



## Spatially and temporally complete Landsat reflectance time series modelling: The fill-and-fit approach



Lin Yan<sup>a,\*</sup>, David P. Roy<sup>a,b</sup>

<sup>a</sup> Center for Global Change and Earth Observations, USA

<sup>b</sup> Department of Geography, Environment, & Spatial Sciences, Michigan State University, East Lansing, MI 48824, USA

### ARTICLE INFO

Edited by Marie Weiss

**Keywords:**

Landsat

Time series

Gap filling

Harmonic model fitting

### ABSTRACT

Statistical time series models are increasingly being used to fit medium resolution time series provided by satellite sensors, such as Landsat, for terrestrial monitoring. Cloud and shadows, combined with low satellite repeat cycles, reduce surface observation availability. In addition, only a single year of data can be used where there is high inter-annual variation, for example, over many croplands. These factors reduce the ability to fit time series models and reduce model fitting accuracy. In solution, we propose a novel fill-and-fit (*FF*) approach for application to medium resolution satellite time series. In the 'fill' step, gaps are filled using a recently published algorithm developed to fill large-area gaps in satellite time series using no other satellite data. In the 'fit' step, a linear harmonic model is fitted to the gap-filled time series. The *FF* approach, and the conventional harmonic model fitting without gap filling, termed the *F* approach, are demonstrated using seven months of Landsat-7 and -8 surface reflectance Analysis Ready Data (ARD) over agricultural regions in North Dakota, Minnesota, Michigan, and Kansas. Synthetic model-predicted reflectance for days through the growing season are examined, and assessed quantitatively by comparison with an independent Landsat surface reflectance data set. The six Landsat reflective band root-mean-square difference (RMSD) between the predicted and the independent reflectance, considering millions of pixel observations for each ARD tile, show that the *FF* approach is more accurate than the *F* approach. The mean *FF* RMSD values varied from 0.025 to 0.026 for the four tiles, whereas the mean *F* RMSD values varied from 0.026 to 0.047. These mean *FF* RMSD values are < 0.03 which is comparable to the uncertainty specification for the Landsat 8 OLI TOA reflectance, but greater than the atmospheric correction uncertainty in any Landsat 8 OLI band. The greatest RMSD values were found over the Minnesota tile and occurred due to a long period of missing data early in the growing season, and the smallest RMSD values were found for the Kansas tile because of the high availability of clear surface observations. The *F* approach could not be applied where there were insufficient clear surface observations to fit the model, and where the model was applied, the fitting was often sensitive to issues including gaps in the Landsat time series and the presence of undetected cloud- and shadow-contaminated observations. The *FF* approach could be applied to every ARD tile pixel location and the predicted reflectance was spatially-coherent and natural looking. Examples are shown that illustrate the potential of using *FF* predicted synthetic reflectance time series for land surface monitoring.

### 1. Introduction

Spatially and temporally complete satellite time series are needed for reliable terrestrial monitoring. Landsat data provide the longest environmental satellite record and have been used widely for science and applications (Roy et al., 2014a; Wulder et al., 2019). Detailed examination of the recently released 30 m Landsat Analysis Ready Data (ARD) long-term record (Dwyer et al., 2018) revealed complex patterns of surface observation availability in space and time (Egorov et al.,

2019), due to the geographically variable Landsat overpass coverage (Li and Roy, 2017), variable cloud and shadow obscuration at the time of overpass (Kovalsky and Roy, 2013; Egorov et al., 2019), past Landsat sensor acquisition and health issues (Goward et al., 2017), and the changing relative orientation of the fixed ARD tiles with respect to the Landsat orbit paths (Egorov et al., 2019). For these kinds of reasons, methods to fill gaps in Landsat, and other satellite time series, have been developed. A recent review of gap filling approaches categorized them into alternative similar pixel (ASP) and temporal interpolation

\* Corresponding author.

E-mail address: [yanlin@msu.edu](mailto:yanlin@msu.edu) (L. Yan).

(TI) approaches (Yan and Roy, 2018a). ASP approaches fill missing pixel observations with sensed values derived from one or more alternative pixels selected from non-missing pixels in the same or a temporally close image (Malambo and Heatwole, 2015; Brooks et al., 2018). TI approaches fit statistical time series models to predict reflectance or vegetation index values on a given day. Some TI approaches use Moderate Resolution Imaging Spectroradiometer (MODIS) coarse resolution data to fill Landsat gaps, but they are less reliable when surface changes occur at scales finer than the coarser resolution data (Gao et al., 2006; Roy et al., 2008; Zhu et al., 2010), and cannot be used in the pre-MODIS era. Furthermore, although MODIS provides near-daily coarse resolution data, the data also contain gaps due to clouds and shadows that may not be reliably filled because of temporal variations in reflectance caused by directional illumination and viewing variations over the wide 110° MODIS field of view (Ju et al., 2010).

There is considerable interest in modelling and predicting the evolution of vegetation using medium resolution (i.e., Landsat or Sentinel-2) satellite time series, for example, to understand and monitor anthropogenic, weather and climate effects on vegetation and to quantify biomass production and crop yields (Gitelson et al., 2012; Vicente-Serrano et al., 2013; Azzari et al., 2017; Zhang et al., 2018; Basso and Liu, 2018). A variety of TI approaches have been developed to predict Landsat reflectance or vegetation index values on a daily or other regular temporal basis. For example, linear regression models have been used for multi-annual forest monitoring and crop activity mapping (Kennedy et al., 2010; Schmidt et al., 2016). TI approaches may be applied to multiple years of data by combining different years of data into a single-year period sorted by their annual acquisition dates as if they had been acquired in the same year, or conventionally by applying a model with a fixed annual frequency. However, this is inappropriate if there are changes in the land cover or condition, for example, crop rotations, between years, and model fits are less reliable if there are inter-annual variations, particularly over croplands that have short growing seasons or several growing seasons per year (Carrão et al., 2010; Roy and Yan, 2020; Griffiths et al., 2019; Rufin et al., 2019). Linear regressions are inappropriate for many natural surfaces, and other statistical models, such as harmonic models that are composed of superimposed sequences of sines and cosines, are more suitable (Olsson and Eklundh, 1994; Jakubauskas et al., 2001; Carrão et al., 2010; Zhu and Woodcock, 2014; Roy and Yan, 2020). Higher-order models are needed to fit more complex temporal variations but increase the incidence of spurious oscillations in the model fit, particularly if there are long periods of missing data (Hermance, 2007; Carrão et al., 2010; Brooks et al., 2012). Recent research over cropland Landsat ARD tiles demonstrated that there can be insufficient observations to apply harmonic models reliably, even when growing season data from two Landsat sensors are used together (Roy and Yan, 2020). Abrupt surface changes reduce the reliability of TI approaches. Statistical models can be fitted independently before and after the change, but this requires correct identification of the change date which may be non-trivial, particularly if there are missing data (DeVries et al., 2016; Brown et al., 2019).

Two solutions to improve temporal model fitting over dynamic surfaces with a single year of data present themselves, (1) use more satellite observations with different overpass dates, for example, by combined use of Landsat and Sentinel-2 data (Roy et al., 2019), or, (2) fill the missing data gaps prior to fitting the time series model. The former solution is preferred but it cannot be adopted if the different sensor data are not available, for example, before the launch of the Sentinel-2 satellites. This paper is concerned with the latter solution.

A fill-and-fit (*FF*) approach for reliable land surface time series modelling using a single year or growing season of Landsat data is presented. The *FF* approach follows the above second suggested solution that fills missing data prior to fitting a time series model. It is implemented in two steps. In the first step, gaps in Landsat acquisitions are filled using the spectral-angle-mapper (SAM) based spatio-temporal

similarity (SAMSTS) alternative similar pixel (ASP) gap filling algorithm (Yan and Roy, 2018a). The SAMSTS finds ASPs using spatio-temporal information in the Landsat time series and can fill spatially extensive gaps, and is robust to land cover changes that occur in the time series. In the second step, a harmonic model is fitted independently per-pixel to the gap-filled time series (Roy and Yan, 2020). The *FF* approach does not use coarser spatial resolution data and works with only a single year or growing season of data, meaning that the spatial fidelity of the gap filled data is maintained and the resulting data are well suited for monitoring agriculture or other dynamic land covers. It is demonstrated using Landsat 8 Operational Land Imager (OLI) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) ARD surface reflectance for a growing season over three cloudy and one relatively non-cloudy agricultural ARD tiles. For comparative purposes, the harmonic model is fitted conventionally to the Landsat time series without any gap filling, referred to as the fit (*F*) approach. The performance of the *F* (fit) and *FF* (fill-and-fit) approaches are qualitatively assessed by visual inspection and are quantitatively assessed by comparison with an independent Landsat ARD set not used by the *F* or *FF* approaches.

## 2. Data and study areas

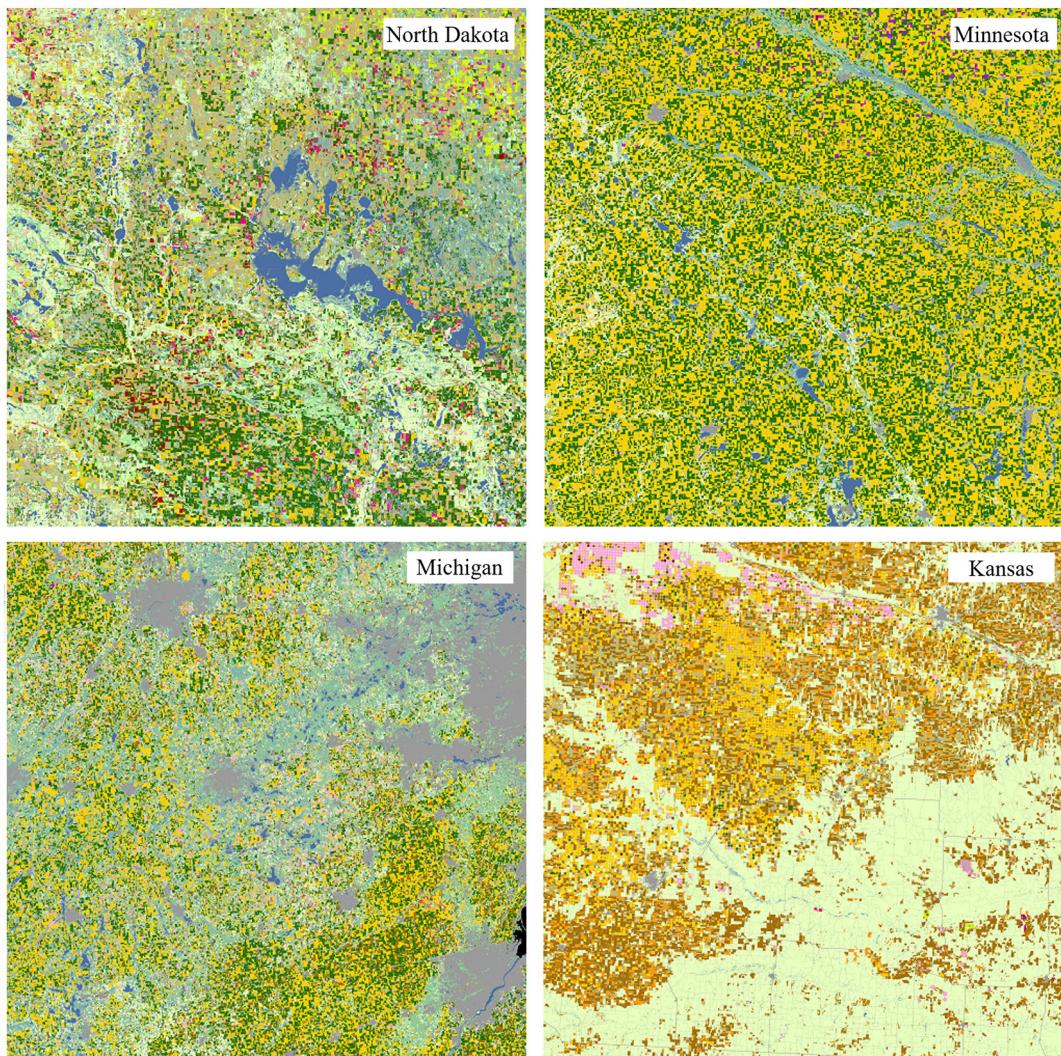
### 2.1. Study areas

Four Landsat ARD tiles in North Dakota (tile 015003), Minnesota (tile 017006), Michigan (tile 023007) and Kansas (tile 014011) were considered (Fig. 1). Each tile is composed of 5000 × 5000 30 m pixels and covers 150 × 150 km, and is defined in the Albers Equal Area projection (Dwyer et al., 2018). The first three tiles are located in areas with high cloud obscuration at the time of Landsat overpass compared to other conterminous United States (CONUS) ARD tiles (Egorov et al., 2019), and so are used to demonstrate the performance of the *FF* method in cloudy regions. The Kansas tile has less cloud obscuration but is also predominantly agricultural and so provides a reference for more moderate CONUS cloud cover conditions.

Fig. 1 illustrates the 2013 Cropland Data Layer (CDL) for the four ARD tiles. The CDL is generated by United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS). The CDL defines about 110 classes at 30 m for all the CONUS in the Albers Equal Area projection (Johnson and Mueller, 2010; Boryan et al., 2011). The main CDL classes in each tile are summarized in Table 1. The four tiles include the major U.S. crops, namely, corn, soybean, spring wheat and winter wheat that have significant seasonal variation (Sacks and Kucharik, 2011; Lark et al., 2017) and other non-crop land covers, including grass/pasture, wetlands, forest, water and developed land.

### 2.2. Landsat ARD

All the Landsat 7 ETM+ and Landsat 8 OLI ARD for a 7-month growing season from April 1st to October 31st 2013 over each tile were used. Fig. 2 shows the percentage of valid pixel Landsat 7 ETM+ and Landsat 8 OLI observations for each ARD tile acquisition over the 7 months. In 2013, Landsat 8 OLI CONUS growing season data were available after the February 2013 Landsat 8 launch, and provided, with the Landsat 7 ETM+, a nominal 8-day overpass frequency (Roy et al., 2014a). At CONUS latitudes, these two Landsat sensors never overpass the same location on the same day and so there is a maximum of one ARD acquisition per day (Roy et al., 2016a). Note that ARD tiles are not produced on days when the tile contains no sensor observations (Dwyer et al., 2018). The differences among days are evident in Fig. 2, and reflect primarily variations in cloud and the variable spatial coverage of the Landsat 7 and 8 swaths over each tile (Egorov et al., 2019). Over the 7 months, the average number of valid observations per 30 m ARD pixel location was 12.2 (Michigan), 14.8 (North Dakota), 16.8 (Minnesota), and 25.1 (Kansas).



**Fig. 1.** The four study ARD tiles,  $5000 \times 5000$  30 m pixels, showing the 2013 Cropland Data Layer (CDL) 30 m data colored in the same way as the CDL web site (<https://nassgeodata.gmu.edu/CropScape/>). The majority CDL crop classes are Corn (yellow), Soybean (green), Winter Wheat (brown), Spring Wheat (wheat yellow), Sorghum (orange), and Alfalfa (pink), non-crop classes include Developed/Open Space (gray), Grassland/Pasture (cream), Deciduous Forest (lime green), Herbaceous Wetlands (light blue), and Open Water (dark blue). The four ARD CONUS tile locations are shown in Roy and Yan, 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The Landsat ARD are described in detail in Dwyer et al. (2018). In this study, the atmospherically corrected ARD, i.e., the surface reflectance, for the reflective wavelength bands common to both sensors i.e., the blue, green, red, Near-Infrared (NIR) and two Short Wave Infrared (SWIR), were used. The Landsat 7 ETM+ surface reflectance for each band was adjusted to be more consistent with the Landsat 8 OLI surface reflectance using the band-specific surface reflectance linear regression parameters provided for this purpose in Roy et al. (2016b). All ARD pixel observations labelled as cloudy, snow, shadow, or saturated in any of the reflective wavelength bands, were discarded. Valid observations are referred to as those that were sensed by a Landsat sensor and were not labelled as cloudy, snow, shadow, or saturated.

### 3. Methods

The fit (*F*) approach is implemented at each ARD tile pixel location by fitting a harmonic model to the Landsat surface reflectance time series. The fill-and-fit (*FF*) approach is implemented in two steps, first gaps in the Landsat surface reflectance time series are filled, and then the harmonic model is fitted to the gap-filled surface reflectance time series. For both approaches, the resulting harmonic model is used to

predict the surface reflectance for each day. To evaluate the *F* and *FF* approaches, the ARD data are divided into two portions. One portion is used to apply the *F* and the *FF* approaches (referred to as the training data) and the other is used to evaluate the results (referred to as the test data). The differences between the *F* or *FF* predicted reflectance and the test data reflectance are quantified.

#### 3.1. Filling algorithm

There are many gap filling algorithms. Alternative similar pixel (ASP) algorithms fill missing pixel observations in an image using one or multiple alternative pixels selected from non-missing pixels in the same or temporally close image, whereby the selections are based on a similarity measure. ASP algorithms typically make no assumptions concerning land surface change trends and so are usually appropriate for gap filling images that are acquired in periods of surface change. In this study, the spectral-angle-mapper (SAM) based spatio-temporal similarity (SAMSTS) ASP gap filling algorithm is used because it can fill spatially extensive gaps and is robust to land cover changes that occur in the time series (Yan and Roy, 2018a). In the SAMSTS algorithm, ASP locations are identified by comparing the multispectral reflectance time

**Table 1**  
Percentage of crop and non-crop land cover classes covering the four ARD tiles derived from the 2013 CDL. Only CDL classes covering > 2% in each tile are summarized.

	North Dakota	Michigan	Minnesota	Kansas
Crop CDL class	Spring Wheat (20.3%), Soybeans (13.5%), Canola (5.2%), Corn (3.2%), Dry Beans (2.9%)	Corn (44.7%), Soybeans (31.9%)	Corn (16.9%), Soybeans (15.4%), Winter Wheat (4.6%), Alfalfa (3.3%)	Winter Wheat (18.7%), Corn (8.4%), Sorghum (7.3%)
Non-crop CDL class	Grassland/Pasture (19.4%), Herbaceous Wetlands (13.4%), Fallow/Idle Cropland (8.2%), Open Water (6.4%), Non-alfalfa Hay (4.0%), Developed, Open Space (3.9%)	Deciduous Forest (6.6%), Herbaceous Wetlands (5.0%), Developed, Open Space (4.9%)	Deciduous Forest (14.3%), Grassland/Pasture (13.3%), Woody Wetlands (10.6%), Developed, Open Space (8.4%), Developed, Low Intensity (6.2%), Developed, Medium Intensity (2.4%)	Grassland/Pasture (49.0%), Fallow/Idle Cropland (8.4%), Developed, Open Space (3.7%)

series of each gap pixel location with non-gap pixel locations using a version of the SAM metric revised to be adaptive to missing temporal observations. Rather than exhaustively compare all pixels, which is computationally expensive, the image time series are first segmented by merging spatially-adjacent similar pixels in space and time, and the segments are subsequently grouped into different clusters. Gap pixel locations are only compared with other pixels that belong to the same cluster to identify ASPs. In this way, the comparisons are undertaken efficiently for the whole image rather than constrained to a spatial neighborhood like most ASP gap filling algorithms (Shen et al., 2015). Once an ASP is found, the gap is filled with the Landsat band reflectance values of the ASP. The SAMSTS was demonstrated over agricultural landscapes and provided gap-filled data with five-band Landsat reflective-wavelength root-mean-square differences < 0.02 and significantly smaller errors than linear harmonic model fitting applied to the same data (Yan and Roy, 2018a). The SAMSTS source code is available publically (Yan and Roy, 2018b).

### 3.2. Fitting algorithm

Linear harmonic models have been used to model the temporal variation of satellite normalized difference vegetation index (NDVI) data for several decades (Menenti et al., 1993; Olsson and Eklundh, 1994). Non-linear harmonic models have also been developed to model non-symmetric waveforms without requiring a large number of model parameters (Carrão et al., 2010). Recently, we applied 5- and 7-parameter linear and non-linear harmonic models to a growing season of Landsat 5 Thematic Mapper (TM) and Landsat 7 ETM + NDVI data over different CONUS cropland ARD tiles (Roy and Yan, 2020). We found that if there were between 15 and 20 Landsat valid observations of a pixel, a 5-parameter non-linear harmonic model could reliably fit the NDVI crop time series, and if there were  $\geq 21$  valid observations, then either the 5-parameter non-linear or the 7-parameter linear model was reliable. However, the 7-parameter linear model had the smallest model fitting errors compared with the other models when there were  $\geq 21$  valid observations. Therefore, in this study, the 7-parameter linear harmonic model was used. In addition, we expect that other researchers will be more easily able to reproduce this research due to the simplicity of the linear harmonic model:

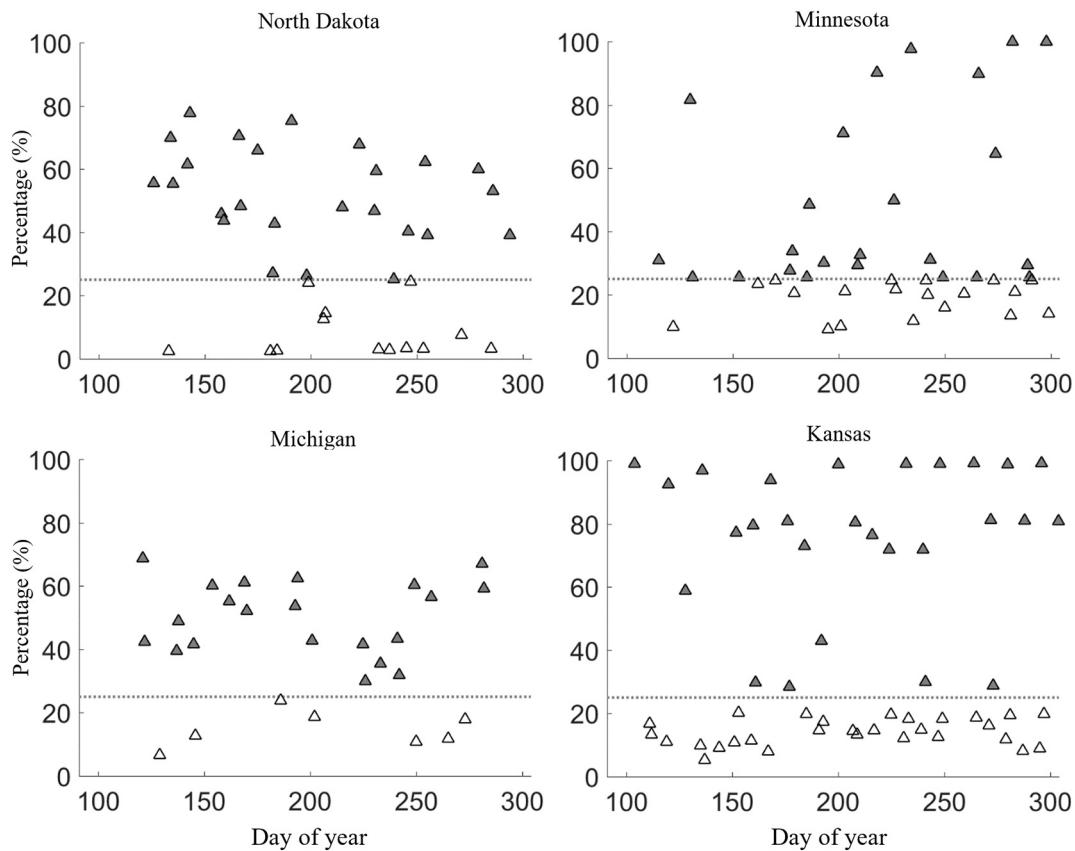
$$\widehat{\rho}_\lambda(t) = a_{o,\lambda} + \sum_{m=1}^M \left( a_{m,\lambda} \cos \frac{2\pi t}{L} + b_{m,\lambda} \sin \frac{2\pi t}{L} \right) \quad (1)$$

where  $\widehat{\rho}_\lambda(t)$  is the model-predicted surface reflectance for spectral band  $\lambda$ , at a single ARD 30 m pixel location for day  $t$ ;  $a_{o,\lambda}$  is the Fourier coefficient describing the mean of  $\widehat{\rho}_\lambda(t)$  over the time series;  $a_{m,\lambda}$  and  $b_{m,\lambda}$  are Fourier coefficients for harmonic component  $m$ ;  $M$  is set as 3 to provide a 7-parameter linear harmonic model; and  $L$  is set as 214 as there were this many days in the 7-month study period (days 91 to 304 of 2013). The model was solved by least-squares adjustment. First the model parameters were initialized with an initial guess (all parameters were set to zero except  $a_{o,\lambda}$  was set to the median of the valid surface reflectance time series values for the spectral band  $\lambda$ ), and then solved iteratively by least-squares adjustment of the differences between the observed times series values and model-predicted values (Ghilani, 2010; Roy and Yan, 2020). The model cannot be solved by least-squares adjustment when there are < 8 valid observations and generally requires 21 or more valid observations for reliable fitting (Zhu and Woodcock, 2014; Roy and Yan, 2020).

### 3.3. Experiments

#### 3.3.1. Training and test data

The ARD data for each tile were divided into two portions. One portion was used to apply the gap filling and harmonic model, referred to as the training data, and the other portion was used to evaluate the



**Fig. 2.** The percentage of valid Landsat 7 ETM+ and Landsat 8 OLI observations across each ARD tile for different days in the 2013 growing season from April 1st (day 91) to October 31st (day 304). The horizontal dashed line shows the 25% level. The training (shaded) and test (unshaded) ARD tile acquisitions have percentages  $\geq 25\%$  and  $< 25\%$ , respectively.

results, referred to as the test data.

The training data were defined by the Landsat ARD tile acquisitions containing  $\geq 25\%$  valid observations (shaded triangles in Fig. 2). A 25% threshold was used because it ensured that for the four ARD tiles, there were at least 21 days that contained data, recall that 21 is the minimum recommended number of observations to fit the 7-parameter linear harmonic model reliably. The Michigan tile is the cloudiest, and over the 7 months there were 21 ARD tile acquisitions containing  $\geq 25\%$  valid observations. There were 27, 25 and 24 for the Kansas, North Dakota, and Minnesota tiles, respectively.

The test data were defined by the Landsat ARD tile acquisitions containing  $< 25\%$  valid observations (unshaded triangles in Fig. 2), and thus the test data were independent of the training data. The test data were in addition aggressively filtered to minimize the incidence of undetected cloud or shadow pixels. The ARD cloud and shadow masks are known to be imperfect (Egorov et al., 2019; Qiu et al., 2019), and so the cloud and shadow masks in each test ARD tile acquisition were dilated (increased in size) by 20 pixels and ARD pixel observations under the dilated masks were removed. In addition, if any test ARD tile acquisition contained  $< 2\%$  valid observations, then all the tile acquisition data for that day were discarded as the data typically had a very small proportion of valid data after the cloud and shadow mask dilations.

### 3.3.2. Fit (F) experiment

For the F approach, the 7-parameter linear harmonic model was fitted to the training data (i.e., only to the ARD acquisitions containing  $\geq 25\%$  valid observations) at each tile pixel location where there were  $\geq 8$  valid observations over the 7 months. These pixel locations are shown colored in Fig. 3. The Landsat orbit paths are quite apparent, so is the high degree of cloudiness in all the tiles except the Kansas tile.

The black pixels show locations where there were  $< 8$  valid observations over the 7 months, i.e., where the F approach could not be implemented, and account for 12.0% (North Dakota), 5.0% (Minnesota), 26.5% (Michigan) and 0.4% (Kansas) of the  $5000 \times 5000$  30 m pixels in each ARD tile.

### 3.3.3. Fill-and-Fit (FF) experiment

For the FF approach, the training data were first gap filled using the SAMSTS algorithm, i.e., any missing pixel values in each ARD tile acquisition containing  $\geq 25\%$  valid observations were gap filled. The 7-parameter linear harmonic model was then fitted at each ARD pixel location where there were  $\geq 8$  valid and gap filled observations. This meant that the FF approach was applied to more ARD pixel locations than in the F approach. In fact, as shown later, after the gap filling, the 7-parameter linear harmonic model could be fitted at every ARD tile pixel location for the four ARD tiles.

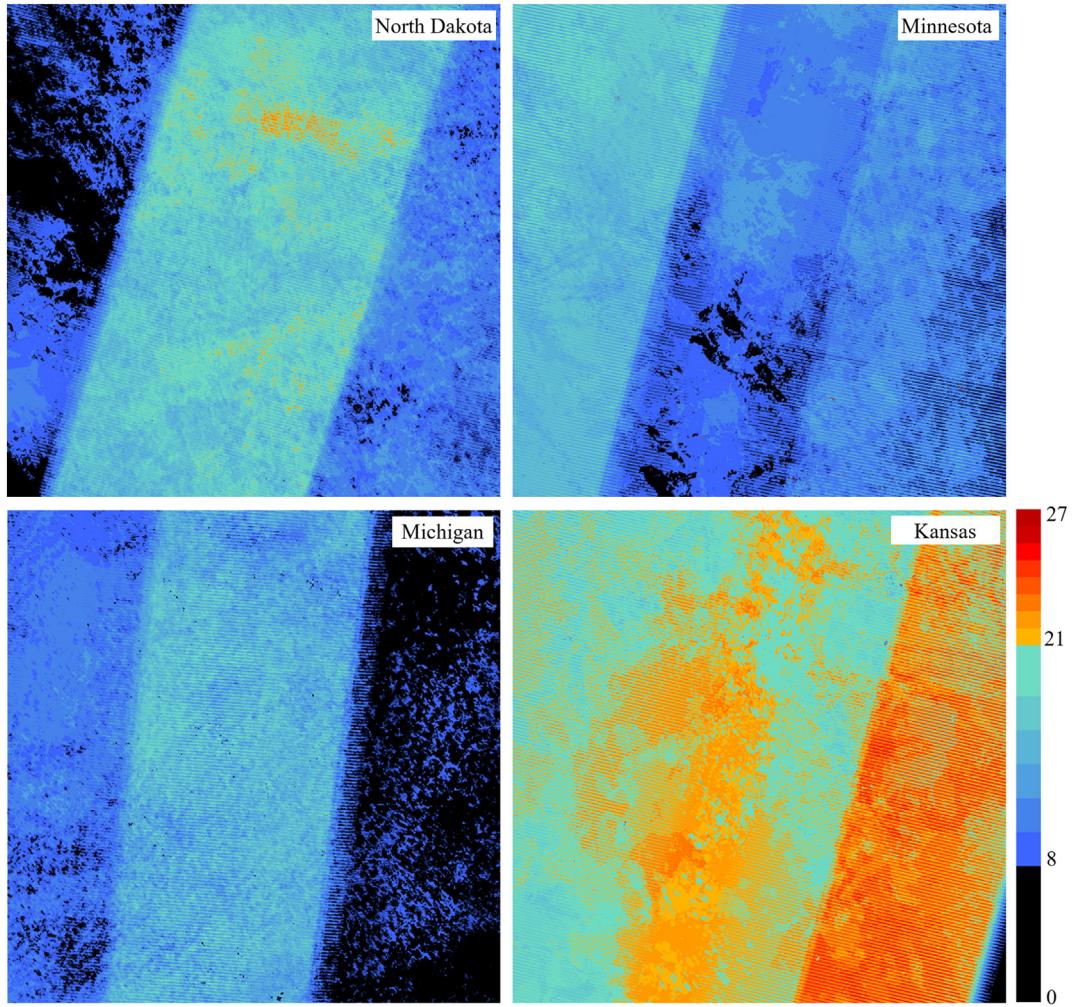
### 3.4. Quantitative validation

To validate the results, the difference between the test data and the contemporaneous F or FF predicted reflectance was derived as:

$$\text{RMSD}(i, j, t) = \sqrt{\frac{\sum_{\lambda=1}^6 (\hat{\rho}_\lambda(i, j, t) - \rho_\lambda(i, j, t))^2}{6}} \quad (2)$$

where  $\text{RMSD}(i, j, t)$  is the six-band (blue, green, red, NIR, and two SWIR) root-mean-square difference,  $\hat{\rho}_\lambda(i, j, t)$  is the predicted surface reflectance derived as Eq. (1) using the seven fitted linear harmonic model parameters found by the F or FF approach, and  $\rho_\lambda(i, j, t)$  is the test Landsat surface reflectance at ARD pixel location  $(i, j)$  on day  $t$ .

The RMSD can only be defined at pixel locations and days where



**Fig. 3.** Training data number of valid (cloud-free, snow-free, shadow-free, unsaturated) Landsat-7 and -8 observations at each 30 m pixel location over the 7-month growing season, for the four ARD tiles (Fig. 1). Training data are defined only by the ARD tile acquisitions with  $\geq 25\%$  valid observations (shaded triangles in Fig. 2). ARD pixel locations with  $< 8$  valid observations are shown in black because at these locations the 7-parameter linear harmonic model cannot be fitted.

there is a valid test ARD value  $\rho_\lambda(i,j,t)$  and if the predicted surface reflectance  $\hat{\rho}_\lambda(i,j,t)$  is defined. In this study, in order to compare meaningfully the *F* and *FF* results, the RMSD values were further only derived for the same pixel locations and dates, i.e., only at pixel locations with  $\geq 8$  valid training data observations over the 7 months (i.e., at the colored pixel locations illustrated in Fig. 3), and where and when there was a valid test  $\rho_\lambda(i,j,t)$  observation.

Summary statistics of the RMSD values were derived to provide accuracy information at three different levels of detail, and are described below. The mean RMSD for each test ARD tile acquisition was derived as:

$$\overline{RMSD}(t) = \frac{\sum_{(i,j)}^{n(t)} RMSD(i,j,t)}{n(t)} \quad (3)$$

where  $\overline{RMSD}(t)$  is the mean of the RMSD (Eq. (2)) values for the test ARD tile acquisition on day  $t$ , and  $n(t)$  is the number of pixels considered across the tile.

The overall mean RMSD considering all the test ARD tile acquisitions was derived as:

$$\overline{RMSD}_{all} = \frac{\sum_{t=1}^N \left( \sum_{(i,j)}^{n(t)} RMSD(i,j,t) \right)}{n_{all}} \quad (4)$$

where  $\overline{RMSD}_{all}$  is the overall mean RMSD,  $N$  is the number of test ARD tile acquisitions,  $n(t)$  is the number of pixels considered across the day  $t$

tile acquisition, and  $n_{all} = \sum_{t=1}^N n(t)$ . Similarly, the overall RMSD standard deviation considering all the test ARD tile acquisitions was derived as:

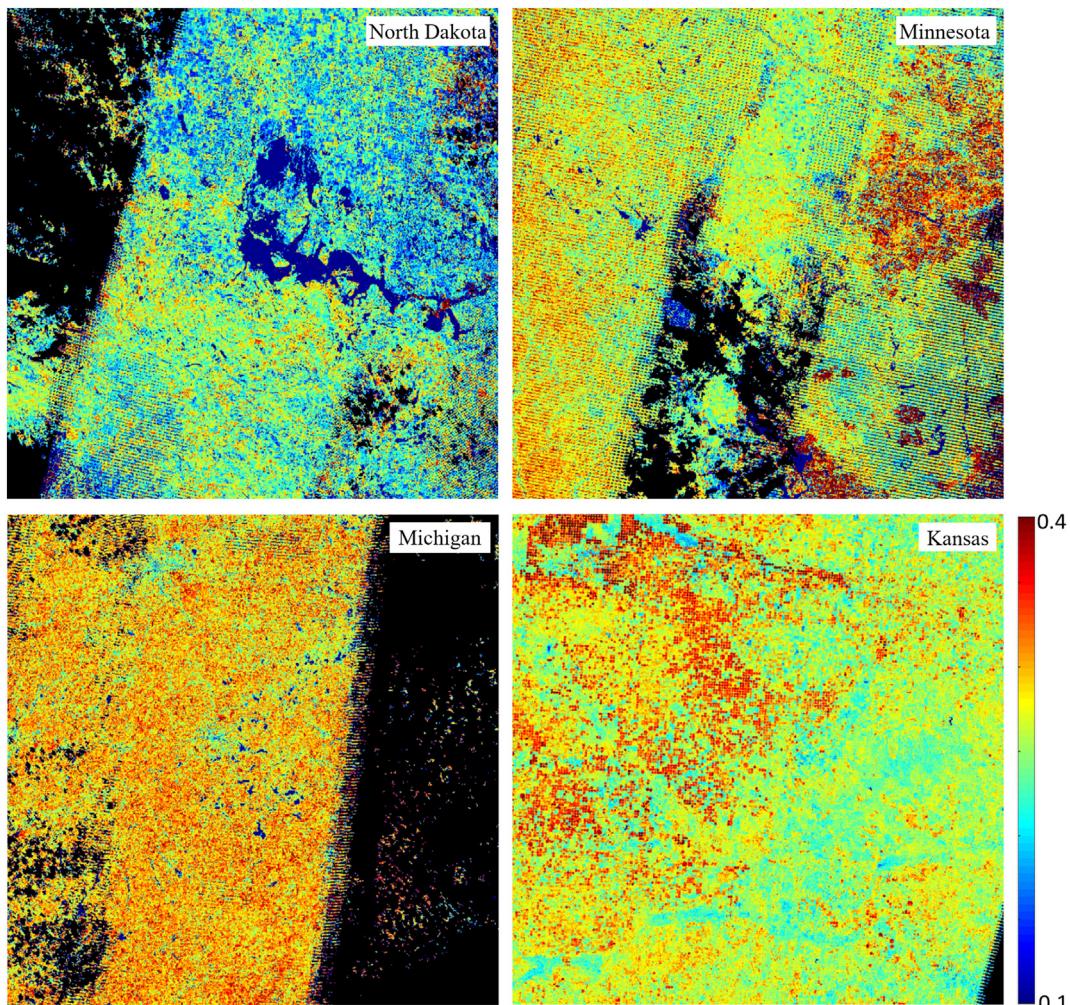
$$\sigma_{all} = \sqrt{\frac{\sum_{t=1}^N \left( \sum_{(i,j)}^{n(t)} [RMSD(i,j,t) - \overline{RMSD}_{all}]^2 \right)}{n_{all}}} \quad (5)$$

where  $\sigma_{all}$  is the overall standard deviation RMSD and the other variables are defined as for Eq. (4).

The dependency of the mean RMSD on the number of valid observations used for the *F* or *FF* approaches was also investigated as:

$$\overline{RMSD}(nobs) = \frac{\sum_{t=1}^N \left( \sum_{(i,j)}^{n(nobs)} RMSD(i,j,t) \Big|_{invalid(ij)=nobs} \right)}{\sum_{t=1}^N n(t, obs)} \quad (6)$$

where  $\overline{RMSD}(nobs)$  is the mean RMSD considering all the test ARD tile acquisitions but only at tile pixel locations that have exactly  $nobs$  valid training data observations over the 7 months (i.e. the count values illustrated in Fig. 3),  $nobs$  was set to vary from 8, 9, 10, ..., to 25 (North Dakota), 24 (Minnesota), 21 (Michigan) and 27 (Kansas) (the maximum unique values for each tile in Fig. 3); and  $n(t, obs)$  is the number of pixels considered for the tile on day  $t$ . To avoid unreliable values,  $\overline{RMSD}(nobs)$  was only derived if there were at least 1000 pixel observations for the  $nobs$  value, i.e., if  $\sum_{t=1}^N n(t, obs) \geq 1000$ .



**Fig. 4.** Linear harmonic model  $a_{0,NIR}$  parameter retrieved at each 30 m ARD pixel location derived by fitting Eq. (1) to the training data i.e., *F* model parameter results. Black shows the 30 m ARD tile pixel locations where there were insufficient valid training data observations ( $< 8$ ) to fit the 7-parameter linear harmonic model. Each tile is  $5000 \times 5000$  30 m pixels.

## 4. Results

### 4.1. Illustrative ARD tile results

#### 4.1.1. Linear harmonic model parameter examples

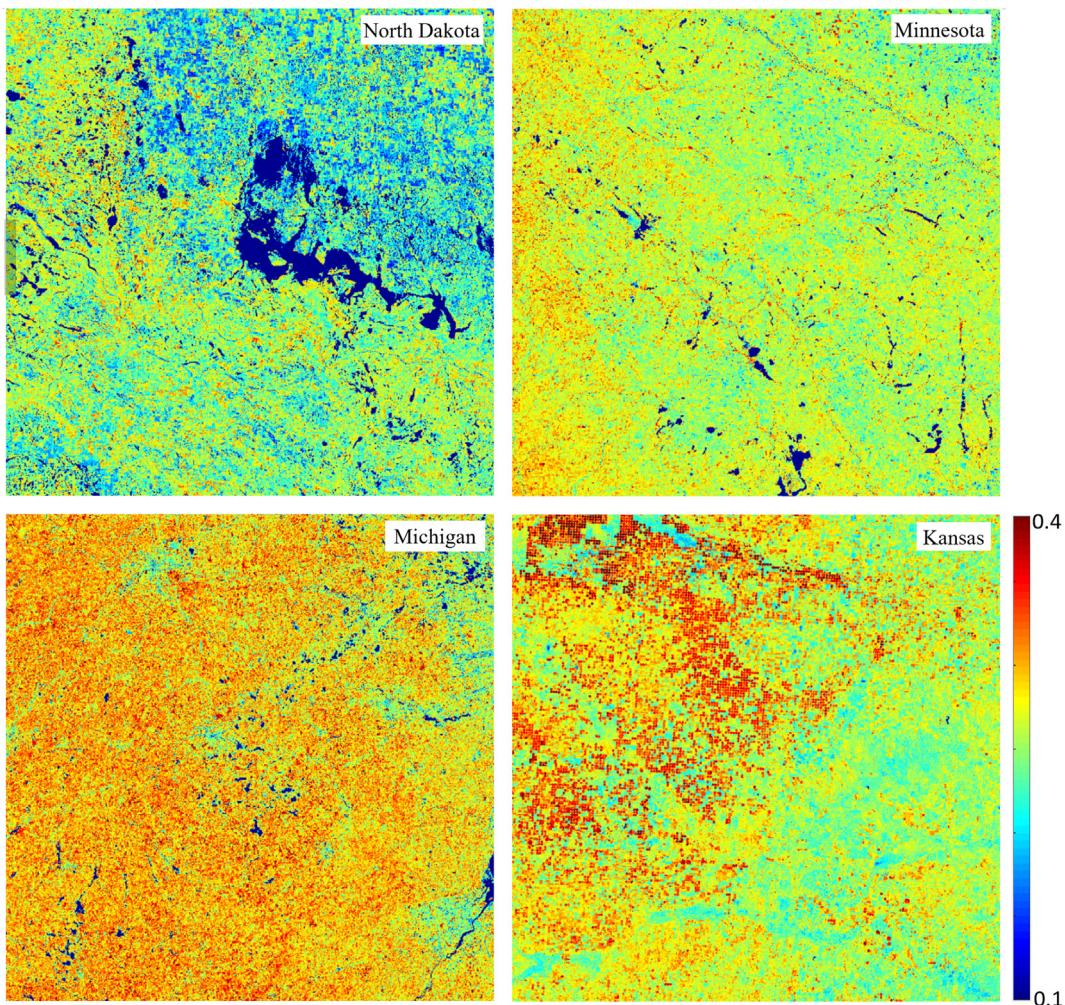
To illustrate the linear harmonic model fitting, Figs. 4 and 5 show the  $a_{0,NIR}$  linear harmonic model parameter values retrieved by fitting Eq. (1) at each ARD pixel location to the NIR training data. Fig. 4 shows the parameter values derived by fitting the model to the valid training data observations, i.e., *F* results. Fig. 5 shows the parameter values derived by fitting the model to the valid and gap-filled training data observations, i.e., *FF* results. Of the 7 model parameters for each of the six bands, the  $a_{0,NIR}$  parameter is illustrated here as it corresponds to the mean 7-month modelled NIR reflectance and so is straightforward to interpret.

In the *F* results (Fig. 4), the black pixels illustrate where the harmonic model could not be applied as there were  $< 8$  valid training data observations (they occur at the same locations as the black pixels in Fig. 3). The *F*  $a_{0,NIR}$  values have some discontinuities due to cloud and shadow contamination that are discussed later. The striping artifacts are due to the Landsat 7 ETM+ scan line corrector (SLC) failure that caused 22% of each Landsat 7 ETM+ image to be missing (Markham et al., 2004), and are particularly evident in the North Dakota, Minnesota, and Michigan tiles. The *FF*  $a_{0,NIR}$  values (Fig. 5) are more natural looking and are spatially more coherent. There are no black pixels

because the gap-filling applied to the training ARD tiles meant that there were  $\geq 8$  observations at every pixel location after the gap filling, and so the 7-parameter harmonic model can be applied everywhere. Comparing the *FF*  $a_{0,NIR}$  values with the CDL (Fig. 1) indicates that the  $a_{0,NIR}$  parameter values are high over crops such as soybean, corn and alfalfa that should have high NIR reflectance due to their vigorous vegetation growth, and, as expected, are lower over water bodies and many of the developed/open space locations. For example, across the Minnesota tile, the water class pixels mean *FF*  $a_{0,NIR}$  was 0.08 ( $\sigma = 0.06$ ), whereas for corn the mean *FF*  $a_{0,NIR}$  was 0.26 ( $\sigma = 0.02$ ) and for soybean the mean *FF*  $a_{0,NIR}$  was 0.27 ( $\sigma = 0.03$ ).

#### 4.1.2. Linear harmonic model predicted reflectance examples

Fig. 6 shows the linear-harmonic-model predicted reflectance for the SWIR (1.6  $\mu\text{m}$ ), NIR, and red bands for three days (rows), derived using the *F* (left column) and the *FF* (right column) model parameters. The *F*-predicted false colour images have no predictions where there were  $< 8$  valid training data observations and thus no fitting results. The *FF*-predicted images have predicted reflectance for every ARD tile pixel. The results are shown for the first days of July, August, and September in 2013 over the Minnesota ARD tile. This tile is illustrated because although it has an average level of missing training data where the harmonic model could not be fitted (Fig. 3), many of the missing observations occurred near the beginning of the growing season and resulted in large differences between the *F* and *FF*-predicted data which



**Fig. 5.** As Fig. 4, but showing the  $a_{0,NIR}$  parameter values derived by fitting the 7-parameter linear harmonic model to the gap-filled training data (valid observations and the SAMSTS-filled missing observations), i.e., FF model parameter results. Each tile is 5000 × 5000 30 m pixels.

is illustrated later.

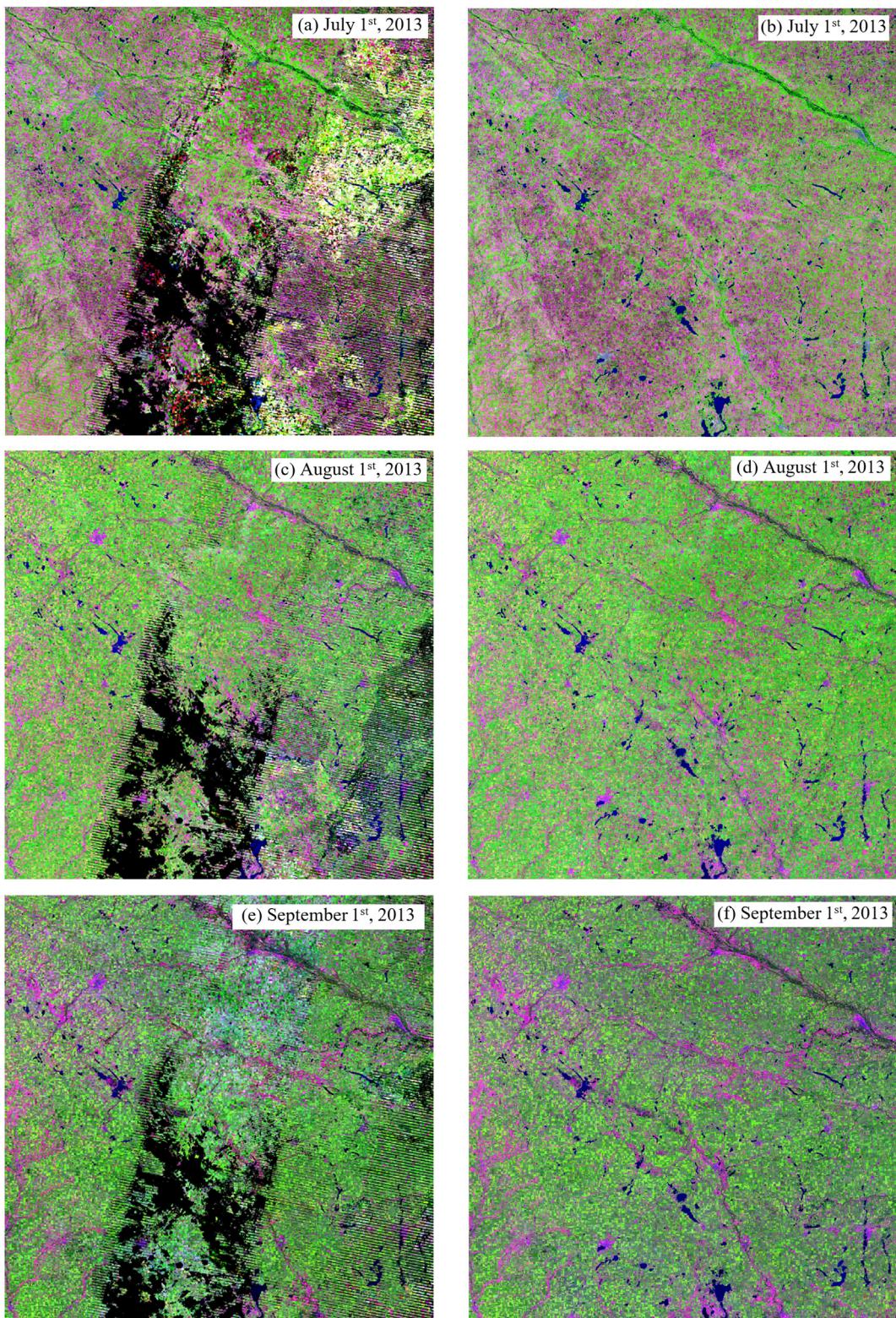
The  $F$ -predicted images appear to have poorer quality than the FF-predicted images. There are abnormally bright areas in the  $F$  images, most evident on the eastern side of the July image (Fig. 6a), caused by the occurrence of undetected clouds over the 7 months that had high SWIR, NIR and red reflectance and so inflated predicted reflectance. The overly dark areas in the south-east of the August image (Fig. 6c) are due to undetected shadows over the 7 months. Landsat 7 ETM+ SLC stripe artifacts are also quite evident in the  $F$ -predicted images. In contrast, the three FF-predicted false colour images (Fig. 6 right column) have no missing predicted reflectance values, do not appear to have the cloud and shadow contamination issues, and are natural looking with reflectance variations that generally correspond to the 2013 CDL class variations (Fig. 1).

Fig. 7 shows predicted reflectance derived using the  $F$  (left) and the FF (right) derived linear harmonic parameters for a 2000 × 5000 30 m pixel subset of the Minnesota tile. The predicted reflectance for May 1st 2013 (day 121) is shown and occurred in a period of missing observations between days 131 and 185 (Fig. 2) when the growing season was starting. Despite this, the FF predicted reflectance is coherent, especially compared with the  $F$  predicted reflectance. The FF predicted reflectance results are particularly convincing given the large number of fields and small lakes apparent across the tile subset.

#### 4.2. Illustrative single-pixel-location time series $F$ and FF examples

Fig. 8 shows six single-pixel  $F$  and FF fitting examples for different crop types and illustrate the typical quality, availability and seasonality of the Landsat ARD valid training data (solid circles) and the gap filled observations (open circles) for the 7 months. The NIR surface reflectance is shown. The red lines show the linear harmonic model fits of the valid training data, i.e., the fit ( $F$ ) results. The blue lines show the model fits of both the valid training-data observations and the gap filled observations, i.e., the fill-and-fit (FF) results.

Fig. 8(a) shows the 7-month time series for a cropland pixel in the least-cloudy Kansas tile. This location had sufficient valid Landsat observations (20) distributed through the growing season, which allowed the  $F$  approach (red line) to capture the characteristic mid-west double cropping of winter wheat and soybean (Wardlow et al., 2007). There were seven additional gap filled observations but they resulted in only slight differences between the  $F$  and FF fits, which implies that the 20 valid observations were sufficient to fit the 7-parameter linear harmonic model at this location. In contrast, Fig. 8(b) illustrates a spring wheat time series for a pixel in the North Dakota tile. Spring wheat typically has a single summer-growing-season vegetation peak. At the Fig. 8(b) pixel location, an anomalously low NIR reflectance value occurred on day 191 due to a shadow-contaminated pixel that was not labelled as such in the ARD. This outlier was the only observation in the middle of the rapid green-up period between days 175 and 215, and consequently it significantly reduced the reliability of the  $F$  fit. The FF

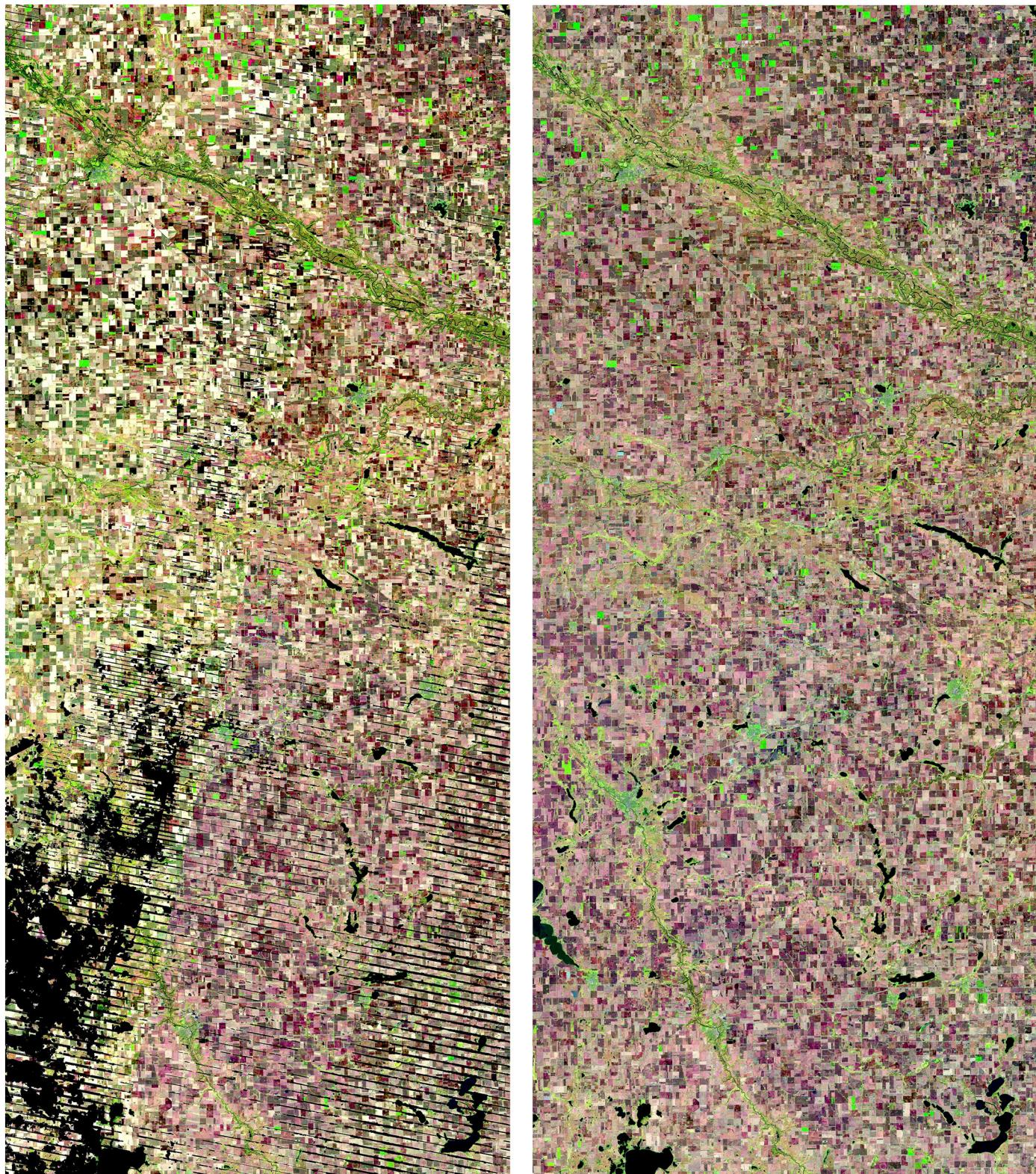


**Fig. 6.** Predicted Minnesota ARD tile SWIR (1.6  $\mu\text{m}$ ), NIR and red surface reflectance for July 1st, August 1st, and September 1st 2013. The predicted reflectance was derived with Eq. (1) using the  $F$  (left column) and the  $FF$  (right column) linear harmonic model parameters, respectively. Black shows the 30 m ARD tile pixel locations where there were insufficient ( $< 8$ ) training data observations to fit the 7-parameter linear harmonic model. Each tile is  $5000 \times 5000$  30 m pixels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fit was less affected because there were three SAMSTS gap-filled observations between days 175 and 215 (on days 182, 183 and 198).

Fig. 8(c) and (d) illustrate the results for two adjacent soybean field pixels. Both pixels had long periods of missing observations in the

growing period between days 170 and 226. On day 193, there was a valid observation at the Fig. 8(c) pixel location, but the observation was missing at the adjacent Fig. 8(d) pixel location because it fell within a Landsat 7 ETM+ SLC gap. The absence of a valid observation on day



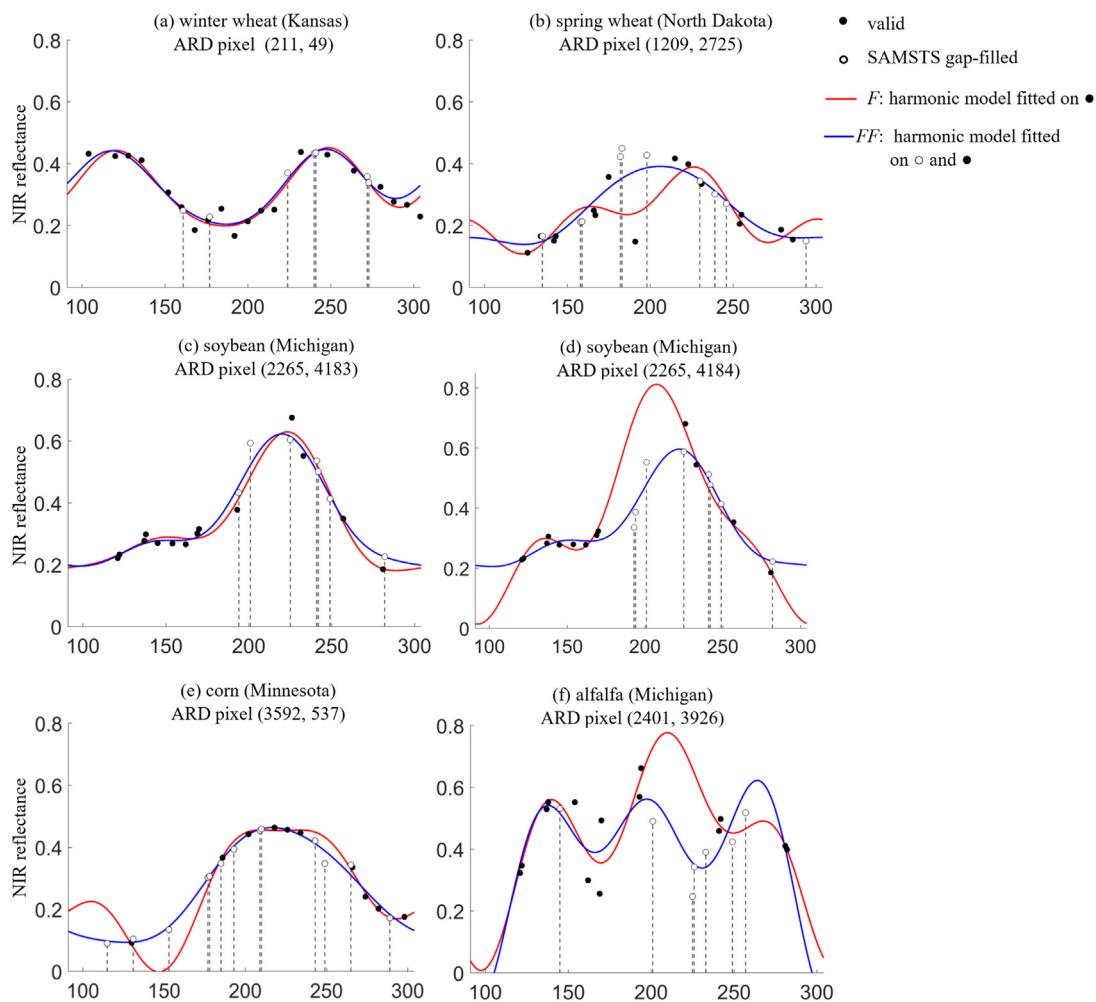
**Fig. 7.** Predicted SWIR ( $1.6 \mu\text{m}$ ), NIR and red surface reflectance for May 1st 2013 for a  $2000 \times 5000$  30 m pixel central portion of the Minnesota tile. The predicted reflectance was derived with Eq. (1) using the  $F$  (left column) and the  $FF$  (right column) linear harmonic model parameters, respectively.

193, when the soybean field was growing rapidly, combined with the long periods of missing observations between days 170 and 226, resulted in an anomalous spike in the  $F$  fit (Fig. 8d). In contrast, the  $FF$  fits for both pixels appear similar and plausible owing to the gap-filled missing observations between days 170 and 226.

Fig. 8(e) shows a corn time series for a pixel location in the

Minnesota tile. The  $F$  fit has a significantly incorrect oscillation at the beginning period of the growing season because there was only one valid observation before day 186, whereas the  $FF$  fit does not because of the six gap-filled observations that were included in this period.

Fig. 8(f) shows an alfalfa pixel time series. It is well established that alfalfa time series are difficult to fit because they have multiple cycles



**Fig. 8.** Examples of fit ( $F$ , red lines) and fill-and-fit ( $FF$ , blue lines) predicted NIR surface reflectance time series at six different ARD 30 m pixel locations. The solid circles show the valid training data NIR surface reflectance. The open circles show SAMSTS gap-filled NIR surface reflectance, on days where there was a training ARD tile acquisition (Fig. 2 shaded triangles) but no valid observation at the pixel location on that day. Results are shown for the 2013 growing season from April 1st (day 91) to October 31st (day 304) (as Fig. 2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Houborg and McCabe, 2018; Roy and Yan, 2020). In this example, neither the  $F$  nor  $FF$  approaches could model the alfalfa variation reliably. Nevertheless, the  $FF$  outperforms  $F$  as it is able to better capture a harvest that occurred between days 194 and 241 (between the second and third peaks in the blue line), which is missed by the  $F$  fit. The occurrence of this approximate harvest date is confirmed later in Fig. 10.

#### 4.3. Quantitative validation

##### 4.3.1. Detailed single-day quantitative validation examples

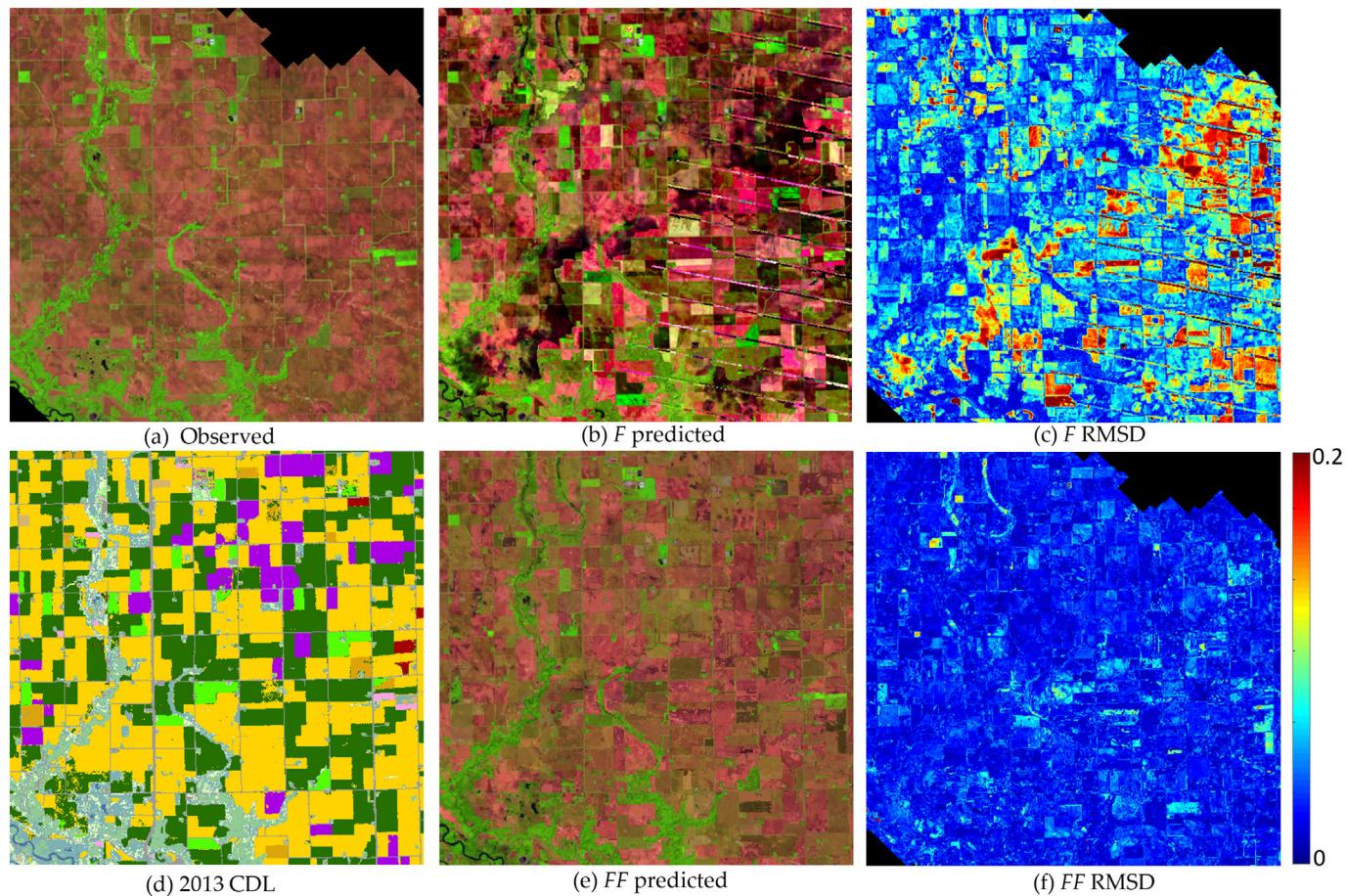
Figs. 9 and 10 show false-colour Landsat surface reflectance for test ARD tile acquisitions in the summer (a), the  $F$  and  $FF$  predicted surface reflectance for selected tile acquisition days (b) and (e), respectively, and the corresponding RMSD values (c) and (f). The 2013 CDL is also shown (d) to provide land cover context. Small subsets are shown to illustrate the results in detail.

Fig. 9 shows Minnesota results for day 170 that occurred in a period with few valid observations (Figs. 2, 8e). The subset is composed of predominantly corn and soybean, and also sugarbeet and pea (shown in purple and light green respectively in Fig. 9d). The  $F$ -predicted day 170 surface reflectance (Fig. 9b) exhibit abnormal dark areas due to undetected shadows that occurred on day 178, and also stripes caused by linear harmonic model fitting errors at locations with fewer

observations due to the Landsat 7 ETM+ SLC-off issue. The  $F$  RMSD values (Fig. 9c) are large (0.072 mean) and spatially variable (0.049 standard deviation). In contrast, the  $FF$ -predicted day 170 surface reflectance (Fig. 9e) do not exhibit these issues, and the  $FF$  RMSD values (Fig. 9f) are smaller (0.028 mean) and spatially quite uniform (0.017 standard deviation).

Fig. 10 shows Michigan tile subset results for day 202 over an area composed of predominantly corn, soybean and winter wheat. The day-202 test data (Fig. 10a) was sensed by the Landsat 7 ETM+. The  $F$ -predicted surface reflectance (Fig. 10b) has no gaps because for this subset there were  $\geq 8$  valid training observations over the 7 months at each pixel location. The  $F$ -predicted surface reflectance is reasonably coherent but patterns of SLC-off striping issue are apparent. The  $F$  RMSD values (Fig. 10c) for this subset (0.031 mean and 0.037 standard deviation) are smaller than those for the Minnesota subset (Fig. 9c). The  $FF$ -predicted surface reflectance (Fig. 10e) does not have any apparent SLC striping, and the  $FF$  RMSD values are smaller (0.025 mean) and less variable (0.017 standard deviation) than the  $F$  RMSD values.

These detailed examples illustrate the ability of the  $FF$  approach to provide coherent predicted reflectance time series. There are some isolated issues however. In the Minnesota day-170 example, the greatest  $FF$  RMSD values occur over several small fields in the northwest (Fig. 9f, yellow RMSD  $\sim 0.12$ ). They occurred because in an ARD tile acquisition a week later (day 178), the SAMSTS algorithm filled



**Fig. 9.** Minnesota  $500 \times 500$  30 m pixel subset results: (a) Landsat 8 OLI surface reflectance (SWIR 1.6  $\mu\text{m}$ , NIR, red) sensed in the test ARD tile acquisition on day 170 (June 19th 2013) after additional cloud and shadow masking, (b) the  $F$ -predicted SWIR 1.6  $\mu\text{m}$ , NIR, and red surface reflectance for day 170, (c) the  $F$  RMSD derived for the six-bands (Eq. (2)) between the  $F$ -predicted reflectance values and the observed reflectance values for day 170, (d) the 2013 CDL (colored as Fig. 1), (e) the  $FF$ -predicted surface reflectance for day 170, (f) the  $FF$  RMSD derived for the six bands between the  $FF$ -predicted reflectance values and the observed reflectance values for day 170. Pixels with no data (a) and thus no RMSD results are shown black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

missing field pixels with the values of shadow-contaminated pixels that were not labelled as such in the ARD. This kind of issue may occur with any ASP gap filling algorithm (Shen et al., 2015). In the Michigan example, the  $FF$ -predicted reflectance values for an alfalfa field (red cross in the north-east corner of Fig. 10d) were not fitted reliably. This was previously illustrated for a single pixel selected from the same alfalfa field (Fig. 8f). The field was harvested on or before day 202 (bare ground is evident across the field in Fig. 10a), and the linear harmonic fits were not able to capture this change in the presence of the other complex alfalfa growing season variations. Notably, however, the alfalfa field pixels had lower  $FF$  RMSD values (about 0.11 over the field) compared to the  $F$  RMSD values (about 0.17). As noted earlier, alfalfa has rapid temporal variations that are difficult to fit reliably using Landsat time series.

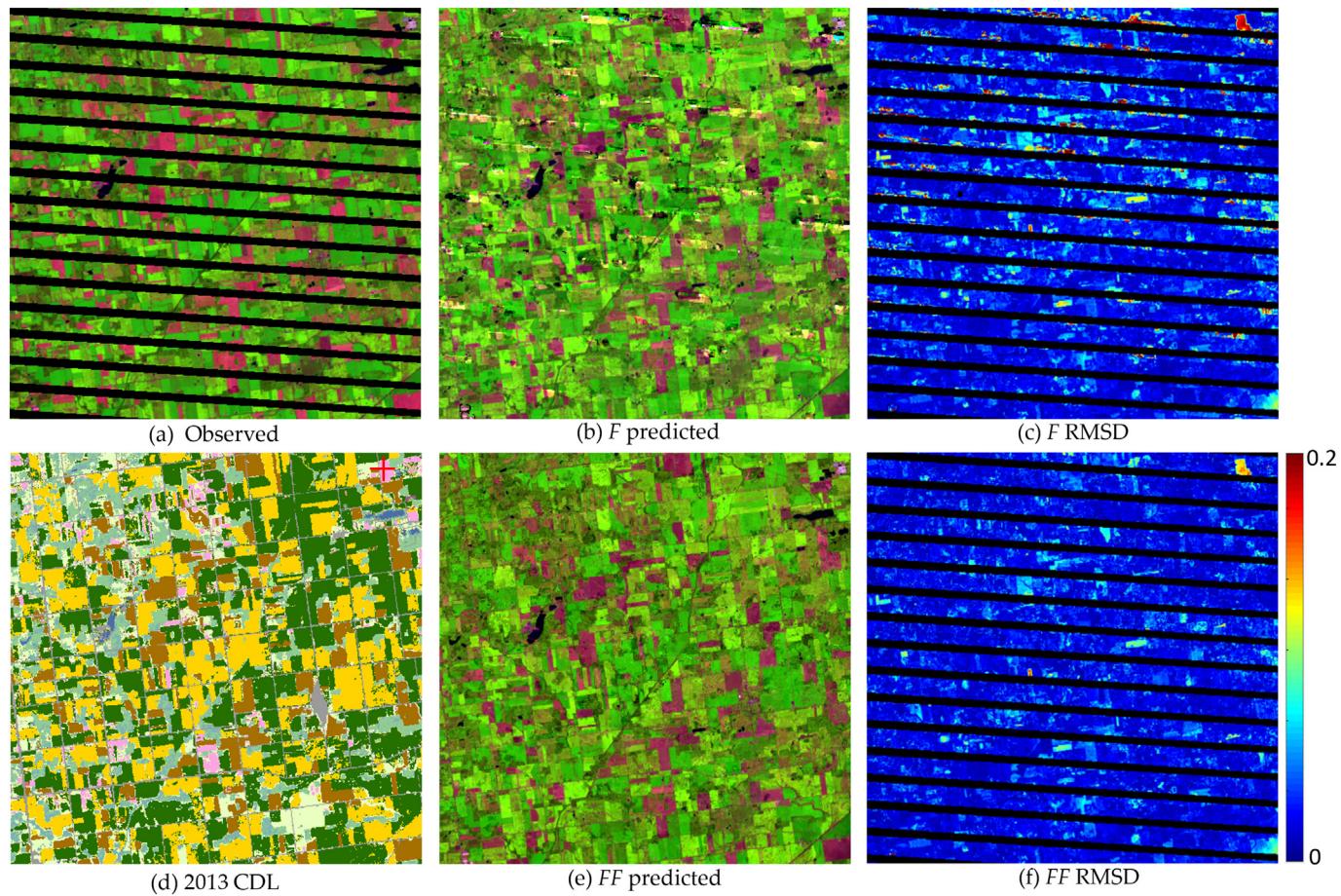
#### 4.3.2. Summary quantitative validation results

Fig. 11 shows the mean daily RMSD values derived as Eq. (3). Recall that in order to compare meaningfully the  $F$  and  $FF$  results, the mean RMSD values are defined for the same pixel locations and dates, i.e., only at pixel locations with  $\geq 8$  valid training data observations over the 7 months (i.e., at the colored pixel locations illustrated in Fig. 3) and where and when there was a valid ARD test pixel observation.

In Fig. 11, the majority of the  $FF \overline{\text{RMSD}}(t)$  values are  $< 0.03$  which is the uncertainty specification for the Landsat 8 OLI top of atmosphere reflectance (which is more stringent than the Landsat 7 ETM +

uncertainty specification) (Mishra et al., 2014). In general, for each day, the  $F \overline{\text{RMSD}}(t)$  values are greater than the  $FF \overline{\text{RMSD}}(t)$  values or they are within 0.01 of each other. Particularly large differences between the  $F$  and  $FF \overline{\text{RMSD}}(t)$  values occur for the Minnesota tile prior to day 255, which is due to the small number of valid training data between days 131 and 185, as remarked upon earlier. The North Dakota tile has large differences between the  $F$  and  $FF \overline{\text{RMSD}}(t)$  values on day 171 due to a long period of missing training data around day 171, and has large differences on day 271 due to undetected cloud- and shadow-contaminated ARD training data on day 266. The Kansas tile has generally the smallest differences between the  $F$  and  $FF \overline{\text{RMSD}}(t)$  values which, as noted earlier with respect to the single-pixel Kansas example (Fig. 8a), was because there were sufficient valid training data observations for the  $F$  fitting without the need for prior gap-filling.

Table 2 summarizes the overall mean ( $\overline{\text{RMSD}_{all}}$ ) and standard deviation ( $\sigma_{all}$ ) RMSD values derived considering all of the test ARD tile acquisitions for the 7 months. More than 7 million (Michigan), 10 million (North Dakota), 49 million (Minnesota), and 88 million (Kansas) pixel observations ( $n_{all}$ ) were considered. The Michigan and Kansas tiles had the smallest and largest  $n_{all}$  respectively, reflecting the different number and degree of cloudiness of the ARD acquisitions over these tiles (Figs. 2 and 3). The results show unambiguously that the  $FF$  approach provides more accurate reflectance prediction than the  $F$  approach. Among the four tiles, the  $FF \overline{\text{RMSD}_{all}}$  and  $\sigma_{all}$  values are small and similar – the  $\overline{\text{RMSD}_{all}}$  values range from 0.025 to 0.026 and the  $\sigma_{all}$



**Fig. 10.** Michigan 500 × 500 30 m pixel subset results, caption details as Fig. 9, but for a test Landsat 7 ETM + acquisition sensed on day 202 (July 21st 2013). The red cross in the north-east of (d) marks the Fig. 8(f) alfalfa pixel location. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

values range from 0.018 to 0.020. In contrast, the  $F \overline{RMSD}_{all}$  values are greater and range from 0.026 (Kansas) to 0.047 (Minnesota) and the  $\sigma_{all}$  values range from 0.020 (Kansas) to 0.085 (Minnesota). The differences between the overall  $FF$  and  $F$  statistics follow the same tile patterns observed in Fig. 11.

Fig. 12 shows the dependency of the mean RMSD on the number of observations used, i.e.,  $\overline{RMSD}(nobs)$  (Eq. (6)) where  $nobs$  is the number of valid training data observations over the seven months and is defined for each unique value  $\geq 8$  in Fig. 3. The Kansas  $F$  and  $FF \overline{RMSD}(nobs)$  values are similar for all  $nobs$  which is because there were already sufficient valid observations for the  $F$  fits. The Minnesota tile  $F$  and  $FF \overline{RMSD}(nobs)$  values are large compared to the other tiles for  $nobs < 11$ . This is because of the long period of missing observations at the beginning of the growing season that was discussed previously.

The following observations are made concerning the results shown in Fig. 12. First, the  $FF \overline{RMSD}(nobs)$  values are all  $\leq 0.03$  and are generally smaller than the  $F \overline{RMSD}(nobs)$  values (or are no greater than 0.002 of the  $F \overline{RMSD}(nobs)$  values).

Second, the  $FF \overline{RMSD}(nobs)$  values have less variation with respect to  $nobs$ . These two observations indicate that the  $FF$  approach can more effectively model the reflectance time series to generate daily synthetic reflectance images in a robust manner with respect to the amount of available valid input data.

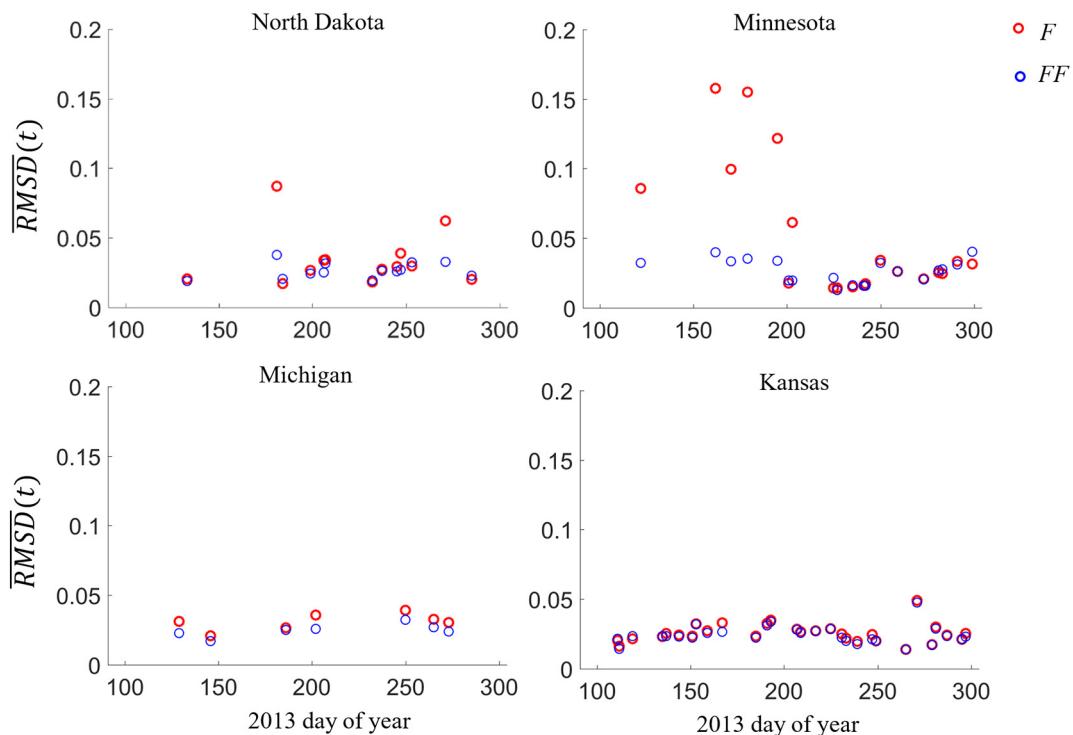
Third, the  $\overline{RMSD}(nobs)$  values generally decrease with  $nobs$ . This is expected as greater model fitting errors will occur at locations where there are fewer available observations, which increases the likelihood of there being longer periods of missing data and also increases model fitting sensitivity with respect to anomalous reflectance values such as

undetected cloud and shadow contaminated data (Roy and Yan, 2020). This pattern is not always apparent however, for example, most notably the  $F \overline{RMSD}(nobs)$  North Dakota values increase when  $nobs = 13$  and  $nobs = 14$ . This occurred because the training pixel locations with  $nobs = 11$  and 12 were clustered in the similar region of the tile and were fit relatively reliably by the linear harmonic model. We also note the convergence between the  $FF$  and  $F \overline{RMSD}(nobs)$  values with increasing  $nobs$ , which is expected because with more temporally-rich observations, there are fewer SAMSTS gap-filled observations used in the  $FF$  approach. Thus, given sufficient valid observations,  $FF$  may not outperform  $F$ .

Fourth, the  $F$  and  $FF \overline{RMSD}(nobs)$  are never zero. This reflects the fact the 7-parameter linear harmonic model (and probably all other models) cannot perfectly model land surface time series observations, and that there is a limit to the agreement between a fitted theoretical harmonic model and the satellite time series. Notably, Landsat surface reflectance are not error free due to several issues particularly those associated with atmospheric correction and sensor calibration (Helder et al., 2018).

## 5. Conclusions

Characterizing land surfaces by fitting time series models to satellite time series has an established heritage using coarse-resolution daily or near-daily data sensed from polar orbiting satellites (Moody and Johnson, 2001, Sakamoto et al. 2005, White et al. 2009, Zhang et al., 2018). The recently available Landsat ARD are defined in fixed geolocated temporally “stackable” tiles that increase the ease of applying



**Fig. 11.** Mean six-band surface reflectance root-mean-square difference (RMSD) values for each test ARD tile acquisition calculated as Eq. (3). The plotted  $\overline{RMSD}(t)$  values were derived for the *F*-predicted (red) and the *FF*-predicted (blue) surface reflectances. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Overall mean ( $\overline{RMSD}_{all}$ , Eq. (4)) and standard deviation ( $\sigma_{all}$ , Eq. (5)) RMSD values derived considering  $n_{all}$  different pixel location and day comparisons for all the test ARD tile acquisitions for the *F* and *FF* results.

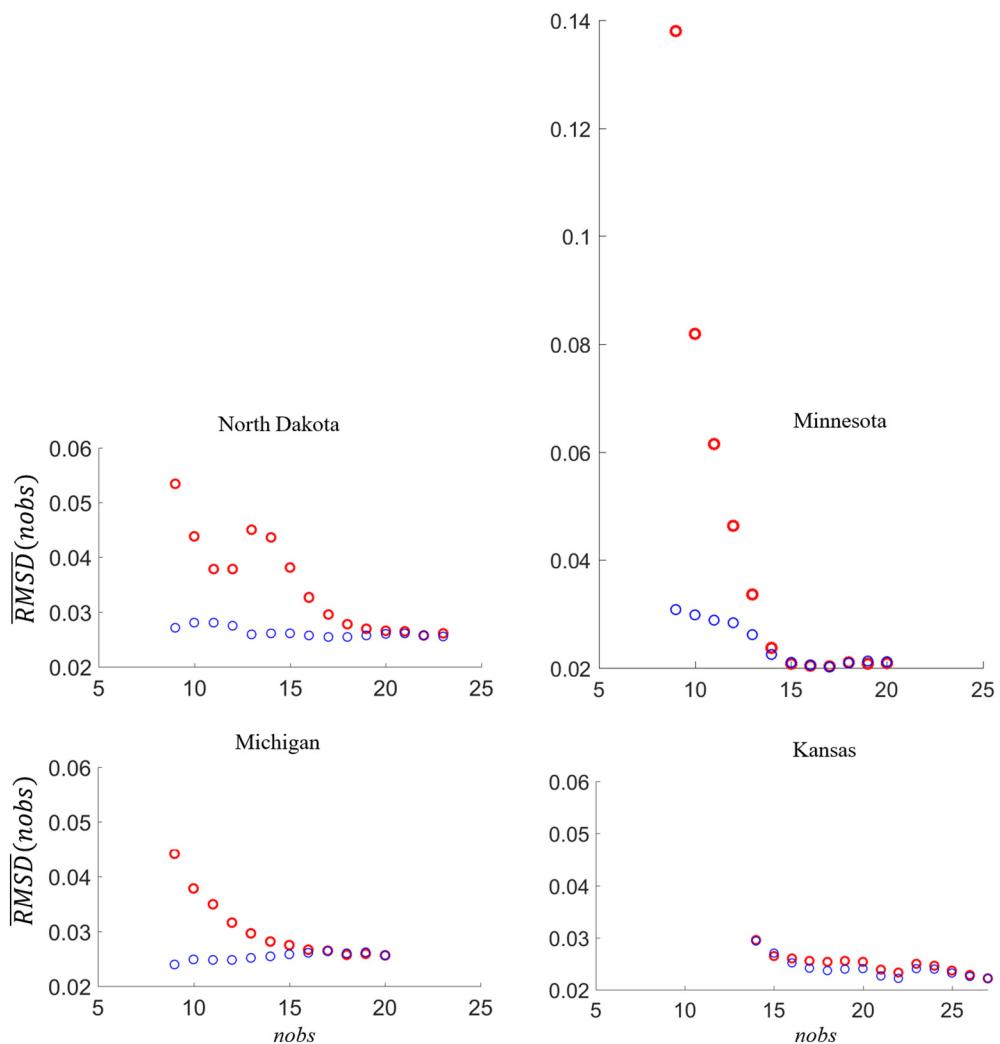
		Fit (F)	Fill-and-fit (FF)
North Dakota	$\overline{RMSD}_{all}$	0.036	0.026
	$\sigma_{all}$	0.044	0.019
	$n_{all}$	10,353,382	10,353,382
Minnesota	$\overline{RMSD}_{all}$	0.047	0.026
	$\sigma_{all}$	0.085	0.020
	$n_{all}$	49,953,136	49,953,136
Michigan	$\overline{RMSD}_{all}$	0.031	0.025
	$\sigma_{all}$	0.043	0.018
	$n_{all}$	7,033,421	7,033,421
Kansas	$\overline{RMSD}_{all}$	0.026	0.025
	$\sigma_{all}$	0.020	0.019
	$n_{all}$	88,250,360	88,250,360

time series models to Landsat data. However, Landsat observations are far less frequent than coarse resolution data, even over the CONUS where Landsat data availability is greater than elsewhere (Wulder et al., 2016). If there are too few available observations, then time series models cannot be fitted, or low order time series models must be used that may not reliably capture surface variations. For example, our recent research over CONUS cropland Landsat ARD tiles demonstrated that there can be insufficient observations to apply 5 and 7 parameter harmonic models reliably, even when growing season data from two Landsat sensors are used together (Roy and Yan, 2020). If there are sufficient observations to fit a model, then the fitting accuracy is dependent upon the number of observations used, the model complexity, and on when the observations occurred relative to the surface variations (Carrão et al., 2010; Brooks et al., 2012; Roy and Yan, 2020).

The fill-and-fit (*FF*) approach described in this paper was developed to model medium resolution satellite time series in a way that is robust to missing observations. It employs a gap filling step before the

application of a linear harmonic time series model. The model parameters are used to predict the surface reflectance for any day. The *FF* approach does not use other satellite data and so the spatial resolution of the predicted reflectance is not degraded. It was demonstrated using seven months of Landsat 7 ETM+ and Landsat 8 OLI ARD over four agricultural ARD tiles. The predicted surface reflectance examples generated using the *FF* approach were spatially coherent and natural looking. In contrast, conventional harmonic model fitting without prior gap filling, i.e., the fit (*F*) approach, provided less reliable predicted surface reflectance. This was because fewer observations were used to fit the linear harmonic time series model and so the results were more sensitive to the occurrence of undetected cloud and shadow contaminated observations in the ARD. In addition, long periods of missing data occurring in periods of surface change resulted in spurious oscillations in the *F* model fit results, which were typically less pronounced or not present in the *FF* fit results because missing observations were gap filled with values from alternative similar pixels.

In principal, any reliable alternative similar pixel (ASP) gap-filling algorithm could be used in the *FF* approach. ASP gap-filling algorithms may provide less coherent results across tile boundaries, but this is a generic ASP gap-filling issue, and using nested groups of tiles may help overcome it (Lassalle et al., 2015; Yan and Roy, 2018a). In this study, the SAMSTS algorithm was used and the code is publically available. The SAMSTS algorithm finds ASPs by considering spatio-temporal similarity in multispectral reflectance time series. It was demonstrated recently to be able to fill spatially extensive gaps in Landsat ARD, and is robust to land cover changes occurring in the time series, and is sufficiently computationally efficient that it can be applied to ARD time series on standard desktop computers (Yan and Roy, 2018a). Typically, for any of the four study tiles, it took five hours to undertake the SAMSTS gap filling and only a few minutes to fit the harmonic model. Other time series fitting models could be used instead of the 7-parameter linear harmonic model in the fill-and-fit approach, although previously, we demonstrated the capability of the 7-parameter linear harmonic model to fit complex cropland Landsat time series (Roy and



**Fig. 12.** Mean RMSD plotted with respect to the number of observations, i.e.  $\overline{RMSD}(nobs)$  calculated by Eq. (6), where  $nobs$  ( $\geq 8$ ) is the number of valid Landsat observations over the 7 months of ARD training data (shown colored in Fig. 3). The plotted  $\overline{RMSD}(nobs)$  values were derived using the *F*-predicted (red) and the *FF*-predicted (blue) surface reflectance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Yan, 2020).

The accuracy of the *FF* and the conventional *F* approaches were quantified in this study using an independent Landsat data set composed of millions of multi-temporal pixel observations for each ARD tile. The six Landsat reflective band root-mean-square difference (RMSD) between the predicted and the independent reflectance showed that the *FF* approach was more accurate than the *F* approach. The mean *FF* RMSD values, considering  $> 7$  million (Michigan) to  $> 88$  million (Kansas) independent pixel observations, varied from 0.025 to 0.026 for the four tiles, whereas the mean *F* RMSD values varied from 0.026 to 0.047. Similarly, the standard deviation of the RMSD was smaller for the *FF* results, ranging from 0.018 to 0.020 for the four tiles, compared to the *F* results that ranged from 0.020 to 0.085. The mean daily RMSD values also reflected this pattern (Fig. 11). The greatest differences between the daily mean RMSD values for the two approaches were for the Minnesota tile data, due to a long period of missing early-growing-season data, and the smallest differences were for the Kansas tile that was the least cloudy and had sufficient valid observations for reliable fitting with either approach.

The RMSD metric is derived with respect to the six Landsat surface reflectance bands and considering Landsat 8 OLI and Landsat 7 ETM+ time series. The noise characteristics of Landsat surface reflectance depends primarily on the sensor design, the calibration, and the atmospheric correction (Mishra et al., 2014; Roy et al., 2016b; Helder

et al., 2018). The Landsat uncertainty specification for the top of atmosphere (TOA) OLI and ETM+ reflectance is 0.03 and 0.05, respectively (Mishra et al., 2014; Markham et al., 2004). The atmospheric correction uncertainty is hard to generalize because the impact of the atmosphere on TOA reflectance depends on the atmospheric constituents and the surface reflectance, and varies spectrally (Kaufman and Senda, 1988; Roy et al., 2014b). The Landsat ARD used in this study was atmospherically corrected using the Landsat Surface Reflectance Code (LaSRC) and the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithms for Landsat 8 OLI and Landsat 7 ETM+, respectively (Dwyer et al., 2018). The LaSRC Landsat 8 OLI atmospheric correction uncertainty varies, from 0.012 in the OLI blue band, to 0.004 and 0.003 in the two OLI SWIR bands (Doxani et al., 2018), and the LEDAPS atmospheric correction uncertainty is lower (Ju et al., 2012). Notably, the mean *FF* RMSD values for the four tiles were  $< 0.03$ , which is comparable to the uncertainty specification for the Landsat 8 OLI TOA reflectance, but greater than the atmospheric correction uncertainty in any Landsat 8 OLI band.

The optimal number of observations for reliable time series model fitting is poorly specified for Landsat. This is because, as illustrated in Fig. 7, Landsat observations are not usually sensed uniformly in time, and the same or similar number of observations may provide different fitting results depending on when the observations were acquired relative to the surface changes. This is complicated further because

satellite observations are not noise-free and may have additive and multiplicative noise (Helder et al., 2018), and include isolated and large errors, for example, due to undetected cloud and shadow contamination, and errors due to a variety of other sources (Roy et al., 2002). The minimum number is dictated by the model complexity. Fundamentally, there must be at least as many observations as model parameters to undertake the model fitting, and in this study, at least 8 observations were required because a 7-parameter linear harmonic model fitted by least squares adjustment was used. The *F* approach could not be applied to every ARD tile pixel location, particularly over the cloudiest Michigan tile. In contrast, the *FF* approach could be applied to every ARD tile pixel location because of the gap-filling.

The optimal minimum number of observations for reliable time series modelling may be broadly specified using heuristic recommendations, such as there be at least three times more surface observations than the number of linear harmonic model coefficients (Zhu and Woodcock, 2014). This heuristic was supported by the reported dependency of the *F* and *FF* results with respect to the number of valid (but not gap filled) surface observations used to fit the model (Fig. 12). The *F* RMSD values declined with the number of observations but the *F* RMSD values became similar to each other when the number of observations was greater than about 19 (North Dakota), 15 (Minnesota), 16 (Michigan) and 21 (Kansas). This supports the recommendation that a 7-parameter linear harmonic model should use  $\geq 21$  observations to reliably fit Landsat time series data if prior gap-filling is not undertaken. A recent examination of the Landsat ARD record found that the CONUS average annual number of non-cloudy observations at each 30 m ARD pixel location was 4.85 (Landsat 4 TM), 16.41 (Landsat 5 TM), and 15.03 (Landsat 7 ETM+) (Egorov et al., 2019). Therefore, conventional linear harmonic time series fitting will not be reliable for many CONUS regions when fitting a single year of single Landsat sensor data, as there are rarely  $\geq 21$  observations. In contrast, the *FF* RMSD values were not sensitive to the number of valid observations used ( $\geq 8$ ). This indicates the robustness of the *FF* approach and that it has potential for wall-to-wall CONUS application to the historical Landsat record back to the start of the Landsat 5 TM era.

This study illustrates the potential of using *FF* predicted synthetic reflectance time series data for large area land surface monitoring. The ability to fit time series models to a single year or growing season of data is particularly useful to reduce the effects of inter-annual variations and for monitoring of croplands as different crops are rotated between years. Consequently, in this study we focused on the challenging issue of Landsat reflectance time series modelling using a single year (crop growing season) of satellite data. However, there is no reason why the fill-and-fit approach could not be adapted to multiple years of satellite data. Further, our expectation is that similar results would be obtained if the algorithm was applied to other medium resolution satellite data such as Sentinel-2. For many regions, the increased temporal observation coverage provided by the two Sentinel-2 satellites, particularly when combined with Landsat (Li and Roy, 2017), may reduce the need for the *FF* approach. When sufficient valid observations are available and when there are no long periods of missing observations, the *FF* approach may not outperform the *F* approach. This was demonstrated for the Kansas data that were acquired under relatively non-cloudy conditions and so there were sufficient observations for reliable linear harmonic modelling without the need for prior gap-filling. However, in cloudy regions, and over certain highly temporally variable land covers such as alfalfa crops, this may not be the case. Further work to examine this is recommended. Finally, we note that validation of the described approach using ground-based or higher spatial resolution observations may be helpful, although challenging given the need for a large sample of daily observations, and this is also recommended for future research.

## CRediT authorship contribution statement

**Lin Yan:** Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft. **David P. Roy:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The USGS Landsat program management and staff, are thanked for the free provision of the Landsat ARD. This research was funded by the U.S. Geological Survey Landsat science team (grant 140G0119C0009).

## References

- Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: testing multiple methods and satellites in three countries. *Remote Sens. Environ.* 202, 129–141.
- Basso, B., Liu, L., 2018. Seasonal crop yield forecast: methods, applications, and accuracies. In: *Advances in Agronomy*. Academic Press, pp. 154.
- 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 International* 26 (5), 341–358.
- Brooks, E.B., Thomas, V.A., Wynne, R.H., Coulston, J.W., 2012. Fitting the multitemporal curve: a Fourier series approach to the missing data problem in remote sensing analysis. *IEEE Trans. Geosci. Remote Sens.* 50 (9), 3340–3353.
- Brooks, E., Wynne, R., Thomas, V., 2018. Using window regression to gap-fill Landsat ETM + post-SLC-off data. *Remote Sens.* 10 (10), 1502.
- Brown, J.F., Tollerud, H.J., Barber, C.P., Zhou, Q., Dwyer, J.L., Vogelmann, J.E., Loveland, T.R., Woodcock, C.E., Stehman, S.V., Zhu, Z., Pengra, B.W., 2019. Lessons learned implementing an operational continuous United States national land change monitoring capability: the Land Change Monitoring, Assessment, and Projection (LCMAP) approach. *Remote Sens. Environ.* 111356.
- Carrão, H., Gonçalves, P., Caetano, M., 2010. A nonlinear harmonic model for fitting satellite image time series: analysis and prediction of land cover dynamics. *IEEE Trans. Geosci. Remote Sens.* 48 (4), 1919–1930.
- DeVries, B., Prathist, A.K., Verbesselt, J., Kooistra, L., Herold, M., 2016. Characterizing forest change using community-based monitoring data and Landsat time series. *PLoS One* 11 (3), e0147121.
- Doxani, G., Vermote, E., Roger, J.C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., 2018. Atmospheric correction inter-comparison exercise. *Remote Sens.* 10 (2), 352.
- Dwyer, J.L., Roy, D.P., Sauer, B., Jenkinson, C.B., Zhang, H.K., Lymburner, L., 2018. Analysis ready data: enabling analysis of the Landsat archive. *Remote Sens.* 10 (9), 1363.
- Egorov, A.V., Roy, D.P., Zhang, H.K., Li, Z., Yan, L., Huang, H., 2019. Landsat 4, 5 and 7 (1982 to 2017) analysis ready data (ARD) observation coverage over the conterminous United States and implications for terrestrial monitoring. *Remote Sens.* 11 (4), 447.
- 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 (8), 2207–2218.
- Ghilani, C.D., 2010. *Adjustment Computations: Spatial Data Analysis*, 5th edition. Wiley Press978-0-470-46491-5.
- Gitelson, A.A., Peng, Y., Masek, J.G., Rundquist, D.C., Verma, S., Suyker, A., Baker, J.M., Hatfield, J.L., Meyers, T., 2012. Remote estimation of crop gross primary production with Landsat data. *Remote Sens. Environ.* 121, 404–414.
- Goward, S.N., Williams, D.L., Arvidson, T., Rocchio, L.E., Irons, J.R., Russell, C.A., Johnston, S.S., 2017. *Landsat's Enduring Legacy: Pioneering Global Land Observations from Space*. American Society for Photogrammetry and Remote Sensing.
- Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sens. Environ.* 220, 135–151.
- Helder, D., Markham, B., Morfitt, R., Storey, J., Barsi, J., Gascon, F., Clerc, S., LaFrance, B., Masek, J., Roy, D., Lewis, A., 2018. Observations and recommendations for the calibration of Landsat 8 OLI and sentinel 2 MSI for improved data interoperability. *Remote Sens.* 10 (9), 1340.
- Hermance, J.F., 2007. Stabilizing high-order, non-classical harmonic analysis of NDVI data for average annual models by damping model roughness. *Int. J. Remote Sens.* 28 (12), 2801–2819.
- Houborg, R., McCabe, M., 2018. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. *Remote Sens.* 10 (6), 890.

- Jakabauskas, M.E., Legates, D.R., Kastens, J.H., 2001. Harmonic analysis of time-series AVHRR NDVI data. *Photogramm. Eng. Remote. Sens.* 67 (4), 461–470.
- Johnson, D.M., Mueller, R., 2010. The 2009 cropland data layer. *Photogramm. Eng. Remote. Sens.* 76 (11), 1201–1205.
- Ju, J., Roy, D.P., Shuai, Y., Schaaf, C., 2010. Development of an approach for generation of temporally complete daily nadir MODIS reflectance time series. *Remote Sens. Environ.* 114 (1), 1–20.
- Ju, J., Roy, D.P., Vermote, E., Masek, J., Kovalskyy, V., 2012. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. *Remote Sens. Environ.* 122, 175–184.
- Kaufman, Y.J., Sende, C., 1988. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. *Int. J. Remote Sens.* 9 (8), 1357–1381.
- Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sens. Environ.* 114 (12), 2897–2910.
- Kovalskyy, V., Roy, D.P., 2013. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30 m Landsat product generation. *Remote Sens. Environ.* 130, 280–293.
- Lark, T.J., Mueller, R.M., Johnson, D.M., Gibbs, H.K., 2017. Measuring land-use and land-cover change using the US department of agriculture's cropland data layer: cautions and recommendations. *Int. J. Appl. Earth Obs. Geoinf.* 62, 224–235.
- Lassalle, P., Ingla, J., Michel, J., Grizonnet, M., Malik, J., 2015. A scalable tile-based framework for region-merging segmentation. *IEEE Trans. Geosci. Remote Sens.* 53 (10), 5473–5485.
- Li, J., Roy, D.P., 2017. A global analysis of sentinel-2A, sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens.* 9 (9), 902.
- Malambo, L., Heatwole, C.D., 2015. A multitemporal profile-based interpolation method for gap filling nonstationary data. *IEEE Trans. Geosci. Remote Sens.* 54 (1), 252–261.
- Markham, B.L., Thome, K.J., Barsi, J.A., Kaita, E., Helder, D.L., Barker, J.L., Scaramuzza, P.L., 2004. Landsat-7 ETM+ on-orbit reflective-band radiometric stability and absolute calibration. *IEEE Trans. Geosci. Remote Sens.* 42 (12), 2810–2820.
- Menenti, M., Azzali, S., Verhoeve, W., Van Swol, R., 1993. Mapping agroecological zones and time lag in vegetation growth by means of Fourier analysis of time series of NDVI images. *Adv. Space Res.* 13 (5), 233–237.
- Mishra, N., Haque, M.D., Leigh, L., Aaron, D., Helder, D., Markham, B., 2014. Radiometric cross calibration of Landsat 8 operational land imager (OLI) and Landsat 7 enhanced thematic mapper plus (ETM+). *Remote Sens.* 6 (12), 12619–12638.
- Moody, A., Johnson, D.M., 2001. Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sens. Environ.* 75 (3), 305–323.
- Olsson, L., Eklundh, L., 1994. Fourier series for analysis of temporal sequences of satellite sensor imagery. *Int. J. Remote Sens.* 15, 3735–3741.
- Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L., Zhu, Z., 2019. Making Landsat time series consistent: evaluating and improving Landsat analysis ready data. *Remote Sens.* 11 (1), 51.
- Roy, D.P., Borak, J.S., Devadiga, S., Wolfe, R.E., Zheng, M., Descloitres, J., 2002. The MODIS land product quality assessment approach. *Remote Sens. Environ.* 83 (1–2), 62–76.
- Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multitemporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sens. Environ.* 112 (6), 3112–3130.
- Roy, D.P., Wulder, M.A., Loveland, T.R., C.E., W., 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., Hippel, 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., 2014a. Landsat-8: science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172.
- Roy, D.P., Qin, Y., Kovalskyy, V., Vermote, E.F., Ju, J., Egorov, A., Hansen, M.C., Kommareddy, I., Yan, L., 2014b. Conterminous United States demonstration and characterization of MODIS-based Landsat ETM+ atmospheric correction. *Remote Sens. Environ.* 140, 433–449.
- Roy, D.P., Yan, L., 2020. Robust Landsat-based crop time series modelling. *Remote Sens. Environ.* 238 (1), 110810.
- 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., 2016a. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* 176, 255–271.
- Roy, D.P., Kovalskyy, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016b. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sens. Environ.* 185, 57–70.
- Roy, D.P., Huang, H., Boschetti, L., Giglio, L., Yan, L., Zhang, H.H., Li, Z., 2019. Landsat-8 and Sentinel-2 burned area mapping—a combined sensor multi-temporal change detection approach. *Remote Sens. Environ.* 231, 111254.
- Rufin, P., Frantz, D., Ernst, S., Rabe, A., Griffiths, P., Özdogan, M., Hostert, P., 2019. Mapping cropping practices on a national scale using intra-annual Landsat time series binning. *Remote Sens.* 11 (3), 232.
- Sacks, W.J., Kucharik, C.J., 2011. Crop management and phenology trends in the US Corn Belt: impacts on yields, evapotranspiration and energy balance. *Agric. For. Meteorol.* 151 (7), 882–894.
- Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., Ohno, H., 2005. A crop phenology detection method using time-series MODIS data. *Remote sensing of environment* 96 (3–4), 366–374.
- Schmidt, M., Pringle, M., Devadas, R., Denham, R., Tindall, D., 2016. A framework for large-area mapping of past and present cropping activity using seasonal Landsat images and time series metrics. *Remote Sens.* 8 (4), 312.
- Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015. Missing information reconstruction of remote sensing data: a technical review. *IEEE Geoscience and Remote Sensing Magazine* 3 (3), 61–85.
- Vicente-Serrano, S.M., Gouveia, C., Camarero, J.J., Beguería, S., Trigo, R., López-Moreno, J.I., Azorín-Molina, C., Pasho, E., Lorenzo-Lacruz, J., Revuelto, J., Morán-Tejeda, E., 2013. Response of vegetation to drought time-scales across global land biomes. *Proc. Natl. Acad. Sci.* 110 (1), 52–57.
- 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 US central Great Plains. *Remote Sens. Environ.* 108 (3), 290–310.
- White, M.A., de Beurs, K.M., Didan, K., Inouye, D.W., Richardson, A.D., Jensen, O.P., 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology* 15 (10), 2335–2359.
- Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P., 2016. The global Landsat archive: status, consolidation, and direction. *Remote Sens. Environ.* 185, 271–283.
- Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., Dwyer, J., 2019. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* 225, 127–147.
- Yan, L., Roy, D.P., 2018a. Robust large-area gap filling of Landsat reflectance time series by spectral-angle-mapper based spatio-temporal similarity (SAMSTS). *Remote Sens.* 10 (4).
- Yan, L., Roy, D.P., 2018b. SAMSTS Satellite Time Series Gap Filling Source Codes – Landsat, Version 1.0. [https://openprairie.sdsstate.edu/samsts/satellite\\_gap\\_filling/1](https://openprairie.sdsstate.edu/samsts/satellite_gap_filling/1).
- Zhang, X., Jayavelu, S., Liu, L., Friedl, M.A., Henebry, G.M., Liu, Y., Schaaf, C.B., Richardson, A.D., Gray, J., 2018. Evaluation of land surface phenology from VIIRS data using time series of PhenoCam imagery. *Agric. For. Meteorol.* 256, 137–149.
- Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* 144, 152–171.
- Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* 114 (11), 2610–2623.