Europe. Remote Sens. Environ. 126, 195–209.

Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Glob. Biogeochem. Cycles 22, http://dx.doi.org/10.1029/2007GB002952 (GB1003).

Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS – Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. Remote Sens. Environ. 112, 3112–3130.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., Mccorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172.

Senf, C., Leitão, P.J., Pflugmacher, D., van de 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 Sens. Environ. 156, 527–536.

Tilman, D., Tilman, D., Fargione, J., Wolff, B., Antonio, C.D., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W.H., Simberloff, D., 2009. Forecasting agriculturally driven global environmental change. Science 292, 281–284.

Turner II, B.L., Lambin, E.F., Reenberg, A., 2007. The emergence of land change science for global environmental change and sustainability. Proc. Natl. Acad. Sci. U. S. A. 104, 20666–20671.

Van Niel, T.G., Mcvicar, T.R., 2004. Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. Comput. Electron. Agric. 45, 91–108.

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 Sens. Environ. 117, 381–393.

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 Sens. Environ. 144, 85–97.

Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great Plains. Remote Sens. Environ. 112, 1096–1116.

Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. Remote Sens. Environ. 108, 290–310.

Watts, J.D., Powell, S.L., Lawrence, R.L., Hilker, T., 2011. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. Remote Sens. Environ. 115, 66–75.

Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. Remote Sens. Environ. 144, 42–64.

Zhang, W., Li, A., Jin, H., Bian, J., Zhang, Z., Lei, G., Qin, Z., Huang, C., 2013. An enhanced spatial and temporal data fusion model for fusing landsat and MODIS surface reflectance to generate high temporal Landsat-like data. Remote Sens. 5, 5346–5368.

Zheng, B., Myint, S.W., Thenkabail, P.S., Aggarwal, R.M., 2015. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. Int. J. Appl. Earth Observ. Geoinf. 34, 103–112.

Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. Remote Sens. Environ. 140, 1–13.

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 Sens. Environ. 172, 165–177.