



## Evaluation of Landsat image compositing algorithms

Shi Qiu<sup>a,\*</sup>, Zhe Zhu<sup>a,\*</sup>, Pontus Olofsson<sup>b,c</sup>, Curtis E. Woodcock<sup>c</sup>, Suming Jin<sup>d</sup>

<sup>a</sup> Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 06269, USA

<sup>b</sup> NASA Marshall Space Flight Center, Huntsville, AL 35808, USA

<sup>c</sup> Department of Earth and Environment, Boston University, 685 Commonwealth Avenue, Boston, MA 02215, USA

<sup>d</sup> U.S. Geological Survey Earth Resources Observation and Science Center, Sioux Falls, SD 57198, USA

### ARTICLE INFO

Edited by Dr. Marie Weiss

#### Keywords:

Landsat  
Compositing  
MAX-RNB  
Comparison  
Land cover  
Land change

### ABSTRACT

We proposed a new image compositing algorithm (MAX-RNB) based on the maximum ratio of Near Infrared (NIR) to Blue band (RNB), and evaluated it together with nine other compositing algorithms: MAX-NDVI (maximum Normalized Difference Vegetation Index), MED-NIR (median NIR band), WELD (conterminous United States Web-Enabled Landsat Data), BAP (Best Available Pixel), PAC (Phenology Adaptive Composite), WPS (Weighted Parametric Scoring), MEDOID (medoid measurement), COSSIM (cosine similarity), and NLCD (National Land Cover Database). Each algorithm was applied to time series of Landsat observations collected within two separate years at six locations around the world, to produce monthly (July 1 ± 15 days), seasonal (July 1 ± 45 days), and annual (July 1 ± 180 days) composite images free of cloud, cloud shadow, and snow/ice. By comparing the composite images to reference Landsat images acquired in the growing season (closest to July 1 within ±15 days) for each year, we evaluated the performance of the algorithms in preserving the spectral and spatial fidelity (hereafter referred to as spectral and spatial evaluation, respectively), as well as land cover classification and land change detection (hereafter referred to as application evaluation). The results demonstrated that no single algorithm outperformed all other algorithms in all the evaluations, but that performance depended on compositing intervals and cloud cover. For monthly composites, the MAX-RNB algorithm generally produced the best results in the spectral and application evaluations. For seasonal composites, the NLCD algorithm produced the best results in the spectral and application evaluations. For annual composites, the PAC algorithm produced the best results in the spectral evaluation and change detection, whereas BAP produced the best results in land cover classification. The BAP algorithm also produced the best results in the spatial evaluation for all the compositing periods. This study provides a comprehensive guidance for selecting the most appropriate image compositing algorithm for different Landsat-based applications.

### 1. Introduction

The Landsat program, jointly managed and operated by the National Aeronautics and Space Administration (NASA) and the U.S. Geological Survey (USGS), has provided 50 years of Earth observations at a medium resolution and a relatively high temporal frequency (8 days with two satellites) (Wulder et al., 2022). Landsat data have been used for a multitude of remote sensing applications, such as mapping forests (Hansen et al., 2013), water (Pekel et al., 2016), urban areas (Liu et al., 2020), general land cover (Chen et al., 2015; Gong et al., 2013), and land change (Brown et al., 2020; Jin et al., 2019) at various scales. An obstacle to applications of optical remote sensing is the presence of cloud, cloud shadow, and snow/ice, which obscure the land surface and

cause data gaps (Zhu et al., 2018). Creating seamless clear images at large scales by image compositing can mitigate the problem of data gaps and is therefore a critical processing step (Jin et al., 2023; Roy et al., 2010; White et al., 2014).

The goal of image compositing is to select the “best” available observations from a pool of observations that are free of cloud, cloud shadow, and/or snow/ice within a certain time interval for the same location. To create a pool of high-quality candidate observations to be selected for image compositing, accurate cloud and cloud shadow detection is the fundamental basis (Zhu et al., 2018). For example, overly aggressive cloud masking may result in an insufficient number of available candidate observations (Broich et al., 2011). On the other hand, omission errors in cloud and cloud shadow masks are inevitable,

\* Corresponding authors.

E-mail addresses: [shi.qiu@uconn.edu](mailto:shi.qiu@uconn.edu) (S. Qiu), [zhe@uconn.edu](mailto:zhe@uconn.edu) (Z. Zhu).

<https://doi.org/10.1016/j.rse.2022.113375>

Received 18 March 2022; Received in revised form 13 November 2022; Accepted 18 November 2022

Available online 6 December 2022

0034-4257/© 2022 Elsevier Inc. All rights reserved.

**Table 1**

The 10 Landsat compositing algorithms that were evaluated in this study. Surface reflectance are the inputs of the compositing criterion for most of the algorithms, except for MAX-NDVI and WELD, in which Top-Of-Atmosphere (TOA) reflectance and Brightness Temperature (BT) are the inputs.

| Category            | Algorithm | Data               | Period                                | Region                                       | Description   | Reference                    |
|---------------------|-----------|--------------------|---------------------------------------|--|---|------------------------------|
| Single rule         | MAX-NDVI  | MODIS              | \                                     | \  | Selects observations with maximum NDVI.   | (Holben, 1986)               |
|                     | MAX-RNB   | TM, ETM+, OLI      | Monthly, seasonal, and annual         | Six Landsat WRS-2 path/rows around the world | Selects observations with maximum RNB   | This study                   |
|                     | MED-NIR   | ETM+               | Seasonal                              | European Russia                              | Selects observations with median NIR.   | (Potapov et al., 2011)       |
| Multiple rules      | WELD      | ETM+               | Weekly, monthly, seasonal, and annual | CONUS  | Selects observations by either maximum NDVI or maximum BT   | (Roy et al., 2010)           |
|                     | BAP       | ETM+               | Seasonal                              | Canada, Saskatchewan, and Newfoundland       | Selects observations according to scores of sensors, DOY, cloud distance, and atmospheric opacity.          | (White et al., 2014)         |
|                     | PAC       | TM, ETM+, OLI      | Seasonal                              | Zambia                                       | Selects observations according to weighted scores of DOY, cloud distance, HOT, and spectral correlation.    | (Frantz et al., 2017)        |
|                     | WPS       | NASA HLS V1.2      | 10-day, monthly, and seasonal         | Germany                                      | Selects observations according to weighted scores of sensors, DOY, cloud distance, HOT, and cloud coverage. | (Griffiths et al., 2019)     |
|                     | MEDOID    | TM, ETM+           | Seasonal                              | A Landsat WRS-2 path/row in Australia        | Selects observations with minimum medoid to other data in spectral bands.                                   | (Flood, 2013)                |
| Spectral similarity | COSSIM    | TM                 | Seasonal                              | CONUS  | Selects observations with the most cosine similarity to other data in spectral bands.                       | (Nelson and Steinwand, 2015) |
|                     | NLCD      | TM, ETM+, OLI/TIRS | Seasonal                              | CONUS  | Selects observations with minimum distance to pseudo median data in spectral bands.                         | (Jin et al., 2023)           |

BT: Brightness Temperature. CONUS: Conterminous United States. DOY: Day Of Year. ETM+: Enhanced Thematic Mapper Plus. HLS: Harmonized Landsat-Sentinel. HOT: Haze Optimized Transformation. MODIS: Moderate Resolution Imaging Spectroradiometer. NIR: Near Infrared. SWIR: Short Wave Infrared. OLI: Operational Land Imager. RNB: Ratio of Near Infrared to Blue reflectance. TIRS: Thermal Infrared Sensor. TM: Thematic Mapper. WRS-2: Worldwide Reference System-2.

which may negatively affect the quality of the composite images (Griffiths et al., 2019). Therefore, an efficient image compositing algorithm should minimize the inclusion of cloud, cloud shadow, and/or snow/ice observations and best represent the targeted land surface (White et al., 2014).

Historically, image compositing has been applied to coarse resolution data (> 100 m) such as that collected by the Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, due to their high temporal frequency (Cihlar and Manak, 1994). Temporal compositing of coarse resolution data was typically based on a single band or index, such as the minimum of the blue band (Vermette and Vermeulen, 1999), minimum of the red band (Chuvieco et al., 2005), maximum of the thermal band (Chuvieco et al., 2005), maximum of Normalized Difference Vegetation Index (NDVI) (Holben, 1986), and by view angle constraint to select the best available observations (Huete et al., 2002; Wolfe et al., 1998).

Considering that image compositing involves selecting one observation from many candidate observations collected for the same location, compositing was rarely applied to medium resolution images (e.g., Landsat) before the era of free data (Hansen et al., 2008). Following the free-data policy of Landsat data in 2008 (Woodcock et al., 2008; Wulder et al., 2012; Zhu et al., 2019), the use of image compositing for medium resolution data became more relevant. Based on the criteria used for selecting the “best” available observation, we can categorize the Landsat image compositing algorithms into three major groups: single-rule-based, multiple-rules-based, and spectral-similarity-based (Table 1).

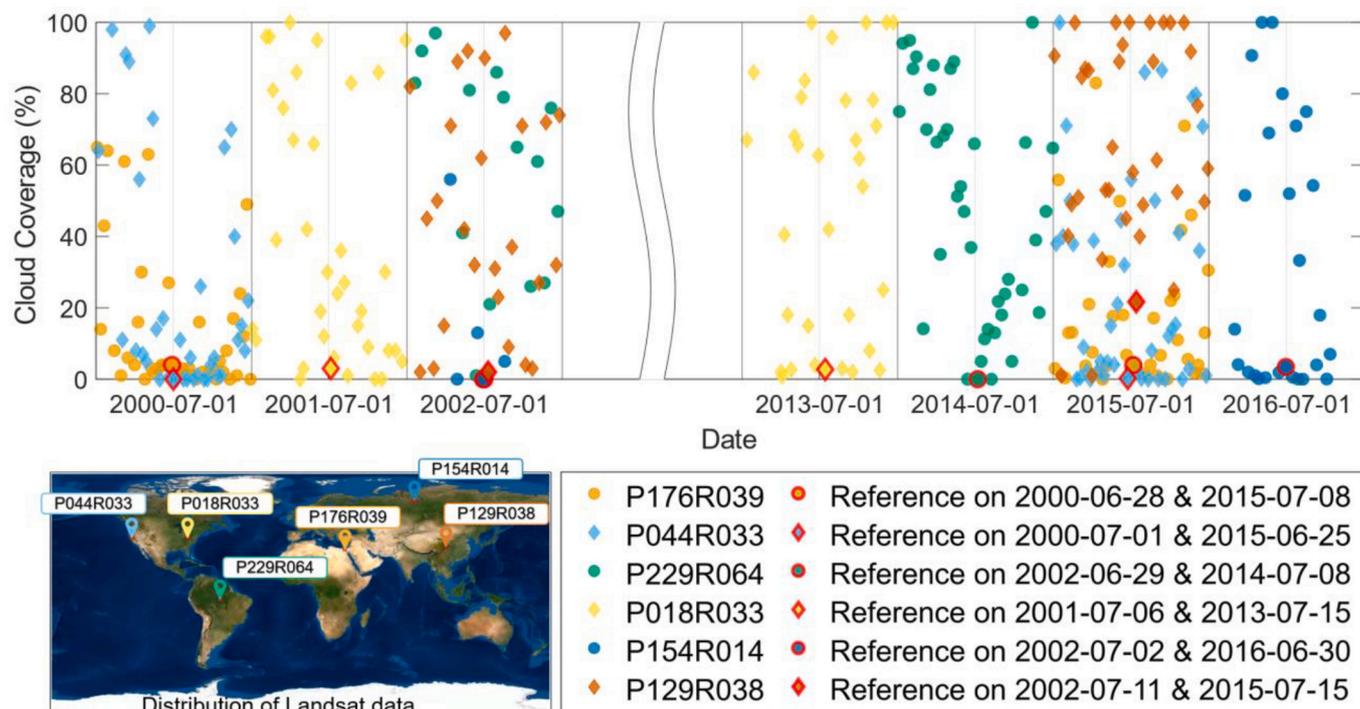
The single-rule-based compositing algorithms apply a single rule (e.g., maximum and median) to select the best available observation based on a spectral band or index. For example, the maximum NDVI-based compositing, that was developed for AVHRR data (Holben, 1986), has been successfully used to composite Landsat data to track cropland changes (Xu et al., 2018) and map spatiotemporal fire trends (Kato et al., 2020). Moreover, median Near Infrared (NIR)-based compositing has been used to create Landsat composite data to detect forest changes (Potapov et al., 2011). Single-rule-based compositing is easy to implement, and particularly useful for large-scale Landsat-based applications.

Multiple-rules-based compositing algorithms select observations

using multiple criteria. For example, the combined use of maximum temperature and maximum NDVI has shown good performance in compositing Landsat data at national scale to produce weekly, monthly, seasonal, and annual Landsat composite images for the conterminous United States (CONUS) (Roy et al., 2010). Further, Landsat data quality can be evaluated based on a scoring system that integrates multiple scores calculated from different rules. This kind of scoring-system compositing algorithm was initially introduced in Griffiths et al. (2013), in which a score for Day-Of-Year (DOY) and a score for cloud distance were proposed. Several other algorithms, such as the “best available pixel” method (White et al., 2014), the phenology-adaptive compositing algorithm (Frantz et al., 2017), and the weighted parametric scoring algorithm (Griffiths et al., 2019), have been developed with similar but more advanced scoring systems.

Spectral-similarity-based compositing algorithms select the observation with the highest spectral similarity among all candidate observations within a compositing period. The spectral similarity is often measured by metrics that are calculated from reflectance of the visible, NIR, and Short-Wave Infrared (SWIR) bands. For example, Flood (2013) used the multi-dimensional median, also known as the medoid (Small, 1990), to create seasonal reflectance images. Nelson and Steinwand (2015) applied the cosine similarity (Qian et al., 2004) to select the observation closest to a given target day in the growing season, which was successfully implemented to generate the LANDFIRE annual disturbance maps (Ryan and Opperman, 2013). Jin et al. (2023) composited Landsat images in multiple seasons to produce the 2019 National Land Cover Database (NLCD) by selecting the observation with the minimum Euclidean distance to a pseudo median value observation.

Even though many Landsat image compositing algorithms are available, few comprehensive comparison studies exist. To our knowledge, Ruefenacht (2016) compared various Landsat compositing algorithms, including maximum NDVI, univariate median, and model regression. However, this study did not include many of the recently developed Landsat image compositing algorithms and only focused on the application of canopy cover estimation based on Landsat 5 data. Moreover, the comparisons did not fully consider the compositing performance in different aspects, such as fidelity in the spectral, temporal,



**Fig. 1.** Landsat data over six path/rows. The cloud coverage was based on the scene cloud over provided in the Landsat metadata. The reference images closest to July 1, highlighted by red edges, were selected to assess the algorithm performances, and the remaining images were used to generate the composite images. ‘P’ and ‘R’ mean ‘Path’ and ‘Row’ of Landsat Worldwide Reference System-2 (WRS-2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and application domains, as well as the effect of varying image compositing periods. Additionally, even rather elaborate image compositing algorithms (Frantz et al., 2017; Griffiths et al., 2019) experience issues related to the omission of cloud, cloud shadow, snow/ice, and haze observations (Griffiths et al., 2019). Therefore, it is imperative to create more efficient and simpler image compositing algorithms that can provide better avoidance of cloud, cloud shadow, snow/ice, and haze observations.

In this study, we: (i) proposed a new single-rule-based compositing algorithm based on the maximum ratio of NIR to blue reflectance; (ii) evaluated the performance of 10 different Landsat compositing algorithms on varying compositing periods (i.e., monthly, seasonal, and annual) and different aspects (i.e., visual, spectral, spatial, and application evaluations); and (iii) provided guidance on the selection of compositing algorithms for different purposes.

## 2. Data

### 2.1. Landsat data

#### 2.1.1. Study area

Landsat Collection 2 data were used to study the different image compositing algorithms (Fig. 1). The Landsat Worldwide Reference System-2 (WRS-2) catalogs Landsat images by Path (afterward “P”) and Row (afterward “R”), and we selected a total of six Landsat path/rows with different cloud coverages, altitudes, and landscapes:

- Low cloud coverage: P176R039 is in central Africa, where the land cover is dominated by bright land surfaces (e.g., desert) with limited land changes; P044R033 is located in western United States and contains different kinds of land covers (e.g., cultivated land, artificial surface, forest, grassland, etc.) and change (e.g., agricultural practices, fire, climatic variability, etc.).

- Intermediate cloud coverage: P229R064 is located in the Amazon rainforest and contains many kinds of forest disturbances such as forest harvest and fire; P018R033, located in northeastern United States, is characterized by large areas of temperate deciduous forest and various kinds of land cover types (e.g., artificial surface, shrubland, cultivated land, etc.), and the main land change processes are forest harvest and insect infestation.
- High cloud coverage: P154R014, located in northern Russia, is mainly covered by wetlands and forests, with large areas of snow/ice coverage in winter and often experienced land changes caused by fire; P129R038 is located in southwestern China, where the main land cover types are cultivated land, artificial surface, grassland, and forest, and the land changes are mostly caused by agriculture and urban expansion.

#### 2.1.2. Two study epochs during the growing season

Two study epochs, the years of 2000 and 2015, were selected in each of the six WRS-2 path/rows (Fig. 1), which allows for evaluation of compositing for change detection and consideration of the effect of using different Landsat sensors. For example, Landsats 4–5 Thematic Mapper and normal Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images are available around 2000, whereas Landsat 7 ETM+ images after the failure of the Scan Line Corrector (SLC-off) and Landsat 8 Operational Land Imager (OLI)/Thermal Infrared Sensor (TIRS) images are available around 2015.

In each epoch and each path/row, a clear “reference” Landsat image was selected during the growing season (closest to July 1 and within  $\pm 15$  days) because growing season data are often used in remote sensing applications with less influence from snow/ice (Ju and Masek, 2016; Pflugmacher et al., 2014; Qiu et al., 2021; Tucker et al., 2004; Vicente-Serrano et al., 2008). As it was difficult to find a good clear Landsat reference image within a short period of July 1  $\pm 15$  days, particularly for the cloudy regions, we adjusted the study year to the closest years ( $\pm 2$  years) with enough valid reference pixels (Fig. 1). For each reference

**Table 2**

Specifications of spectral bands of Landsats 4–8 data. The spectral bands used in this study are highlighted in bold letters. Note that the surface reflectance of the spectral band was provided from the Landsat Collection 2 Level-2 product while the Top-Of-Atmosphere (TOA) reflectance and the Brightness Temperature (BT) were converted from the Landsat Collection 2 Level-1 product. The thermal band was used for calculating composite criterion only, and the thermal band with low gain was used for Landsat 7 data.

| Band ID | Band Name      | Landsats 4–5 TM Bands (μm)     | Landsat 7 ETM+ Bands (μm)      | Landsat 8 OLI/TIRS Bands (μm)                                    |
|---------|----------------|--------------------------------|--------------------------------|--|
| 1       | <b>Blue</b>    | <b>Band 1</b><br>(0.45–0.52)   | <b>Band 1</b><br>(0.45–0.52)   | <b>Band 1</b><br>(0.435–0.451)<br><b>Band 2</b><br>(0.452–0.512) |
| 2       | <b>Green</b>   | <b>Band 2</b><br>(0.52–0.60)   | <b>Band 2</b><br>(0.52–0.60)   | <b>Band 3</b><br>(0.533–0.590)                                   |
| 3       | <b>Red</b>     | <b>Band 3</b><br>(0.63–0.69)   | <b>Band 3</b><br>(0.63–0.69)   | <b>Band 4</b><br>(0.636–0.673)                                   |
| 4       | <b>NIR</b>     | <b>Band 4</b><br>(0.76–0.90)   | <b>Band 4</b><br>(0.77–0.90)   | <b>Band 5</b><br>(0.851–0.879)                                   |
| 5       | SWIR1          | <b>Band 5</b><br>(1.55–1.75)   | <b>Band 5</b><br>(1.55–1.75)   | <b>Band 6</b><br>(1.566–1.651)                                   |
| 7       | SWIR2          | <b>Band 7</b><br>(2.08–2.35)   | <b>Band 7</b><br>(2.09–2.35)   | <b>Band 7</b><br>(2.107–2.294)                                   |
| 6       | <b>Thermal</b> | <b>Band 6</b><br>(10.40–12.50) | <b>Band 6</b><br>(10.40–12.50) | <b>Band 10</b><br>(10.60–11.19)                                  |
|         |                |                                |                                | Band 11<br>(11.50–12.51)   |
|         |                |                                | Band 8<br>(0.52–0.90)          | Band 8<br>(0.503–0.676)  |

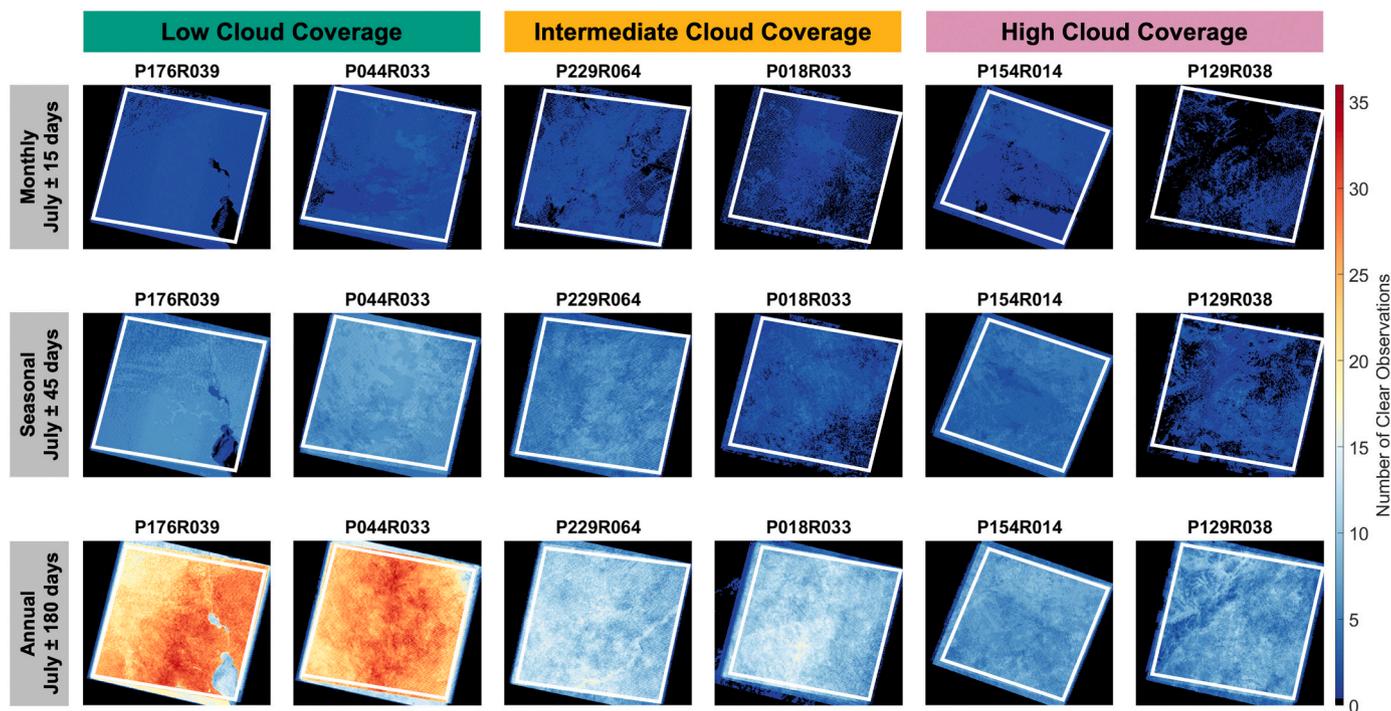
ETM+: Enhanced Thematic Mapper Plus. NIR: Near Infrared. OLI/TIRS: Operational Land Imager/Thermal Infrared Sensor. SWIR: Short Wave Infrared. TM: Thematic Mapper.

image, we manually screened out clouds, cloud shadows, and haze, with help of the Fmask version 4.6 algorithm (Qiu et al., 2019), and only the remaining clear pixels were used in our analysis. All the reference Landsat images were excluded as “candidate” images for image compositing and retained for comparison.

2.1.3. Landsat products

All Landsat Collection 2 images with cloud cover less than 100%, including Level-2 surface reflectance and Level-1 Digital Numbers (DN) products, were downloaded from the USGS Earth Explorer (<https://earthexplorer.usgs.gov>; last access on November 3, 2022). Note that Landsats 4–7 Tier 2 were excluded because of their relatively low radiometric and positional quality (Zhang et al., 2022). The Level-2 product includes the surface reflectance data of Landsats 4–7 generated by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) version 3.2.1 (Masek et al., 2006), and those of Landsat 8 generated by the Landsat Surface Reflectance Code (LaSRC) version 1.3.0 (Vermette et al., 2016). Additionally, we converted the Level-1 DN products to Top-Of-Atmosphere (TOA) reflectance and Brightness Temperature (BT) (Chander et al., 2009), which were used as the criterion in the Landsat compositing algorithms that rely on TOA reflectance and/or BT, such as MAX-NDVI and WELD (Table 1). The mask of clouds, cloud shadows, and snow/ice was generated based on the Level-1 data using the Fmask version 4.6 algorithm (Qiu et al., 2019; Zhu et al., 2015a; Zhu and Woodcock, 2012). Pixels with unrealistic surface reflectance data (i.e., less than 0 or larger than 1) were also excluded in our analysis. In this study, we only focused on the performance of Landsat compositing algorithms on the surface reflectance of the six optical bands, which include blue, green, red, NIR, SWIR1, and SWIR2 (Table 2).

Fig. 2 shows the total number of clear observations free of cloud, cloud shadow, and snow/ice that are available for monthly (July 1 ± 15 days



**Fig. 2.** Total number of available clear observations (free of cloud, cloud shadow, and snow/ice) for monthly (July 1 ± 15 days), seasonal (July 1 ± 45 days), and annual (July 1 ± 180 days) image compositing for the 2015 epoch year at six Landsat path/rows. The white polygons are the Region-Of-Interest (ROI) used for this analysis.

**Table 3**

Land cover and land change training data. The definition of land cover classes is based on the GlobeLand30 product (Chen et al., 2015).

| Land cover/ Land change | Number of training pixels |          |          |          |          |          |
|-------------------------|---------------------------|----------|----------|----------|----------|----------|
|                         | P176R039                  | P044R033 | P229R064 | P018R033 | P154R014 | P129R038 |
| Cultivated land         | 116,247                   | 227,616  | 0        | 28,968   | 0        | 384      |
| Forest                  | 0                         | 73,797   | 664,514  | 88,021   | 64,560   | 67,523   |
| Grass/Shrubland         | 248                       | 244,076  | 17,185   | 6428     | 157      | 0        |
| Wetland                 | 2392                      | 23,568   | 10,763   | 105      | 10,968   | 0        |
| Water                   | 7150                      | 39,648   | 17,819   | 9236     | 31,434   | 18       |
| Tundra                  | 0                         | 0        | 0        | 0        | 2115     | 0        |
| Artificial surface      | 54,921                    | 101,792  | 104      | 26,434   | 1060     | 240      |
| Bareland                | 342,483                   | 387      | 0        | 0        | 0        | 0        |
| Changed                 | 38,264                    | 74,978   | 44,763   | 31,234   | 58,056   | 5129     |
| Stable                  | 529,752                   | 718,658  | 730,539  | 218,343  | 189,158  | 75,018   |

days), seasonal (July 1 ± 45 days), and annual (July 1 ± 180 days) image composites for the 2015 epoch year. The availability of Landsat data and the frequency of cloud/snow cover could affect the number of available clear observations, but it is obvious that the longer the compositing period, the more candidate observations. In addition, to avoid the potential effects of small shifts in the footprint of each Landsat image, we created a Region-Of-Interest (ROI) for each Landsat path/row, which included the footprint of all the Landsat images, that is, only observations within the ROIs were analyzed (see white polygons in Fig. 2).

## 2.2. Land cover and land change training data

Landsat composite data have been widely used to map land cover and land change (Jin et al., 2023; Xu et al., 2018). To explore the performance of different image compositing algorithms on such applications, we manually collected training data of land cover and land change based on the Landsat reference images (see Section 2.1.2 for details), with aid of the high-resolution images from Google Earth and the 2010 30-m global land cover product from GlobeLand30 (Chen et al., 2015). The training data selection was performed opportunistically based on the availability of high-resolution images from Google Earth and careful comparison between the Landsat reference images acquired in two epochs. We interpreted multiple polygons of training data for each land cover (eight classes) and land change categories (two classes – “changed” and “stable”) for each Landsat path/row, and the pixels within the interpreted polygons were used as the training data (Table 3).

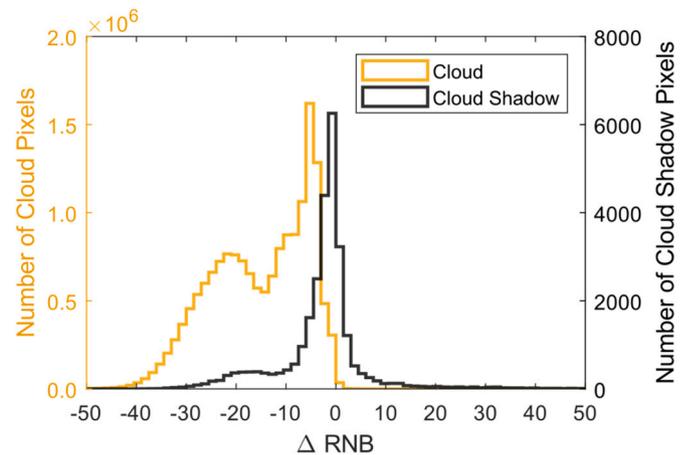
## 3. Landsat image compositing algorithms

A total of 10 Landsat image compositing algorithms were evaluated in this study (Table 1), and each of them is briefly introduced in this section. The final goal of all algorithms is to create clear surface reflectance images for six optical bands (Table 2) on a monthly (July 1 ± 15 days), seasonal (July 1 ± 45 days), and annual (July 1 ± 180 days) basis. Although the MAX-NDVI and WELD algorithms need TOA reflectance and BT images to make decisions in image compositing, other algorithms rely on surface reflectance only (Table 1). To make our comparison fair and consistent, clouds, cloud shadows, and snow/ice were excluded based on the most recent Fmask version 4.6 algorithm for all the image compositing algorithms (Qiu et al., 2019; Zhu et al., 2015a; Zhu and Woodcock, 2012) (see Section 2.1.3 for details). Moreover, the default parameters as specified in the associated papers were used for all the image compositing algorithms compared here.

### 3.1. Single-rule-based compositing algorithms

#### 3.1.1. Maximum NDVI algorithm

The maximum NDVI algorithm (hereafter called ‘MAX-NDVI’) was originally proposed to composite AVHRR data for studying the dynamic processes of terrestrial vegetation. The algorithm selects the observation



**Fig. 3.** Histogram of the differences of surface-reflectance-derived RNB. The  $\Delta$ RNB values were calculated based on the difference of cloud or cloud shadow observation RNB values minus the corresponding clear observation RNB values in the reference images that are collected within  $\pm 8$  days for the same locations. We analyzed a total of 16,319,856 pixels (99.58% over land and 0.42% over water), of which 16,291,995 of them are cloud pixels and 27,861 of them are cloud shadow pixels. Moreover, 99.29% of the total pixels (cloud or cloud shadow pixels) are having  $\Delta$ RNB < 0, and 99.34% of the cloud pixels and 73.05% cloud shadow pixels are having  $\Delta$ RNB < 0. RNB: Ratio of NIR to Blue reflectance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with the highest NDVI from all observations within the composite period, by eliminating cloud (and potentially snow/ice) contamination (Holben, 1986). In this study, Landsat observations with maximum TOA-derived NDVI (Eq. 1) were chosen to generate the surface reflectance composite data.

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

where,

$\rho_{red}$  indicates the TOA reflectance of the red band.

$\rho_{nir}$  indicates the TOA reflectance of the NIR band.

#### 3.1.2. Maximum RNB algorithm

A new Landsat image compositing algorithm based on the maximum Ratio of NIR to Blue reflectance (RNB) is proposed in this study (MAX-RNB). The RNB value can be calculated based on the Landsat surface reflectance according to Eq. 2. For the same location, the higher the RNB value, the more likely the observation is clear. This happens because the presence of clouds (or snow/ice and haze) will increase both the blue and NIR band, making their ratio close to 1, but for clear observations, the RNB values are usually much higher than 1 (particularly after atmospheric correction). The presence of cloud shadow will also decrease

the RNB values, as cloud shadow will usually substantially reduce the NIR band reflectance but only slightly reduce the blue band reflectance due to the much stronger Rayleigh scattering in the blue band that provides illumination in all angles in shadowed areas (Luo et al., 2008). It is worth noting that though the atmospheric correction can correct the Rayleigh scattering effect for the atmosphere, the shadowed areas will be still lightened more in the blue band than they are in the NIR band because of the higher proportion of the diffuse light (caused by Rayleigh scattering) in the shorter wavelength. Thus, selecting Landsat observations with the maximum RNB value for image compositing could exclude not only remaining clouds (or snow/ice and haze) but also cloud shadows. We decided to use surface reflectance instead of TOA reflectance as inputs for the MAX-RNB algorithm because it requires no extra data downloading/preprocessing and can achieve similar results (see Section 6.4 for details).

$$RNB = \frac{\rho_{nir}}{\rho_{blue}} \quad (2)$$

where,

$\rho_{blue}$  indicates the surface reflectance of the blue band.

$\rho_{nir}$  indicates the surface reflectance of the NIR band.

To better demonstrate this unique criterion, we manually collected reference data of cloud and cloud shadow from different times and locations, and compared their RNB values with the ones calculated from the corresponding clear reference image that are collected within  $\pm 8$  days. The histogram of the differences of surface-reflectance-derived RNB for the reference data showed that the presence of cloud and their shadows will decrease the RNB values for most of the observations (Fig. 3). Although the presence of cloud and cloud shadow may occasionally increase RNB values (the tails on the right with  $\Delta RNB > 0$ ), these cases are relatively rare (only 0.71% of the cloud and cloud shadow pixels analyzed in Fig. 3).

### 3.1.3. Median NIR algorithm

The median NIR compositing algorithm (hereafter called 'MED-NIR') selects the Landsat observation with the median value of NIR band surface reflectance of all observations within the composite period (Potapov et al., 2011). This method is used because the NIR band is sensitive to both clouds (extremely bright) and cloud shadows (extremely dark). Moreover, (Potapov et al., 2011) demonstrated that this algorithm had better visual performance and the least noise over boreal forested areas, compared to maximum NDVI, median red band, or median SWIR1 band.

## 3.2. Multiple-rules-based compositing algorithms

### 3.2.1. WELD algorithm

The Web-Enabled Landsat Data (WELD) compositing algorithm was originally developed based on Landsat TOA reflectance to create CONUS-wide Landsat image composites (Roy et al., 2010). The WELD algorithm mainly relies on maximum TOA-derived NDVI and maximum BT criteria. The maximum BT criterion is added for unvegetated pixels such as soil, water, and permanent snow/ice, for which NDVI is not a good indicator. A pixel's temperature is a good indicator of cloud because clouds are usually colder than Earth's surfaces (Roy et al., 2010). The unvegetated pixels are identified using empirical spectral tests (i.e., NDVI  $< 0.09$  and reflectance of SWIR2 band  $< 0.048$ ). For places that are vegetated, the maximum NDVI approach will be used for image compositing. In addition to maximum NDVI and maximum BT, other data quality tests, such as data fill, saturation, and cloud and cloud shadow, are also considered in a priority queue (Roy et al., 2010). In this study, the component of cloud detection was updated based on the

**Table 4**

Scoring weights in the WPS algorithm.

| Score Name     | Monthly | Seasonal (Annual) |
|----------------|---------|-------------------|
| Cloud distance | 1.0     | 1.0               |
| DOY            | 0.8     | 1.0               |
| Sensor         | 0.5     | 0.5               |
| Cloud coverage | 0.5     | 0.75              |
| HOT            | 1.0     | 1.0               |

DOY: Day-Of-Year; HOT: Haze-Optimized Transformation.

Fmask version 4.6 algorithm (Qiu et al., 2019) as we did for the other compositing algorithms.

### 3.2.2. Best available pixel algorithm

The Best Available Pixel (BAP) algorithm selects the "best" Landsat observation according to scores for sensor, DOY, cloud distance, and atmospheric opacity (White et al., 2014). The sensor score was designed to reduce the effect of scan lines in Landsat 7 SLC-off data (a score of 0.5 is assigned to the images with SLC-off as opposed to 1 under normal conditions). To optimize for image seasonality, a DOY score is calculated based on a Gaussian function with a standard deviation of 38 days and a mean value equal to the target date (i.e., July 1) (Eq. 3), and is scaled to a value between 0 and 1 based on the maximum value of all images (Griffiths et al., 2013; White et al., 2014). In addition, to eliminate the potential effect from clouds and cloud shadows that are omitted by the Fmask algorithm (Qiu et al., 2019), a cloud distance score is based on a sigmoidal function of distance to any cloud or cloud shadow (Eq. 4), which assumes that an observation is more unreliable if it is closer to a cloud or cloud shadow than a minimum required distance (i.e., 50 pixels in BAP) (Griffiths et al., 2013; White et al., 2014). Further, the BAP algorithm computes an atmospheric opacity score based on a sigmoidal function to reduce the effects from the aerosol optical thickness (Eq. 5). The available opacity range of this function varies between 0.2 and 0.3, in which the observations with an opacity value  $> 0.3$  will be directly excluded and the ones  $< 0.2$  will be assigned an opacity score of 1. However, this opacity score is only applied to Landsats 4–7 data, for which the atmospheric opacity band is generated by the LEDAPS atmospheric correction algorithm (Vermote and Saleous, 2007). For Landsat 8 data, an opacity score of 1 will be directly assigned in this study because the Landsat 8 surface reflectance product, generated by the LaSRC atmospheric correction algorithm (Vermote et al., 2016), does not provide this kind of band indicating the opacity. Note that Landsat 8 data are less likely to be affected by the aerosol optical thickness because the new cirrus and ultra-blue bands improve cloud detection (Qiu et al., 2019; Zhu et al., 2015a) and atmospheric correction (Vermote et al., 2016).

$$\text{Score for DOY} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \quad (3)$$

where,

$\mu$  and  $\sigma$  are the mean and the standard deviation of all the DOY values, respectively ( $\mu$  is equal to the DOY of the target date in this study, like July 1, and  $\sigma$  can be customized further).

$t$  is the DOY of the Landsat imagery.

Score for cloud distance =

$$1/(1 + \exp(-0.2 \bullet (\min(d, d_{req}) - (d_{req} - d_{min})/2))) \quad (4)$$

where,

$d$  is the Euclidean distance closest to any clouds or cloud shadows.

$d_{min}$  is the minimum distance (i.e., 0 pixel in BAP), less than which

**Table 5**  
Scoring weights in the PAC algorithm.

| Score Name           | Weight |
|----------------------|--------|
| DOY                  | 1.0    |
| Cloud distance       | 0.2    |
| HOT                  | 0.2    |
| Spectral correlation | 0.2    |

DOY: Day-Of-Year; HOT: Haze-Optimized Transformation.

pixels will be assigned a score of 0.

$d_{req}$  are the minimum required distance (i.e., 50 pixels in BAP), beyond which pixels will be assigned a score of 1.

$$Score\ for\ opacity = 1 - 1 / (1 + \exp(-0.2 \bullet (\min(o, o_{max}) - (o_{max} - o_{min}) / 2))) \quad (5)$$

where,  $o$  is the opacity value (only available for LEDAPS Landsats 4–7 surface reflectance products).

$o_{min}$  and  $o_{max}$  are the minimum ( $o_{min} = 0.2$ ) and the maximum opacity value ( $o_{max} = 0.3$ ), respectively (pixels less than  $o_{min}$  will have a score of 1 and pixels beyond  $o_{max}$  will be directly discarded).

### 3.2.3. Weighted parametric scoring algorithm

Based on the success of the first scoring Landsat image compositing algorithm (Griffiths et al., 2013), Griffiths et al. (2019) further proposed a Weighted Parametric Scoring (WPS) algorithm optimized for intra-annual composites (i.e., 10-day, a month, and a season). The WPS algorithm selects the observation with the highest total score, which is calculated based on a weighted linear combination of all the individual scores in Table 4 (Griffiths et al., 2019). The WPS algorithm was originally designed to composite NASA Harmonized Landsat-Sentinel data (Claverie et al., 2018), but in this study, we implemented it to process Landsat data only. Moreover, the WPS algorithm was designed for image compositing with a period less than a season, and we use the same protocols for seasonal composites to create our annual image composites.

In the WPS algorithm, the DOY score (Eq. 3) is calculated using a standard deviation of DOY values of 5 for the monthly composites and 12 for the seasonal composites. The cloud distance score (Eq. 4) is based on an optional minimum required cloud distance of 100 pixels. Because more spectral bands are available (i.e., red edge bands) in the Sentinel-2 data, the original WPS algorithm gives a higher sensor score to Sentinel-2 data (i.e., 1.0 for Sentinel-2 vs. 0.8 for Landsat 8). In this study, we did not use Sentinel-2 data, but because the Landsat 7 images SLC-off images are considered inferior to other images, a sensor score of 0.8 was assigned to SLC-off data, and the sensor score of normal Landsat data was set to 1. Furthermore, a score for cloud-free coverage percentage (range between 0 and 1) and a score for the Haze-Optimized Transformation (HOT) (Zhang et al., 2002) based on a sigmoidal function (Eqs. 6–7) were used to prevent low-quality observations (i.e., high cloud coverage, haze, and thin cloud).

$$Score\ for\ HOT = 1 / \left( 1 + \exp\left(\frac{10}{0.02} \bullet (HOT + 0.075)\right) \right) \quad (6)$$

$$HOT = \rho_{blue} - 0.5 \bullet \rho_{red} - 0.08 \quad (7)$$

where,

$\rho_{blue}$  and  $\rho_{red}$  are the surface reflectance of blue band and red band, respectively.

### 3.2.4. Phenology adaptive compositing algorithm

The Phenology Adaptive Composite (PAC) algorithm was proposed to create MODIS or Landsat composites (Frantz et al., 2017). In this study, we only implemented the PAC algorithm for Landsat data. This algorithm selects the observation by a weighted linear combination of multiple scores, such as DOY, cloud distance, HOT, and spectral correlation (Table 5).

The PAC algorithm is capable of dynamically adjusting the observation selection process according to the land surface phenology based on scoring functions (Frantz et al., 2017). For example, the DOY score is calculated from a two-tailed Gaussian function with three temporal land surface phenology parameters (Eqs. 8–10) that are 1) DOY of peak growing season with maximum vegetation growth ( $\mu_p$ ), 2) DOY of the end of the growing season with a decreasing green stock ( $\mu_e$ ), and 3)

DOY of the minimum greenness with the absolute minimum of vegetation development between two peaks ( $\mu_m$ ). The cloud distance score and the HOT score are calculated according to Eqs. 11 & 12, respectively. Although the previous score functions account for the majority of factors affecting the composite, other factors remain, such as sensor anomalies, residual misregistration, short-term changes caused by flooding and active fire, and missed clouds/shadows. Thus, the PAC algorithm integrates a new spectral correction score (Eq. 13) based on the mean absolute correlation between a given observation and all other observations in the spectral bands (Frantz et al., 2017). Note that the original PAC algorithm also has a view angle score to reduce the weight of off-nadir observations, which is mainly designed for sensor with large view zenith angles, such as MODIS data (Frantz et al., 2017), and this function is not used in this study because of the limited changes in the view zenith angles from the single path/row Landsat images.

$$Score\ for\ DOY = \begin{cases} S_e \bullet \exp\left(\frac{-(t - \mu_e)^2}{2\sigma_e^2}\right), & (t < \mu_e) \\ S_e \bullet \exp\left(\frac{-(t - \mu_e)^2}{2\sigma_e^2}\right), & (t \geq \mu_e) \end{cases} \quad (8)$$

$$\sigma_l = \frac{|\mu_p - \mu_e|}{\sqrt{-2 \bullet \log(s_p/s_e)}} \quad (9)$$

$$\sigma_r = \frac{|\mu_m - \mu_e|}{\sqrt{-2 \bullet \log(s_m/s_e)}} \quad (10)$$

where,

$\mu_p$  is the DOY of peak vegetation growth (default  $\mu_p = 25$ ).

$\mu_e$  is the DOY of the end of the growing season (default  $\mu_e = 174$ ).

$\mu_m$  is the DOY of minimum vegetation development between two peaks (default  $\mu_m = 245$ ).

$s_p$ ,  $s_e$ , and  $s_m$  are the parameters defining the width of the Gaussian curves ( $\sigma_l$  and  $\sigma_r$ ), and  $s_e$  also indicates the amplitude of the Gaussian functions (default  $s_p = 0.01$ , default  $s_e = 1$ , and default  $s_m = 0.01$ ).

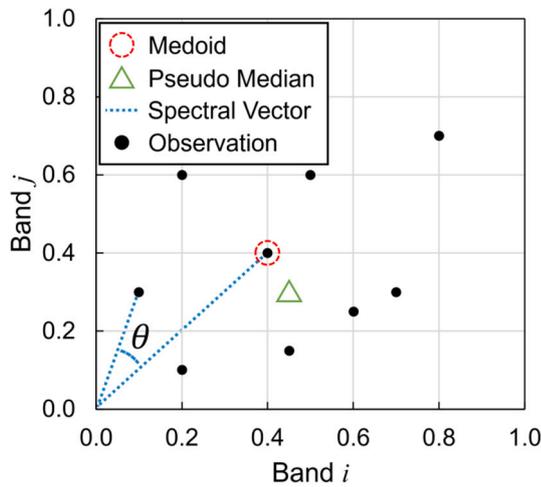
$\sigma_l$  and  $\sigma_r$  are left- and right-tail width for the Gaussian DOY score function.

$$Score\ for\ cloud\ distance = 1 / (1 + \exp(-10/d_{req} \bullet (d - d_{req}/2))) \quad (11)$$

where,

$d$  is the Euclidean distance to any clouds or cloud shadows.

$d_{req}$  is the minimum required cloud distance (default  $d_{req} = 100$  pixels), beyond which the pixel is assumed to be clear sky (cloud



**Fig. 4.** Illustration of spectral-similarity compositing algorithms in two bands. The two blue dashed lines indicate the spectral vectors and the angle between them ( $\theta$ ) indicates the cosine similarity in the COSSIM algorithm. The medoid with red dashed circle presents the MEDOID algorithm. The pseudo median with green triangle indicates the NLCD algorithm. The data are randomly generated for explaining the basic compositing ideas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distance score of 1).

$$\text{Score for } HOT = 1 / \left( 1 + \exp \left( \frac{10}{0.02} \cdot (HOT + 0.015) \right) \right) \quad (12)$$

where,

$HOT$  means the Haze Optimized Transformation (Zhang et al., 2002), which is calculated by Eq. 7.

$$\text{Score for correlation} = 1 / (1 + \exp(-10 / (0.5 \cdot 2/3) \cdot (r_i - 0.5 \cdot 4/3))) \quad (13)$$

where,

$r_i$  is the mean absolute correction between the  $i$ th observation and all other observations (Frantz et al., 2017).

### 3.3. Spectral-similarity compositing algorithms

#### 3.3.1. Cosine similarity algorithm

The cosine similarity compositing algorithm (COSSIM) selects observations according to the cosine similarity value of the six spectral bands between any two valid observations (Eq. 14) (Nelson and Steinwand, 2015). For each individual Landsat pixel, the COSSIM algorithm selects no more than five valid observations acquired closest to the target date (i.e., July 1), and iteratively computes the cosine similarity of each observation to any other observations based on the spectral vectors (see the blue dashed lines in Fig. 4). The pixel with the lowest mean cosine value is chosen.

$$\text{Cosine similarity between } \mathbf{A} \text{ and } \mathbf{B} = 1 - \cos(\theta) = 1 - \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|} \quad (14)$$

where,

$\mathbf{A}$  and  $\mathbf{B}$  are the six spectral band vectors of the observation from two different dates.

$\theta$  means the angle between  $\mathbf{A}$  and  $\mathbf{B}$ .

$\|\bullet\|$  denotes denote the Euclidean norm.

#### 3.3.2. Medoid algorithm

The medoid compositing algorithm (MEDOID) is based on the medoid measurement in the six spectral bands (Flood, 2013). This

**Table 6**

Metrics for comparing composite images and reference images.

| Domain      | Metrics  |
|-------------|--|
| Spectral    | R-square, RMSE, slope and intercept of linear regression |
| Spatial     | Semivariance difference                                  |
| Application | Disagreement of land cover and land change maps          |

algorithm uses a measure of the “center” of a multi-variate set of points (Struyf et al., 1997), in which the selected observation point minimizes the sum of the Euclidean distances to all other points (see the red dashed circle in Fig. 4) as shown in Eq. 15.

$$\text{Medoid}(x) = \underset{y \in x}{\operatorname{argmin}} \sum_{i=1}^n \|y - x_i\| \quad (15)$$

where,

$x$  indicates the  $n$  observations in the six Landsat spectral bands,

$i$  is the  $i$ th observation.

$\|\bullet\|$  denotes denote the Euclidean norm.

#### 3.3.3. NLCD algorithm

The National Land Cover Database (NLCD) compositing algorithm selects the “best” observation using the median value of six spectral bands and was used for creating the NLCD 2019 product (Jin et al., 2023). The idea is similar to the MEDOID algorithm (Flood, 2013), but the NLCD algorithm first creates a pseudo median-value observation based on all valid Landsat observations for each spectral band, and then selects the observation which has the lowest Euclidean distance to the pseudo median-value observation (Eq. 16) (see the green triangle in Fig. 4).

$$\text{Distance to median} = \sum_{i=1}^7 (\rho_i - \tilde{\rho}_i)^2 \quad (16)$$

where,

$\rho_i$  is the surface reflectance of the  $i$ th band ( $i = 1, 2, 3, 4, 5,$  and  $7$  representing to the blue, green, red, NIR, SWIR1, and SWIR2 bands in Table 2, respectively).

$\tilde{\rho}_i$  is the median surface reflectance of the  $i$ th band.

## 4. Evaluation metrics

The purpose of this study is to evaluate the performance of different image compositing algorithms under different conditions. The composite periods were defined by adjusting the date range around the target date (i.e., July 1) for each study year, such that  $July 1 \pm 15$  days,  $July 1 \pm 45$  days,  $July 1 \pm 180$  days are the monthly, seasonal, and annual composite periods, respectively. The growing season Landsat reference images acquired on the date closest to July 1 were selected as the reference data (see Section 2.1.2 for details).

We evaluated 10 different image compositing algorithms based on both qualitative comparison (see the results in Section 5.1) and quantitative analysis in the spectral, spatial, and application (e.g., land cover classification and change detection) domains (Table 6; Sections 4.1–4.3). Note that the BAP algorithm discards observations with spectral opacity  $> 0.3$  (Section 3.2.2) (White et al., 2014), whereas other algorithms do not. To conduct a fair comparison, only pixels, which all compositing algorithms considered valid, were used in our analysis.

### 4.1. Spectral evaluation

The spectral performance was evaluated using the R-square, Root Mean Square Error (RMSE), slope and intercept between the composite data and the reference data. For example, Fig. 5 shows the density plot between the seasonal composite data, generated by the MAX-NDVI algorithm, versus the reference data for each spectral band, which allows

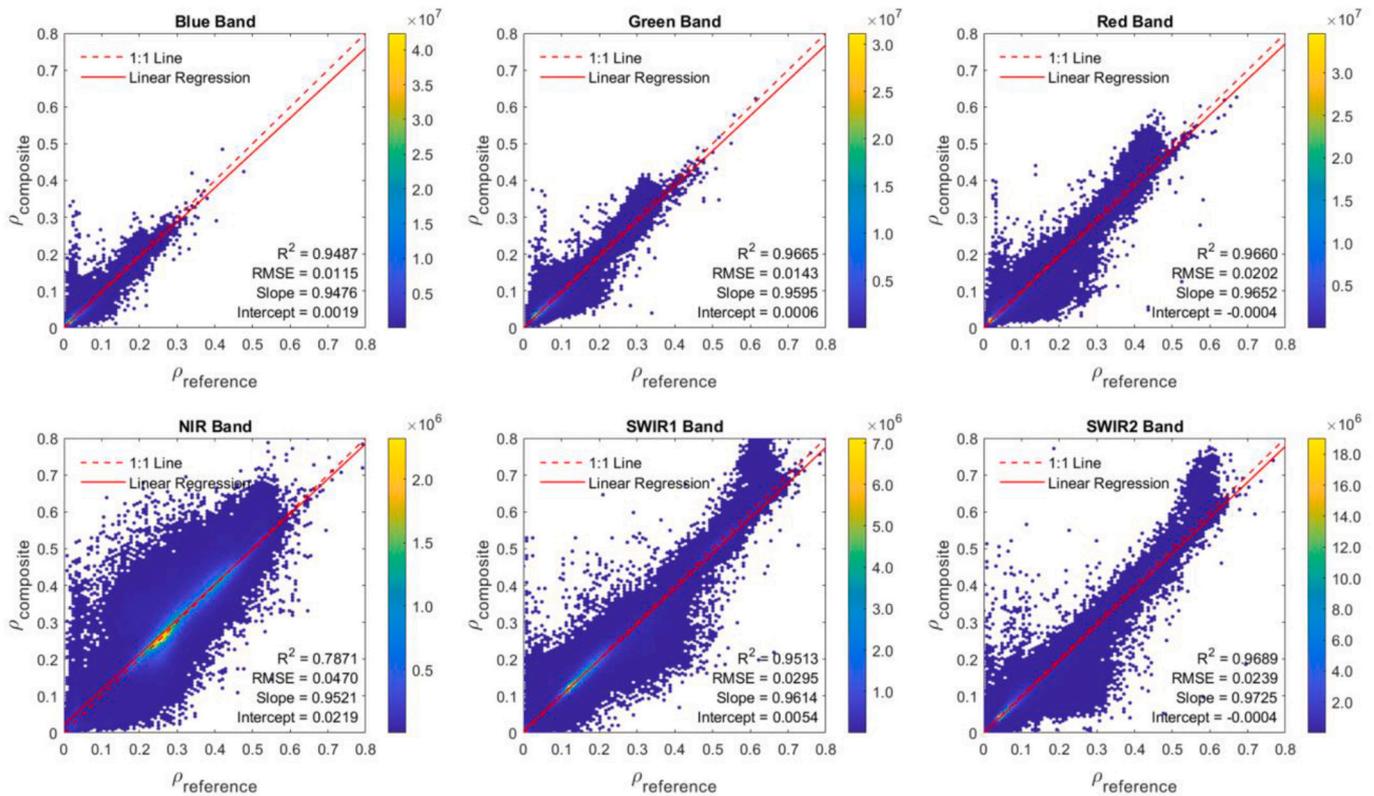


Fig. 5. Illustration of spectral comparison between the seasonal Landsat composite images derived from the MAX-NDVI algorithm (compositing period: July 1 ± 45 days) and the growing season Landsat reference images (closest to July 1). The maximum NDVI criterion was based on TOA reflectance. The R-square with higher value, RMSE with lower value, slope closer to 1:1 line, and intercept closer to 0 indicate a better Landsat compositing algorithm.

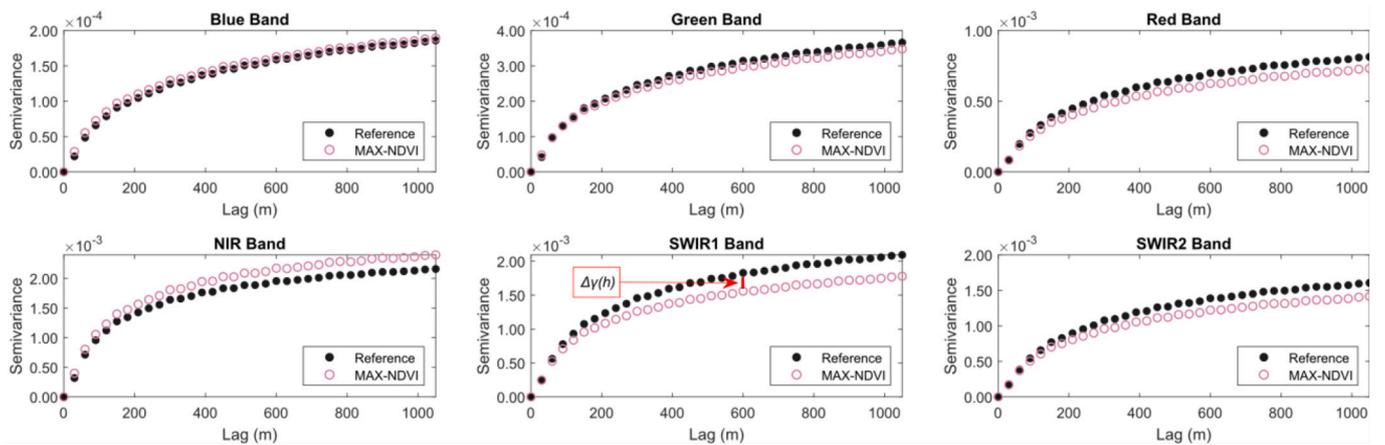


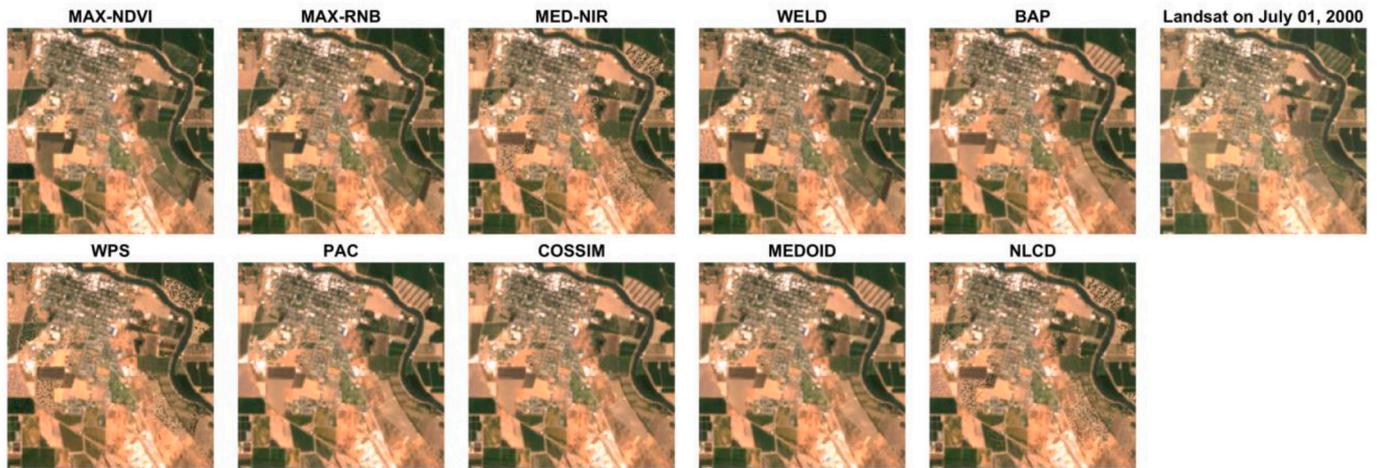
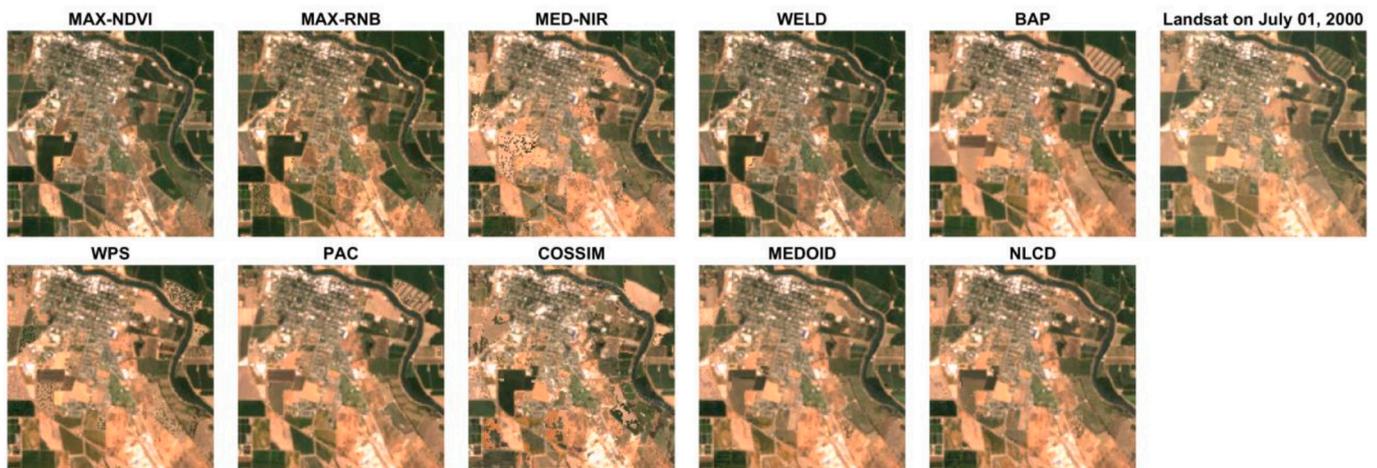
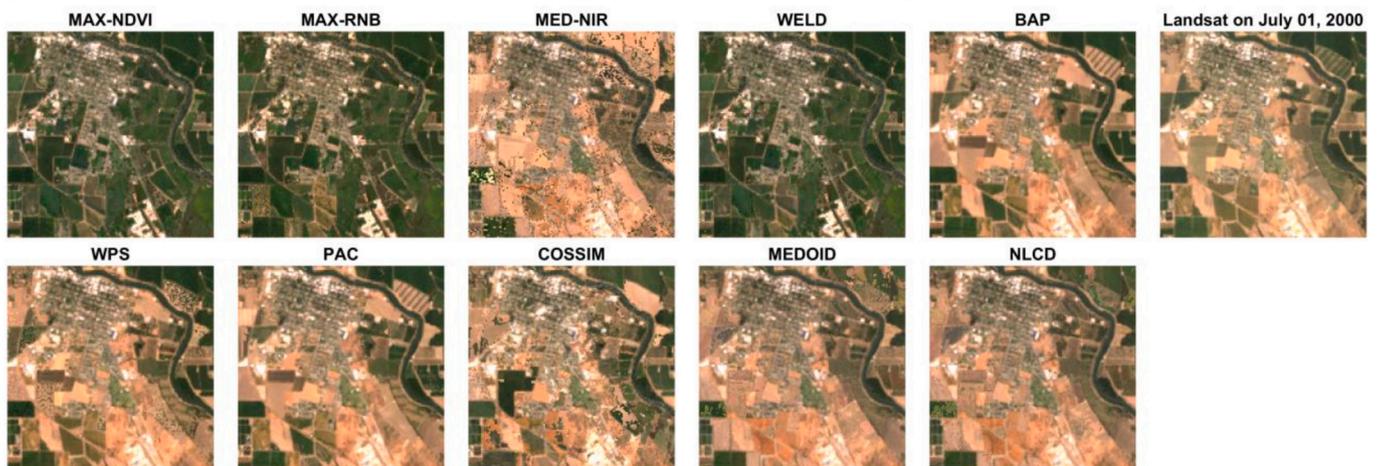
Fig. 6. Illustration of spatial comparison of semivariance between the seasonal composites derived from MAX-NDVI algorithm (compositing period: July 1 ± 45 days) and the growing season Landsat reference images (closest to July 1). The maximum NDVI criterion was based on TOA reflectance. The  $\Delta\gamma(h)$  indicates the semivariance difference between the composite data and the reference data at lag distance  $-h$ .

for calculation of R-square, RMSE, slope, and intercept values. Usually, the higher the R-square value, the lower the RMSE value, the closer to the 1:1 line; and the closer to 0 for the intercept value, the higher the spectral fidelity of the algorithm. Finally, the mean of the absolute value of these metrics calculated from the six spectral bands was used to evaluate the overall algorithm performance quantitatively.

#### 4.2. Spatial evaluation

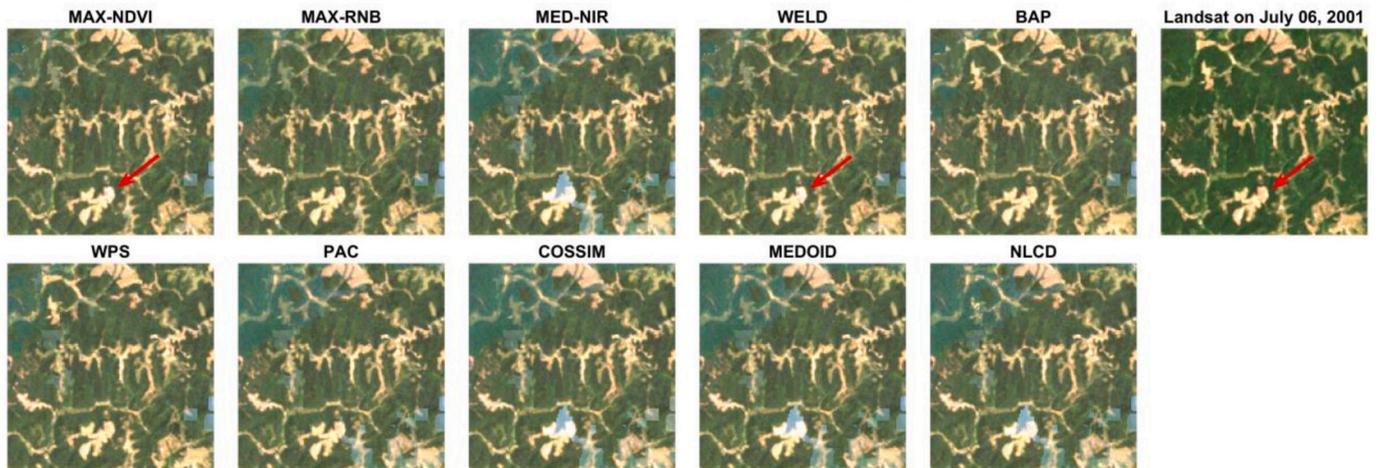
The semivariogram (Eq. 17) was used to characterize the spatial structure and measure the spatial variation in remote sensing imagery

(Curran, 1988). In this study, we varied the lag distance  $h$  between 0 and 1050 m at 30 m steps to analyze the semivariogram of each spectral band of Landsat composite data and that of the reference data, respectively. The semi-variograms were estimated based on a random selection of 1% of the pixels in each spectral band to reduce computation cost. If the shape of semivariogram of a composite Landsat image was more similar to that of the reference data, it was expected to have higher spatial fidelity (Fig. 6). We calculated the absolute difference of semivariance between the composite data and the reference data at each lag distance (see SWIR1 band in Fig. 6). The mean of the absolute differences of all spectral bands was used to evaluate the algorithm

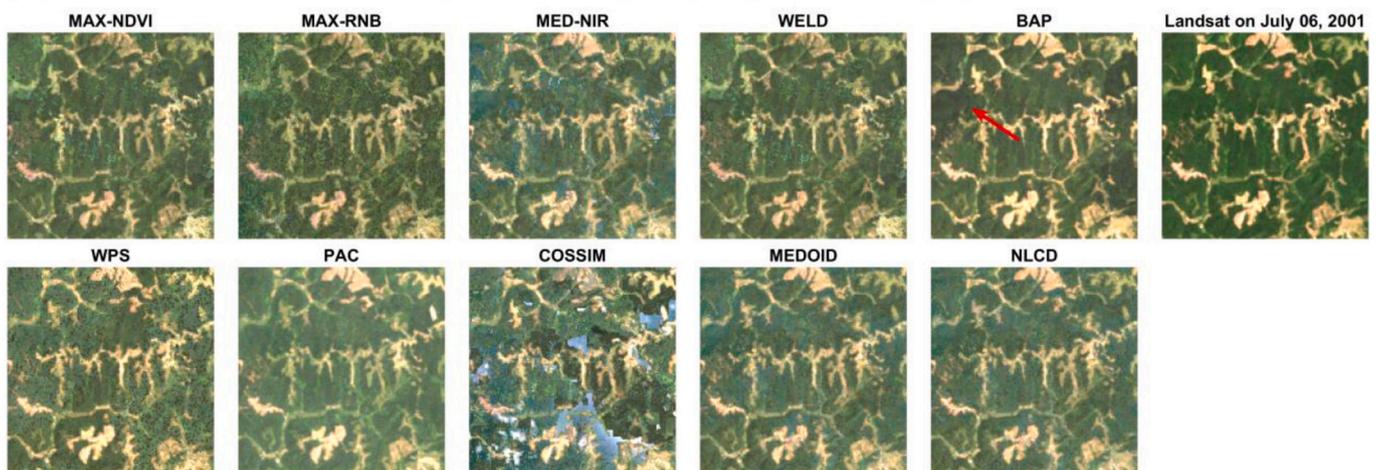
**(a) Monthly Landsat composite images (July 1  $\pm$  15 days)****(b) Seasonal Landsat composite images (July 1  $\pm$  45 days)****(c) Annual Landsat composite images (July 1  $\pm$  180 days)**

**Fig. 7.** Examples of Landsat natural-color composites generated by 10 different algorithms for a low cloudy region (P044R033). All the images are at the central location of lat 39.197669°, lon -122.004008°, with the same extent of 151 pixels by 151 pixels, and they are displayed by red, green, blue bands with the same stretches for display. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

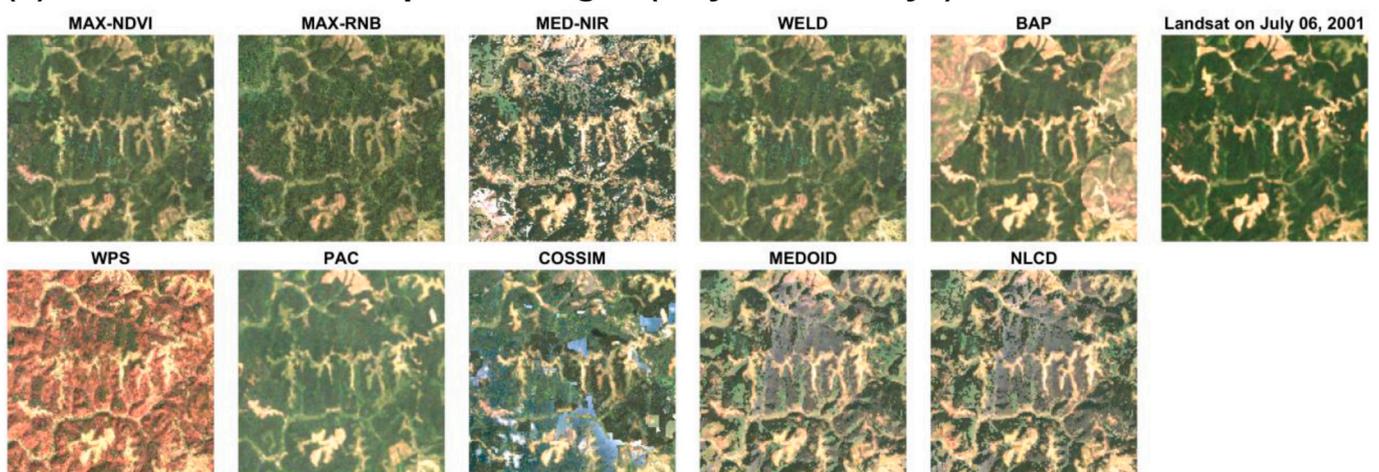
**(a) Monthly Landsat composite images (July 1 ± 15 days)**



**(b) Seasonal Landsat composite images (July 1 ± 45 days)**

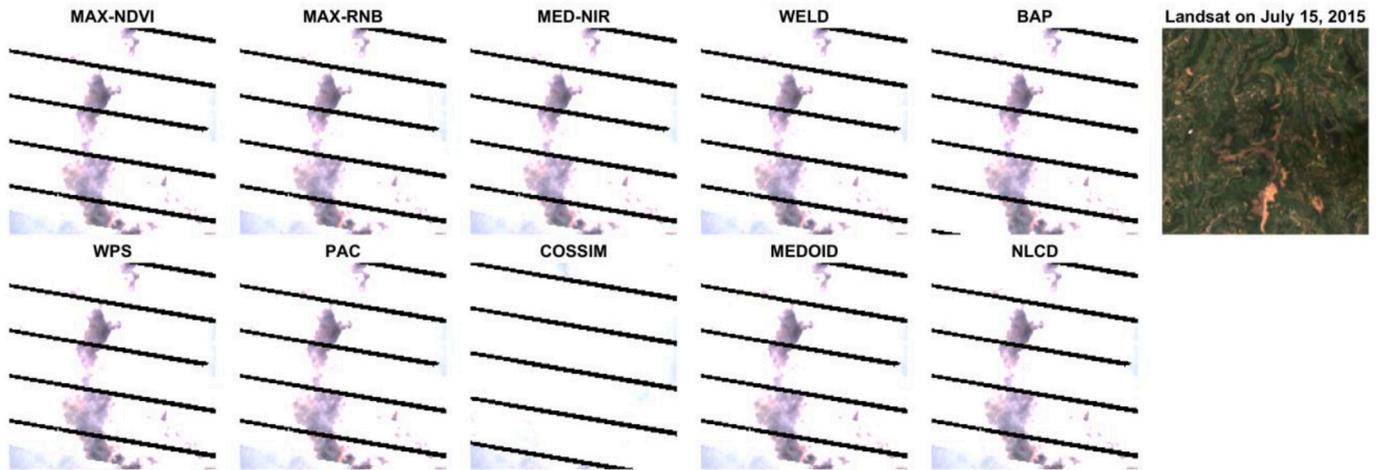


**(c) Annual Landsat composite images (July 1 ± 180 days)**

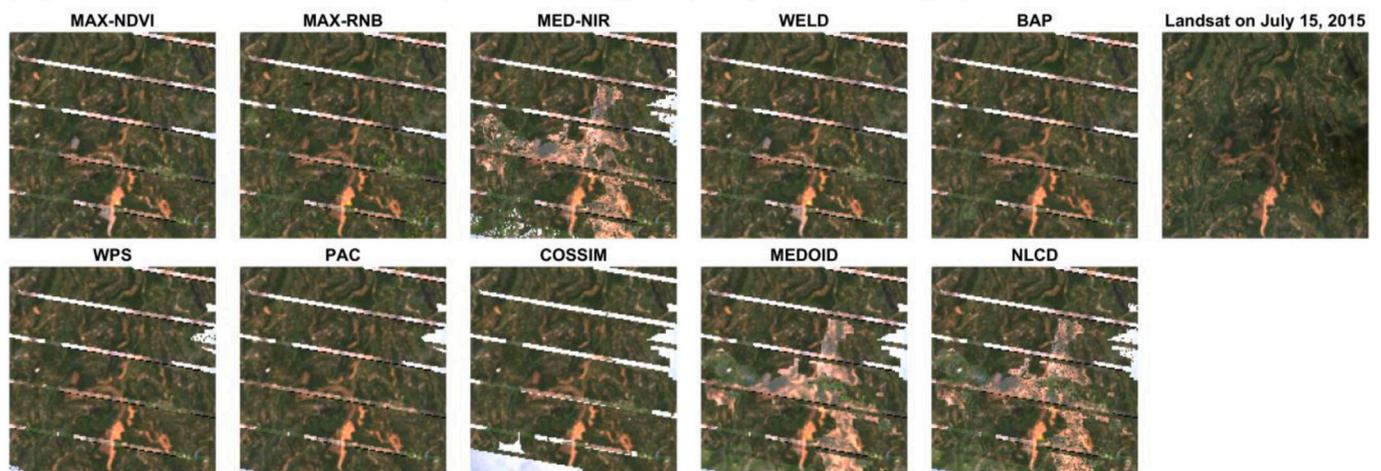


**Fig. 8.** Example of Landsat natural-color composites generated by 10 different algorithms at intermediate cloudy region (P018R033). All the images are at the central location of lat 38.787805°, lon -81.935457°, with the same extent of 151 pixels by 151 pixels, and they are displayed by red, green, blue bands with the same stretches for display. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

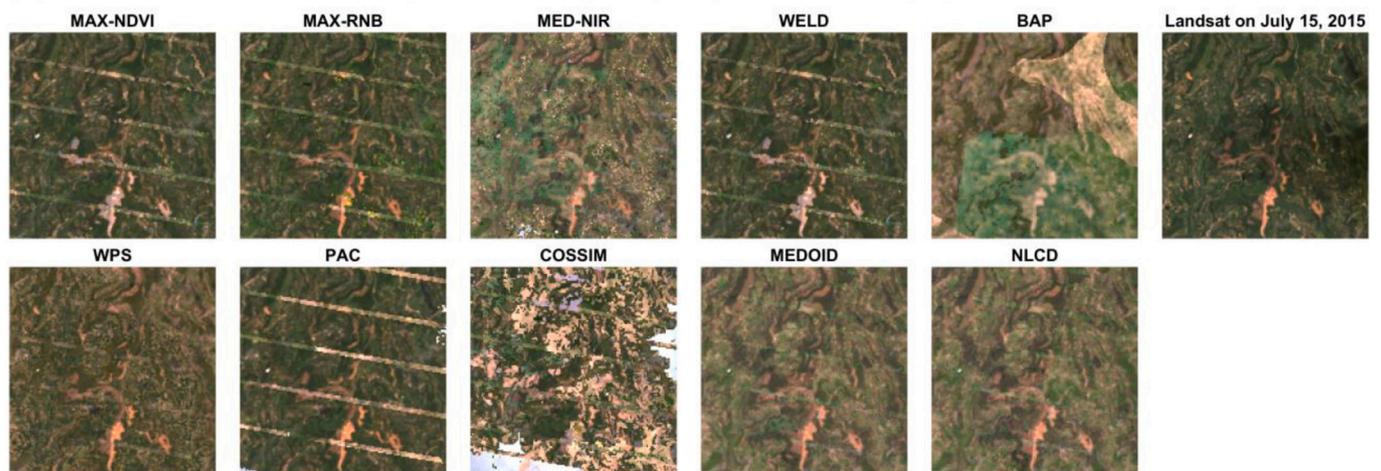
**(a) Monthly Landsat composite images (July 1 ± 15 days)**



**(b) Seasonal Landsat composite images (July 1 ± 45 days)**



**(c) Annual Landsat composite images (July 1 ± 180 days)**



**Fig. 9.** Example of Landsat natural-color composites generated by 10 different algorithms for the heavy cloud region (P129R038). All the images are at the central location of lat 31.874584°, lon 105.155999°, with the same extent of 151 pixels by 151 pixels, and they are displayed by red, green, blue bands with same stretch scales. Note that the stripe artifacts are caused by the Landsat 7 SLC-off issue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



Fig. 10. Quantitative evaluation (spectral domain) of 10 different Landsat image compositing algorithms (for monthly, seasonal, and annual composites) for all study areas based on mean R-square, RMSE, slope, and intercept. Abs. means absolute value. \*\* indicates the best case.

performance in maintaining spatial fidelity of the image. At each lag distance, the smaller the mean absolute difference value, the better the compositing results.

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N (\rho_i - \rho_{i+h})^2 \quad (17)$$

where,

$\rho_i$  is the surface reflectance of the  $i$ th pixel.

$\rho_{i+h}$  is the surface reflectance at the pixel with a lag distance of  $h$  from the  $i$ th pixel.

$h$  is the lag distance between pairs of pixels ( $\rho_i$  and  $\rho_{i+h}$ ).

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

### 4.3. Application evaluation

The Landsat composites generated from different compositing algorithms were used for land cover classification and change detection to evaluate their performance relative to similar results from the reference images. We trained a Random Forest (RF) classification model (Breiman, 2001) using land cover and land change training data (see Section 2.2 for details) to create land cover maps in 2000 and 2015, and land change maps between 2000 and 2015. The maps were created by supervised

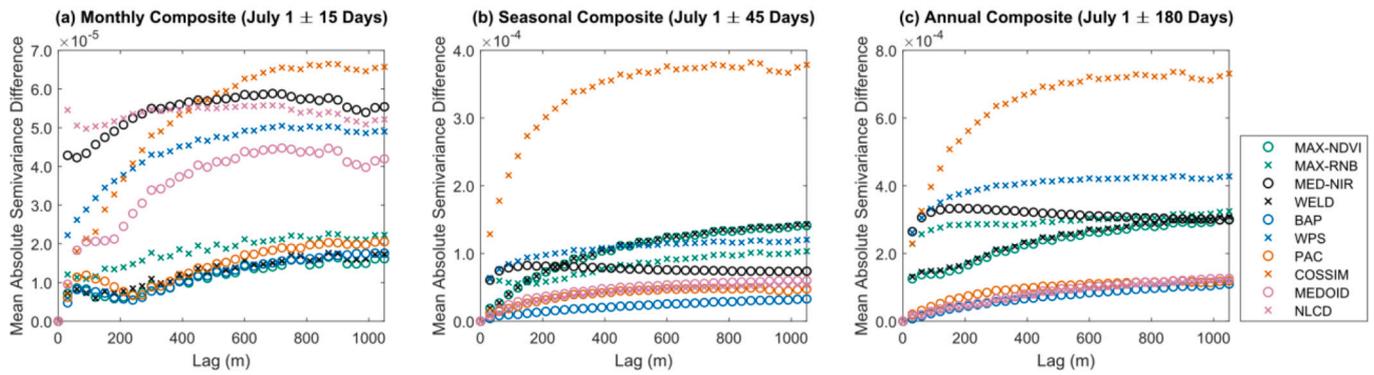
classification using 6 spectral bands from a single date and 12 bands from two dates, respectively. The widely used direct multidate classification approach was used for change detection (Singh, 1989). The number of trees was set to 100 for balancing computation efficiency and classification accuracy. The same land cover classification and land cover change training data were used for evaluating each compositing algorithm. The disagreement percentage between the maps derived from the composite images and that of the reference images were used as indicators of application performance; the smaller the disagreement, the better the compositing performance.

## 5. Results

### 5.1. Qualitative evaluation

Figs. 7-9 show three examples of Landsat composites (true color composites; false color composites are provided in Figs. S1-S3 in the Supplementary Materials) based on the 10 different algorithms in the three compositing periods, that are monthly (July ±15 days), seasonal (July ±45 days), and annual (July ±180 days). The locations in Figs. 7, 8, and 9 have low, intermediate, and high cloud coverage, respectively.

For the monthly Landsat composites, all 10 algorithms looked similar for the study site with low cloud coverage, but the MAX-NDVI, MAX-



**Fig. 11.** Quantitative evaluation (spatial domain) of 10 different Landsat image compositing algorithms (for monthly, seasonal, and annual composites) for all study areas.

RNB, WELD, BAP, PAC, COSSIM, and MEDOID algorithms generally presented better performances with less salt-and-pepper noise and artifacts (Fig. 7a). In the intermediate cloud coverage site (Fig. 8a), however, only the MAX-RNB and BAP algorithms exhibited the best results with very limited residual clouds. The MAX-NDVI and WELD algorithms showed the second-best results but were affected by clouds over unvegetated areas (red arrows in Fig. 8a). In the heavy cloud region, all Landsat image compositing algorithms exhibit similar results due to a very limited number of valid observations in a short period of time (Fig. 9a).

For the seasonal Landsat composites, the BAP algorithm showed the best results in the low cloudy area, and the MEDOID or NLCD algorithm (both are spectral-similarity-based) had similar and the second-best results with less salt-and-pepper noise and artifacts (Fig. 7b). In the intermediate cloudy area (Fig. 8b), the BAP-based image was very similar to the reference image, but selected observations acquired on different dates (the red arrow in Fig. 8b). In contrast, the PAC, MEDOID, and NLCD algorithms generated composites similar to the reference images without any obvious patches and salt-and-pepper noise. In the heavy cloud area (Fig. 9b), the BAP algorithm showed the best performance followed by the MAX-RNB, MAX-NDVI, and WELD algorithms.

Many more usable observations are available for the annual Landsat composites, which explains why the multiple-rules-based scoring algorithms (particularly for BAP and PAC) performed better. For example,

the BAP algorithm performed the best in the low cloudy region (Fig. 7c) but resulted in spatial artifacts in the intermediate cloudy (Fig. 8c) and heavy cloudy areas (Fig. 9c) because of the observations selected from different dates. It is worth noting that the BAP algorithm reduced the stripe artifacts caused by Landsat 7 SLC-off (Fig. 9c). The PAC algorithm presented robust results across all three sites (Fig. 7c, 8c, and 9c).

Judging from visual evaluations, no single Landsat compositing algorithm outperformed all others in all situations, instead, different algorithms worked better in different situations (e.g., MAX-RNB or BAP in monthly composites; BAP or NLCD in seasonal composites; and PAC or BAP in annual composites).

5.2. Quantitative evaluation

The algorithm performances were quantitatively assessed in the spectral, spatial, and application domains, respectively, as follows.

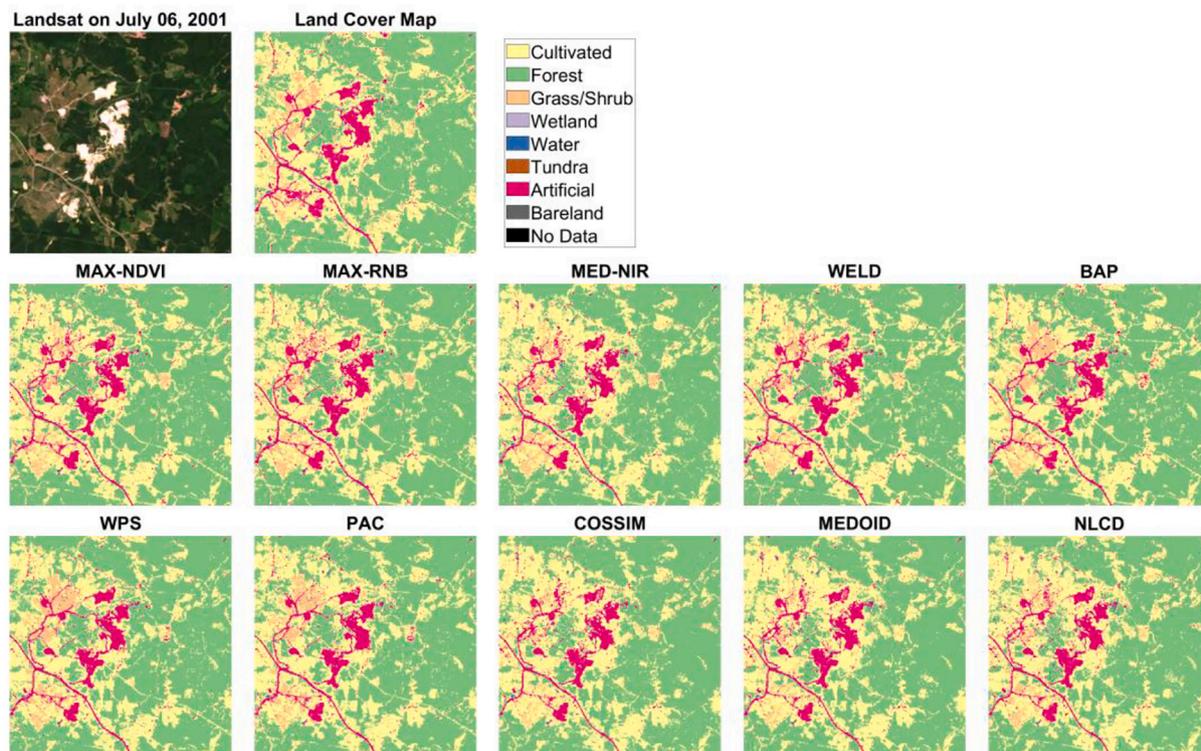
5.2.1. Spectral evaluation

Fig. 10 shows the quantitative evaluation in the spectral domain of the 10 compositing algorithms based on all test data during the monthly (July ±15 days), seasonal (July ±45 days), and annual (July ±180 days) compositing periods. The results demonstrate that the spectral indicators (e.g., R-square, RMSE, slope, and intercept) of the monthly composites were mostly similar to that of the seasonal composites for



**Fig. 12.** Quantitative evaluation (application domain) of 10 different Landsat image compositing algorithms (for monthly, seasonal, and annual composites) for all study areas measured by the land cover and land change maps differences compared to the reference image derived ones. \*\* indicates the best case.

### (a) Land cover maps from monthly Landsat composite images



### (b) Land change maps from monthly Landsat composite images

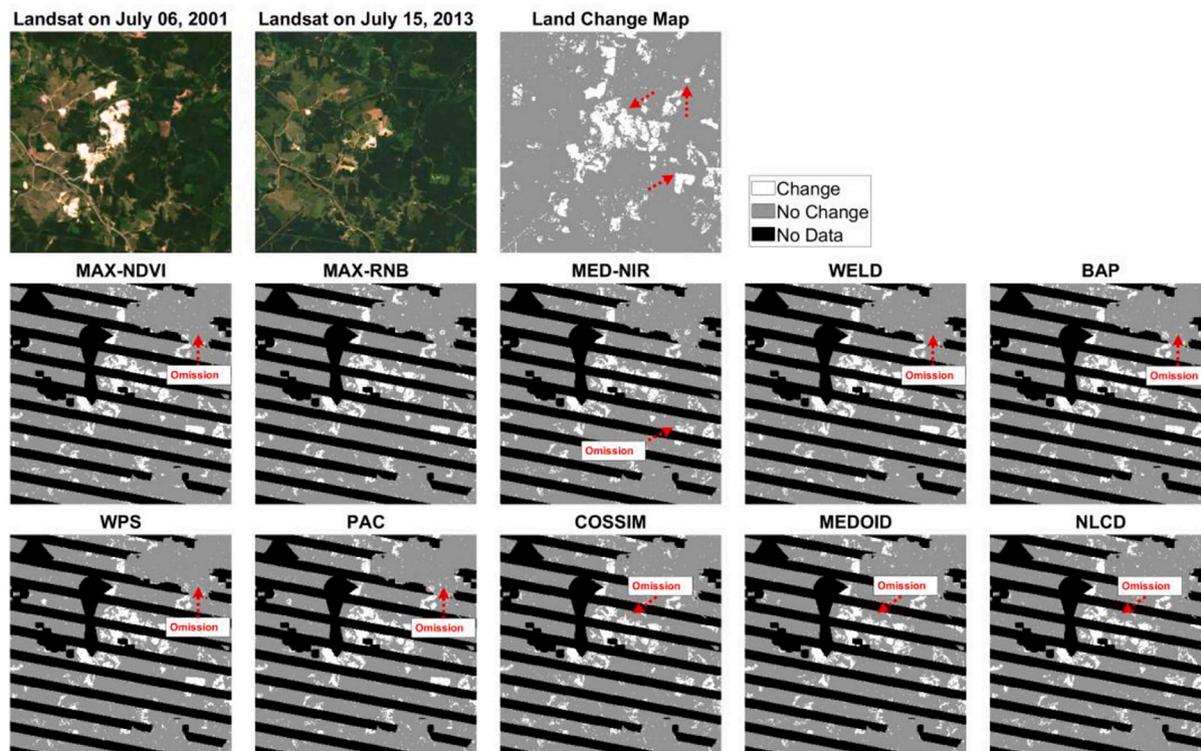
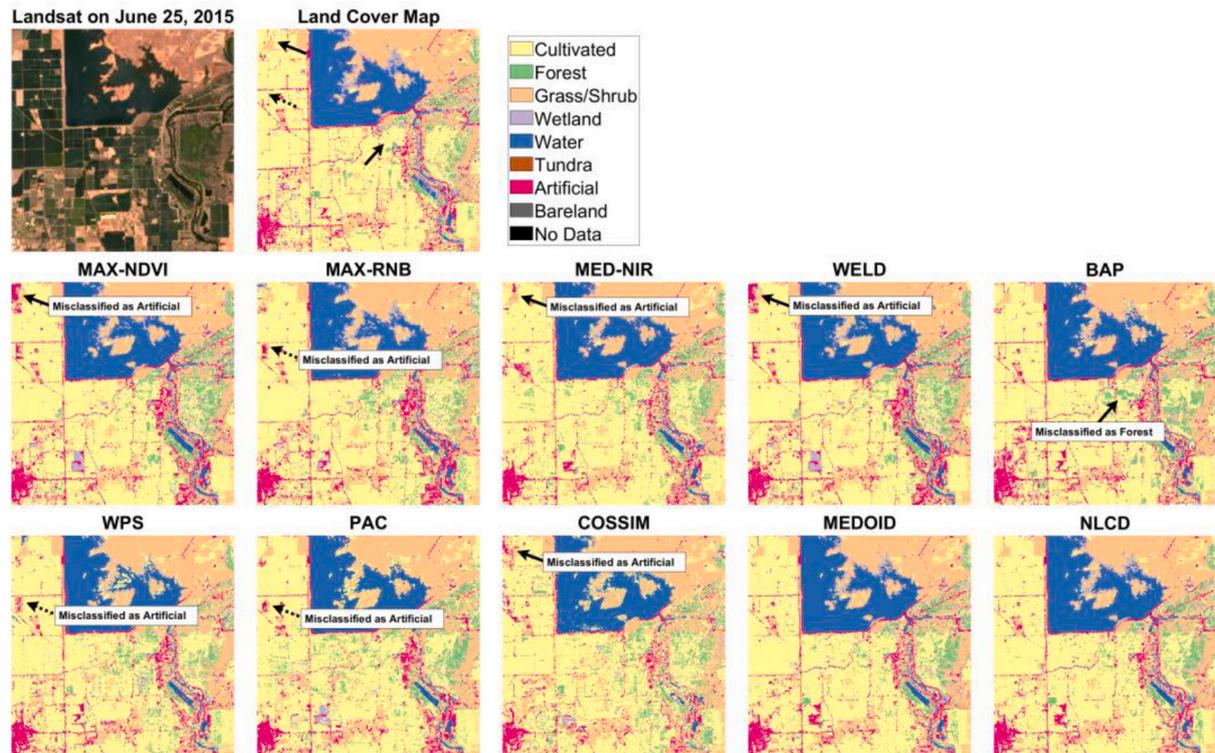


Fig. 13. An example of comparing land cover (a) and land change (b) maps derived from 10 monthly composite images (July 1 ± 15 days). The upper first plane indicates the Landsat reference data and the corresponding land cover or land change map. All the images or maps cover the same extent with 301 pixels by 301 pixels of Landsat WRS-2 P018R033 at the central location of lat 38.984897°, lon -82.481836°. Note that the “no data” stripes in the land change maps are caused by the Landsat 7 SLC-off issue.

### (a) Land cover maps from seasonal Landsat composite images



### (b) Land change maps from seasonal Landsat composite images

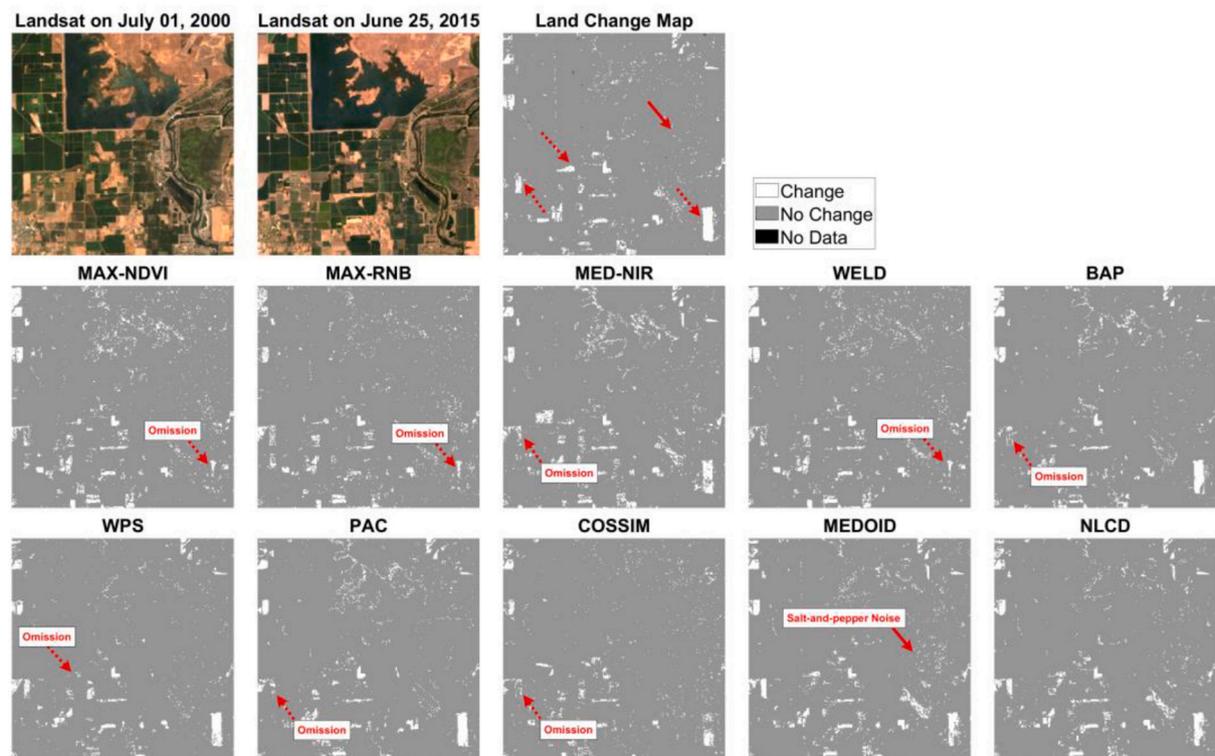
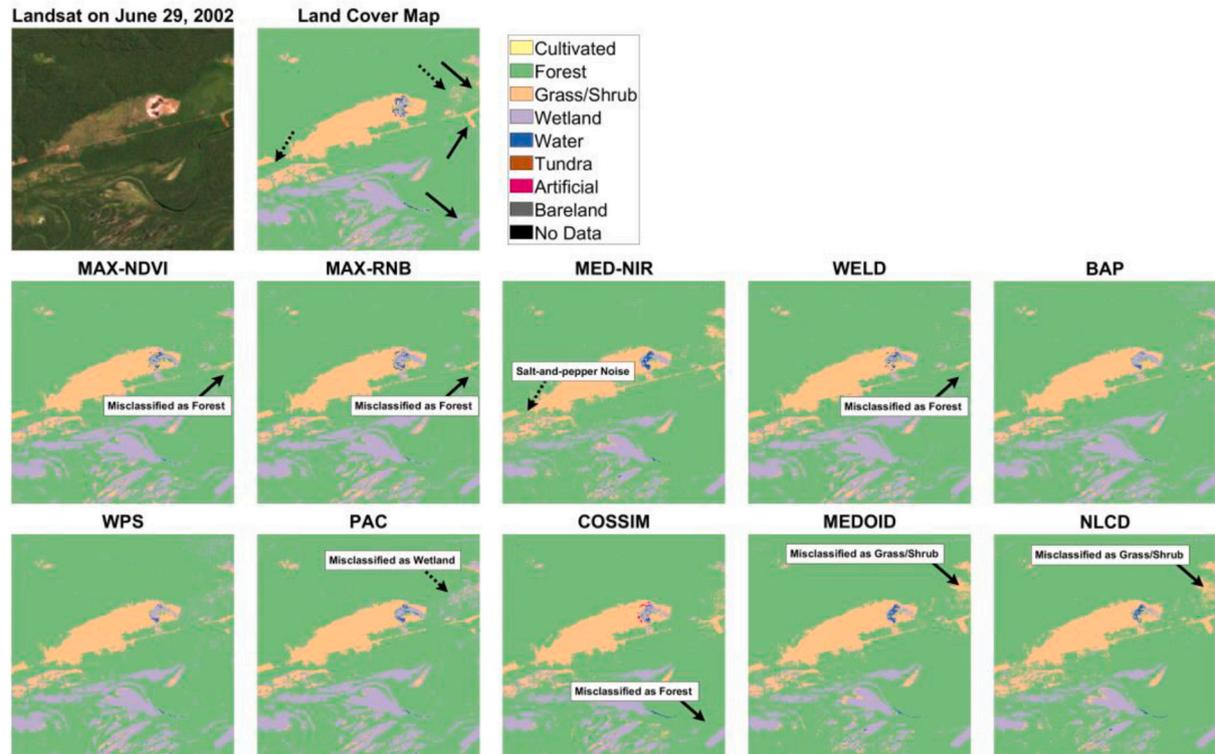


Fig. 14. An example of comparing land cover (a) and land change (b) maps derived from 10 seasonal composite images (July 1 ± 45 days). All the images or maps cover the same extent with 301 pixels by 301 pixels of Landsat WRS-2 P044R033 at central location of lat 39.446156°, lon -121.660556°. The upper first plane indicates the Landsat reference data and the corresponding land cover map.

### (a) Land cover maps from annual Landsat composite images



### (b) Land change maps from annual Landsat composite images

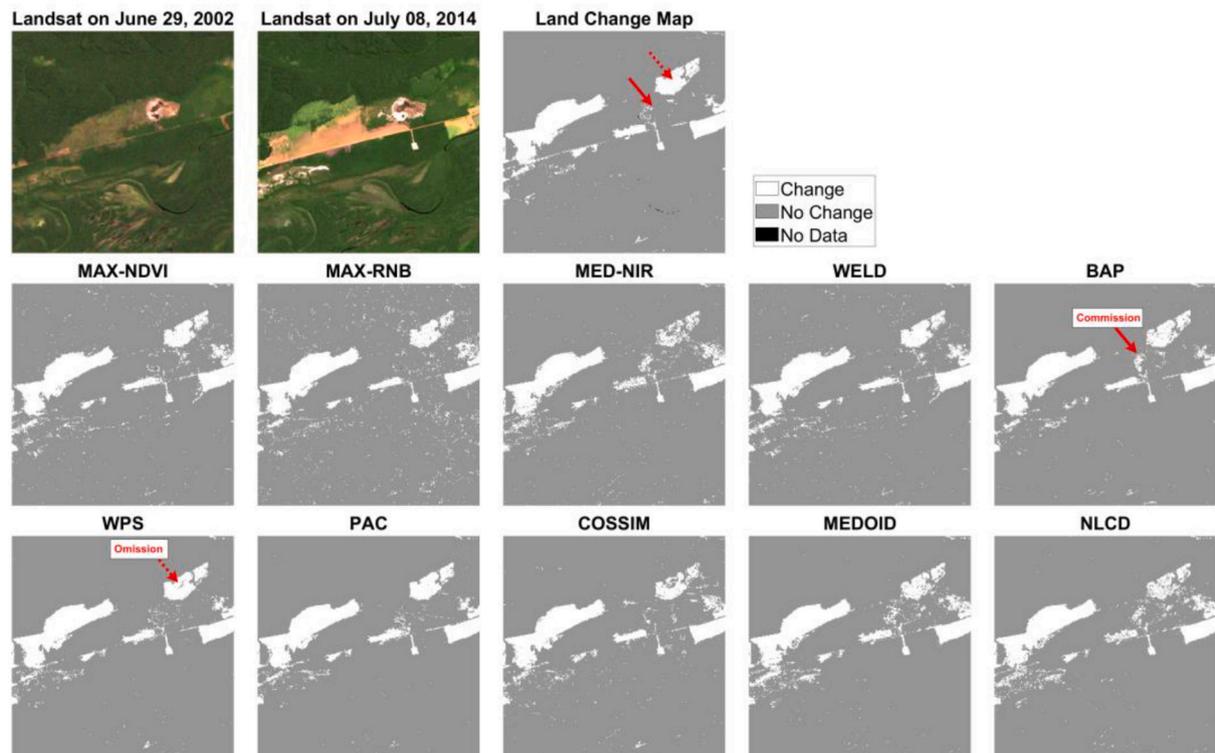


Fig. 15. An example of comparing land cover (a) and land change (b) maps derived from 10 seasonal composite images (July 1 ± 180 days). All the images or maps cover the same extent with 301 pixels by 301 pixels of Landsat WRS-2 P229R064 at the central location of lat -6.3493945°, lon -58.092597°. The upper first plane indicates the Landsat reference data and the corresponding land cover map.

most of the image compositing algorithms, but that the monthly composites produced slightly better results than the seasonal composites. Compared to monthly and seasonal composites, the annual composites from all algorithms had the lowest R-square, highest RMSE, worst slope, and largest intercept, which indicates that if more candidate observations are considered, it is more difficult to find observations that best represent the growing season reference image. In general, the narrower the image compositing window, the higher the spectral fidelity in image compositing.

The performance of each compositing algorithm varied with compositing period. For monthly composites, the MAX-RNB algorithm showed the best performance for spectral fidelity, with the highest R-square of 0.9594, the smallest RMSE of 0.0187, the best slope of 0.9686, and the lowest intercept of 0.0052. For seasonal composites, the NLCD algorithm had the highest R-square (0.9547) and the lowest RMSE (0.0190), while the PAC received the best slope (0.9931). The best intercept was obtained by the BAP algorithm (0.0036). For the annual composites, the PAC algorithm showed the best performance with the highest R-square of 0.9375, the lowest RMSE of 0.0228, the best slope of 0.9860, and the smallest intercept value of 0.0055. The BAP algorithm was the second best with a slightly lower performance compared to the PAC algorithm. The spectral evaluation of the algorithms at each individual Landsat path/row and each spectral band can be found in the Supplementary Materials (Table S1 and Figs. S4-S9).

5.2.2. Spatial evaluation

Fig. 11 shows the quantitative evaluation in the spatial domain of the algorithms based on all test data for the monthly, seasonal, and annual compositing periods. The mean value of absolute semivariance difference of all spectral bands between the composite and reference data increased through monthly, seasonal, and annual compositing periods (see the y-axis range in Fig. 11), which indicates that the narrower the compositing window the higher the spatial fidelity. During each individual compositing period, the BAP algorithm always showed the smallest difference for all lag distances. However, some algorithms were close to the performance of the BAP algorithm, such as the monthly composites of MAX-NDVI, WELD, PAC, and MAX-RNB (Fig. 11a) and seasonal/annual composites of NLCD and MEDOID (Fig. 11b/11c). The spatial evaluation of the algorithms at each individual Landsat path/row and each spectral band can be found in the Supplementary Materials (Figs. S10-S21).

5.2.3. Application evaluation

Fig. 12 shows the quantitative evaluation in the application domain of the algorithms based on the disagreement between land cover and land change maps constructed from composite and reference data.

For monthly Landsat composites, the MAX-RNB algorithm generally had the best results with less disagreement compared to the reference classifications, with about 12.95% disagreement in land cover and about 3.82% in land change. For example, Fig. 13 shows maps generated from different monthly compositing algorithms at a location with cultivated areas, artificial surfaces, and forests. The land cover maps constructed from the MAX-RNB composites generally showed less salt-and-pepper noise (Fig. 13a), and the MAX-RNB composites were able to map forest-harvest activities correctly as opposed to the other algorithms that exhibited more omission errors (see dashed red arrows in Fig. 13b).

For seasonal Landsat composites, the NLCD algorithm agreed well with the land cover and the land change derived from the reference Landsat imagery, with about 12.59% and about 3.43% disagreement, respectively. The MEDIOD algorithm achieved the second-best results in mapping land cover and land change. Fig. 14 shows maps made from seasonal composites, where the land cover is dominated by cultivated areas, grass/shrublands, water, artificial surface, and forests, and with land change driven by agriculture activities. Compared to the NLCD and MEDIOD algorithms, the other algorithms generated composite-based maps with obvious misclassifications of cultivated land as artificial

Table 7

Overall best Landsat compositing algorithms in the spectral, spatial, and application domains.

| Evaluation domains |                           | Composite periods |          |        |
|--------------------|---------------------------|-------------------|----------|--------|
|                    |                           | Monthly           | Seasonal | Annual |
| Spectral           |                           | MAX-RNB           | NLCD     | PAC    |
| Spatial            |                           | BAP               | BAP      | BAP    |
| Application        | Land cover classification | MAX-RNB           | NLCD     | BAP    |
|                    | Land change detection     | MAX-RNB           | NLCD     | PAC    |

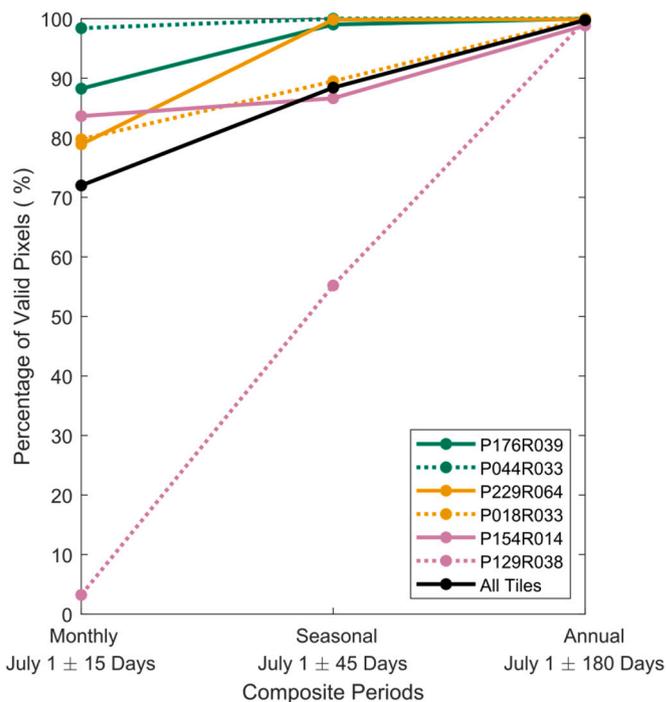


Fig. 16. Percentage of valid pixels in monthly, seasonal, and annual composites.

land or forest (black arrows in Fig. 14a). By carefully comparing the change maps constructed from different compositing algorithms, we observed that only the MEDOID and NLCD algorithms generated satisfactory results with less omission of change (see dashed red arrows in Fig. 14b). Compared to the NLCD, however, the MEDOID-based change map had more salt-and-pepper noise (see solid red arrow in Fig. 14b).

For annual composites, the multiple-scoring algorithms generally outperformed other types of algorithms in creating land cover and land change maps. For example, the BAP algorithm had the lowest disagreements in land cover maps (about 13.46%), while the PAC algorithm had the lowest disagreements in land change maps (about 3.98%). Fig. 15 shows land cover and land change maps made from annual composites of each algorithm, where there is a large area of forest, wetland, and grass/shrubland. The MAX-NDVI, MAX-RNB, WELD, COSSIM, MEDOID, and NLCD algorithms exhibit confusion between grass/shrubland (or wetland) and forest (see solid black arrows in Fig. 15a). Additionally, the MED-NIR algorithm resulted in the salt-and-pepper noise (e.g., forest pixels misidentified as grass/shrubland or wetland) (see dashed black arrows in Fig. 15a). The BAP, WPS, and PAC-based land cover maps look very similar. Still, the PAC-based land cover map had more misclassifications of forest as wetland (see dashed arrows in Fig. 15a), and a careful comparison indicates that the BAP-based land cover map with less salt-and-pepper noise was slightly better than the WPS (Fig. 15a). For the land change maps, the BAP, WPS, and PAC algorithms still showed the best performance. However, the BAP algorithm had more commission errors (solid red arrow in Fig. 15b). The one

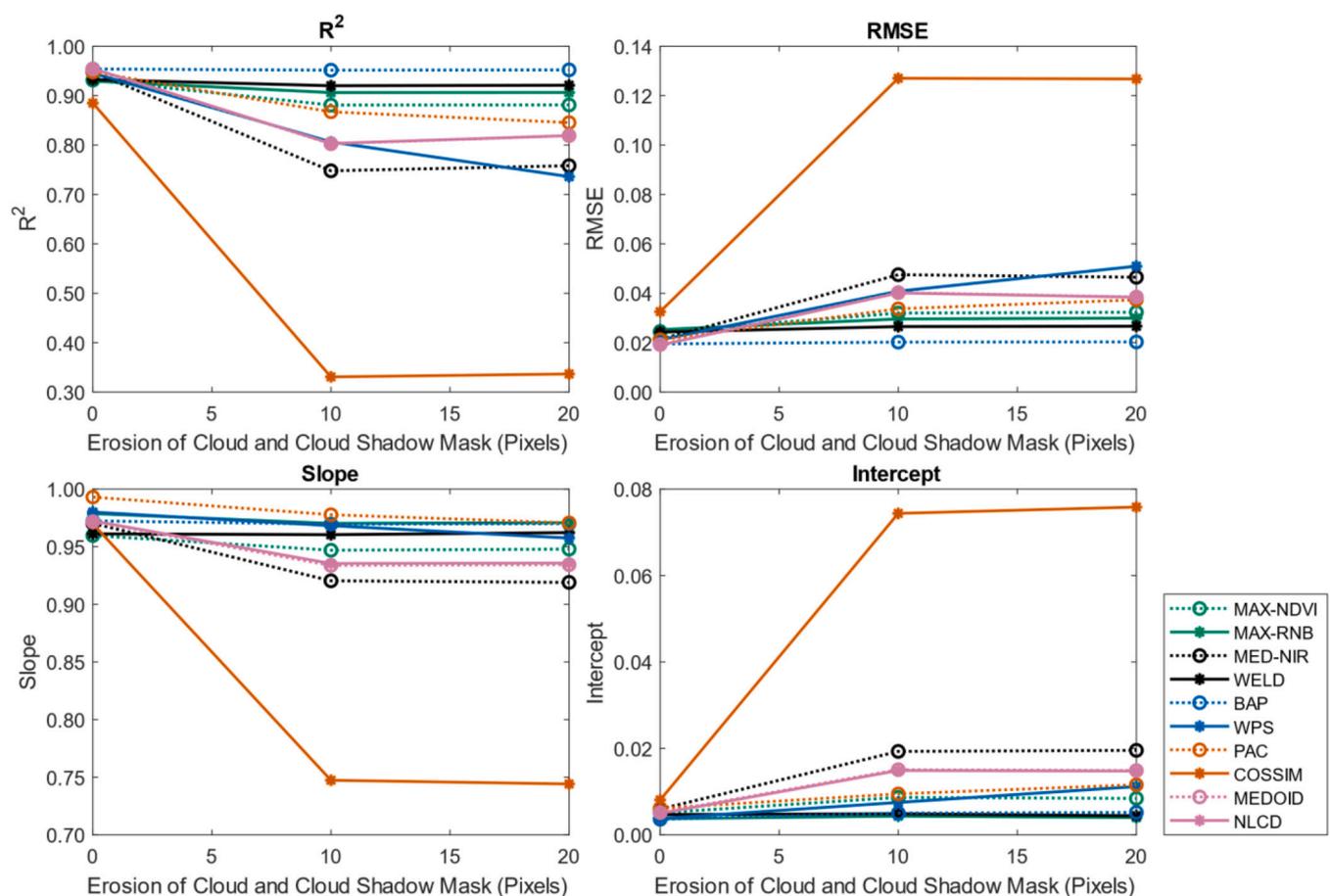


Fig. 17. Effects of cloud and cloud shadow mask on 10 different Landsat compositing algorithms. This analysis was made by seasonal compositing Landsat data based on all test sites.

derived from the PAC compositing algorithm had fewer omission errors of change, compared to the WPS algorithm (see dashed red arrow in Fig. 15b).

A summary of the compositing algorithms that generated the overall best results in each of the quantitative test domains is shown in Table 7.

## 6. Discussion

### 6.1. Compositing periods and data availability

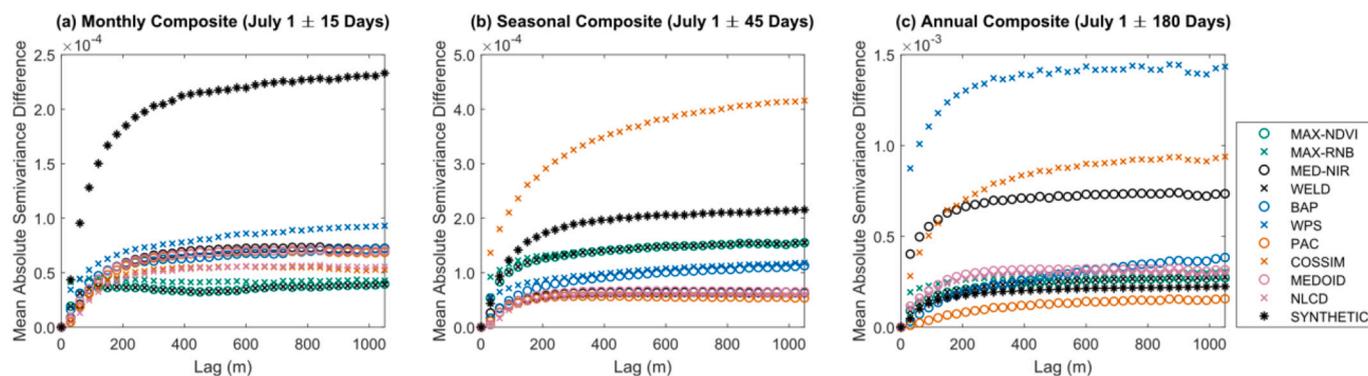
Our results showed that the MAX-RNB algorithm best maintained the spectral fidelity and had the best application performances for creating monthly Landsat composites, but a longer composite period such as seasonal and annual is often required due to clouds (Bleyhl et al., 2017; Griffiths et al., 2014; Hermosilla et al., 2016, 2022; Potapov et al., 2015; Suess et al., 2018). Across the test sites, monthly composites consisted of 71.99% of valid pixels, whereas seasonal and annual composites consisted of 88.42% and 99.75% of valid pixels, respectively (see black line in Fig. 16). This statistic depends to a large extent on the cloud cover of each test site (see colorful lines in Fig. 16). The spectral-similarity-based algorithm (i.e., NLCD) generally performed better than the other compositing algorithms in producing seasonal Landsat composites. On the other hand, a longer compositing period is more challenging for selecting the “best” observations, that are more easily affected by inconsistent reflectance or artifacts from vegetation phenology (Melaas et al., 2016), varying atmospheric condition (Thome, 2001), bi-direction reflectance effects associated with change in solar angles and satellite orbits (Qiu et al., 2021; Roy et al., 2020), Landsat 7 SLC-off (Wulder et al., 2008), and other unknown image quality issues. Thus, when

creating an annual Landsat composite, more comprehensive rules, such as the difference in acquisition date, the distance to cloud or cloud shadow, and penalization function for malfunctioning sensors, seem to improve results. Particularly, the BAP algorithm with penalization function for Landsat 7 SLC-off reduced the consequent artifacts (Hermosilla et al., 2022). The PAC algorithm with more scoring functions like phenology adaptation and spectral correlation (Frantz et al., 2017) maintained spectral fidelity and showed good results in producing land change maps for the annual composites. Additionally, narrowing the specified compositing period is a straight-forward way to reduce the effects of vegetation phenology, but a multi-year composite approach is often required to fill the “no data” pixels (White et al., 2014).

We created composites of six Landsat path/rows and found that algorithm performance varied greatly with location (see spectral evaluations of each path/row in Table S1, spatial evaluations of each path/row in Fig. S16-S21, and application evaluations of each path/row in Table S2). Thus, the choice of compositing algorithm needs to be tailored to data availability of the study area. For example, for places with high cloud coverage (Fig. 9), all compositing algorithms failed to create monthly image composites, but MAX-NDVI, MAX-RNB, WELD algorithms showed similar good performances in the seasonal and annual composites and are better choice for image compositing for places with extremely high cloud coverage.

### 6.2. Cloud and cloud shadow impacts

Cloud and cloud shadow detection is fundamental for Landsat image compositing (Zhu et al., 2018). Most image compositing algorithms were designed to reduce the effect of clouds and their shadows but do so



**Fig. 18.** Quantitative evaluation (spatial domain) of 10 different Landsat image compositing algorithms and the synthetic approach (for monthly, seasonal, and annual composites) at P018R033.

with mixed results. To explore the ability of the algorithms to handle cloud and cloud shadow omitted by Fmask, we included cloud and cloud shadow edge pixels on purpose by eroding the cloud and cloud shadow masks from 0 to 20 pixels (Fig. 17). We observed that the spectral indicators of the BAP, MAX-NDVI, MAX-RNB, WELD algorithms did not vary much for different erosion sizes which indicates that these algorithms are better when omitted clouds and cloud shadows are present. Furthermore, newer time-series-based cloud and cloud shadow detection algorithms, such as Cirrus cloud mask (Cmask) (Qiu et al., 2020) and multiTemporal mask (Tmask) (Zhu and Woodcock, 2014), may improve the quality of Landsat composites.

### 6.3. Composite images vs. synthetic data

A potential issue when selecting the “best” observation for inclusion in a composite is the influence of seasonality, especially if the compositing period is long. A solution to mitigate the effect of seasonality is to create synthetic Landsat imagery based on model regression (Zhu et al., 2015b) or multi-source data fusion (Gao et al., 2006). Such methods often require a large volume of data but have the advantage of having fewer or none “no data” pixels. Studies have found that synthetic Landsat data may significantly improve land cover classifications (Senf et al., 2015; Watts et al., 2011). Here, as a demonstration, we applied a model regression method (Zhu et al., 2015b) to create a synthetic Landsat imagery based on the COntinuous monitoring of Land Disturbance (COLD) algorithm (Zhu et al., 2020) for P018R033, where cloud coverage is intermediate. The COLD algorithm can create synthetic Landsat imagery at any given date based on a harmonic time series model (Zhu et al., 2015b). The Landsat time series models were estimated based on all the available Landsat surface reflectance images between 1985 and 2020, with cloud cover less than 100%. The spectral evaluation showed that the synthetic images, estimated on July 1 in 2001 and 2014, were highly correlated with the reference image:  $R$ -squared = 0.8070, which is better than any of the composite images (see P018R033 in Table S1). The RMSE value of the synthetic data (0.0186) was the lowest among all the annual composites, but higher than almost all monthly composites (except for COSSIM) and the MAX-RNB seasonal composites (see P018R033 in Table S1). However, the slope and intercept values of the synthetic data were 0.7908 and 0.0361, respectively, which were not as good as the monthly, seasonal, and annual composites (see P018R033 in Table S1). On the other hand, the performance of the synthetic imagery did not show good performance in maintaining the spatial fidelity of Landsat data, particularly for monthly and seasonal composites compared to the “traditional” compositing approaches examined (Fig. 18a and b). However, if a longer composite period is considered, most of the traditional compositing algorithms will become more unlike the reference data; hence, the synthetic data maintained the spatial fidelity better than most of the image compositing algorithms

evaluated (Fig. 18c). In addition, the land cover maps derived from the synthetic data had a disagreement of 13.02%, which is much lower than the various compositing algorithms for annual composites (see P018R033 in Table S2). Note that the synthetic data slightly underperformed in mapping land cover, compared to the best monthly (12.31% disagreement for MAX-NDVI) and seasonal (12.23% disagreement for BAP) composites. Meanwhile, the land change maps derived from the synthetic data, with a disagreement of 3.76%, are not as good as those from the best monthly, seasonal, and annual composites (3.16%, 3.13%, and 3.50% disagreement, respectively) in this specific Landsat path/row (see P018R033 in Table S2).

### 6.4. TOA reflectance vs. surface reflectance in image compositing

Although most Landsat compositing algorithms were developed based on surface reflectance, simple spectral-based algorithms, such as MAX-NDVI and WELD, were originally proposed based on inputs from TOA reflectance (Table 1). The MAX-RNB algorithm was designed to use surface reflectance, but the atmospheric correction process may contain errors, particularly for the blue band which is used in calculating the RNB values (Masek et al., 2006; Vermote et al., 2016). Therefore, we compared image compositing results for the three algorithms, including MAX-NDVI, WELD, and MAX-RNB, based on inputs from TOA versus surface reflectance. The comparison demonstrated that the use of TOA reflectance and the use of surface reflectance resulted in diverse but very slight differences in all the spectral, spatial, and application evaluations (Tables S3-S4 and Fig. S22). Compared to TOA reflectance, particularly, the use of surface reflectance in MAX-RNB yielded better results for monthly composites, but slightly worse performance for seasonal and annual composites. The MAX-NDVI and WELD algorithms using TOA reflectance generally showed better performance, but the improvements were very small. On the other hand, as Landsat atmospheric correction and surface reflectance retrieval algorithms are not ideal for water bodies due to the inherently low surface reflectance of water, the image compositing algorithms that use surface reflectance as their inputs (Table 1) may not work well over water areas.

### 6.5. Limitations

The MAX-RNB image compositing approach presented in this study is not without limitations. For example, MAX-RNB may work well for image compositing in land areas but may fail for water areas in which their RNB values are usually close to 1 and may not decrease in the presence of cloud. At the same time, the low blue band surface reflectance caused by overcorrection in atmospheric correction may also result in the increment of RNB values, particularly for cloud shadow pixels which are often very dark in both blue and NIR bands, and a subtle variations in the blue band could lead to large changes in the RNB

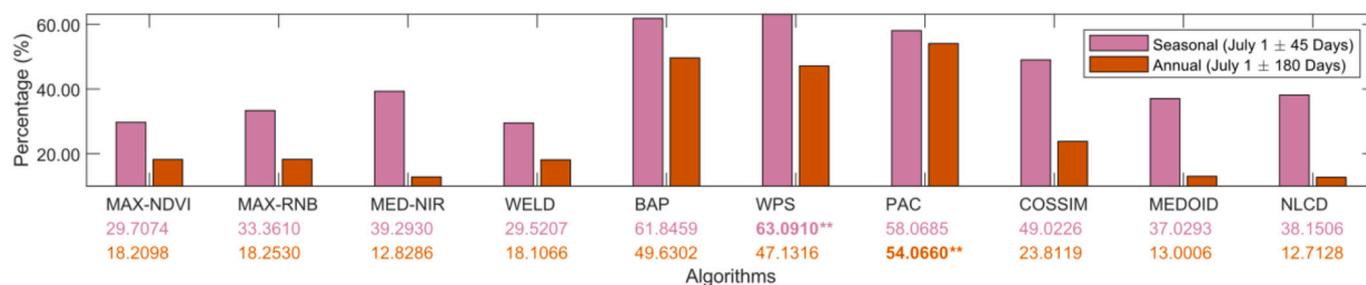


Fig. 19. Percentage of the observations selected from the period of July 1 ± 15 days for the seasonal and annual composites. \*\* indicates the maximum percentage. The BAP, WPS, and PAC algorithms use the target date (July 1) as one of the selection criteria.

values. This explains the patterns in Fig. 3, where a small proportion of the observations exhibit increased RNB when cloud or cloud shadow are present. Note that the limitation over water areas is also true for MAX-NDVI, as water observations usually have small (or negative) NDVI values, and sometimes can be negative, due to the high absorption in the NIR band in the water areas. Another limitation of the MAX-RNB algorithm is that by selecting the maximum value of the RNB values, most of the observations are selected during the peak growing season (highest NIR band and lowest blue band), which may not work well if the target compositing date is during leaf-off period.

Due to the limit of space, we only evaluated 10 Landsat image compositing algorithms that are either newly introduced here or have been widely used in various kinds of studies (Table 1). We do acknowledge that there are also many other image compositing methods could be compared and evaluated, such as Lück and Van Niekerk (2016), Luo et al. (2008), and Roberts et al. (2017).

#### 6.6. Target and reference compositing date

We used clear Landsat reference imagery, acquired on or close to a certain date (i.e., July 1) during the growing season, to compare different compositing algorithms. This approach assumes that the vegetation peak occurs around middle of the year; a common assumption in various kinds of remote sensing applications such as vegetation trend analyses and orbit change evaluations (Ju and Masek, 2016; Pflugmacher et al., 2014; Qiu et al., 2021; Tucker et al., 2004; Vicente-Serrano et al., 2008). This evaluation usually favors the algorithms using the target date as one of their selection criteria, such as BAP, WPS, and PAC, as they are more likely to select the clear observations close to the hidden reference data (Fig. 19).

On the other hand, in certain cases, when the target compositing date (July 1) is not during peak growing season, the conclusions made in this study could be different. Therefore, if peak vegetation image composites are needed, some simple methods, such as maximum NDVI algorithm may produce the best results. For example, to track cropland change (Xu et al., 2018) and map spatiotemporal fire trends (Kato et al., 2020), the maximum NDVI algorithm has proven useful. Meanwhile, leaf-off composites can be beneficial for improving land cover classification and land change detection (Higginbottom et al., 2018; Jin et al., 2023), and methods that select less vegetated pixels may show better performance. For land change mapping, the target compositing date could vary widely based on the specific application (White et al., 2014); in wildfire detection, the target compositing date is usually at the end of the fire season, while for insect infestation detection, the target compositing date usually depends on the specific types of damage required for detection (e.g., defoliator versus bark beetle).

## 7. Conclusions

We proposed a Landsat compositing algorithm based on the maximum RNB values, and qualitatively and quantitatively evaluated a total of 10 different Landsat image compositing algorithms. We

observed that each algorithm is unique, and no algorithm can outperform all other algorithms in all locations and all composite periods. For monthly composites, the MAX-RNB algorithm generated the best results in maintaining the spectral and application fidelity. For seasonal composites, the NLCD algorithm presented the best results for the spectral and application fidelity. For annual composites, the PAC algorithm presented the best results in keeping the spectral fidelity and generating the land change maps, but the BAP algorithm showed best results for mapping land cover. The BAP algorithm also generated data with the best spatial fidelity for all the compositing periods. This information can guide users in selecting appropriate compositing algorithms based on data availability, cloud cover, and compositing periods.

#### CRedit authorship contribution statement

**Shi Qiu:** Conceptualization, Methodology, Formal analysis, Visualization, Data curation, Writing - original draft, Writing - review & editing. **Zhe Zhu:** Conceptualization, Methodology, Formal analysis, Visualization, Funding acquisition, Writing - review & editing. **Pontus Olofsson:** Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Curtis E. Woodcock:** Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Suming Jin:** Conceptualization, Methodology, Formal analysis, 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.

#### Data availability

Data will be made available on request.

#### Acknowledgments

This study was supported by a USGS-NASA Landsat Science Team (LST) contract #140G0119C0008 to the University of Connecticut Storrs (Toward Near Real-time Monitoring and Characterization of Land Surface Change for the Conterminous US). Pontus Olofsson was supported by the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) program, grant number 80NSSC18K0994. We would like to thank the editors and the three anonymous reviewers for their insightful comments and suggestions for this manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2022.113375>.

## References

- Bleyhl, B., Baumann, M., Griffiths, P., Heidelberg, A., Manvelyan, K., Radeloff, V.C., Zazanashvili, N., Kuemmerle, T., 2017. Assessing landscape connectivity for large mammals in the Caucasus using landsat 8 seasonal image composites. *Remote Sens. Environ.* 193, 193–203. <https://doi.org/10.1016/j.rse.2017.03.001>.
- Broich, M., Hansen, M.C., Potapov, P., Adusei, B., Lindquist, E., Stehman, S.V., 2011. Time-series analysis of multi-resolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. *Int. J. Appl. Earth Obs. Geoinf.* 13, 277–291. <https://doi.org/10.1016/j.jag.2010.11.004>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- 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., 2020. Lessons learned implementing an operational continuous United States national land change monitoring capability: the land change monitoring, assessment, and project (LCMAP) approach. *Remote Sens. Environ.* 238, 111356 <https://doi.org/10.1016/j.rse.2019.111356>.
- Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* 113, 893–903. <https://doi.org/10.1016/j.rse.2009.01.007>.
- Chen, Jun, Chen, Jin, Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., Mills, J., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* 103, 7–27. <https://doi.org/10.1016/j.isprsjprs.2014.09.002>.
- Chuvieco, E., Ventura, G., Martín, M.P., Gómez, I., 2005. Assessment of multitemporal compositing techniques of MODIS and AVHRR images for burned land mapping. *Remote Sens. Environ.* 94, 450–462. <https://doi.org/10.1016/j.rse.2004.11.006>.
- Cihlar, J., Manak, D., 1994. Evaluation of compositing algorithms for AVHRR data over land. *IEEE Trans. Geosci. Remote Sens.* 32, 427–437. <https://doi.org/10.1109/36.295057>.
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161. <https://doi.org/10.1016/j.rse.2018.09.002>.
- Curran, P.J., 1988. The semivariogram in remote sensing: an introduction. *Remote Sens. Environ.* 24, 493–507.
- Flood, N., 2013. Seasonal composite Landsat TM/ETM+ images using the medoid (a multi-dimensional median). *Remote Sens.* 5, 6481–6500. <https://doi.org/10.3390/rs5126481>.
- Frantz, D., Röder, A., Stellmes, M., Hill, J., 2017. Phenology-adaptive pixel-based compositing using optical earth observation imagery. *Remote Sens. Environ.* 190, 331–347. <https://doi.org/10.1016/j.rse.2017.01.002>.
- Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the MODIS and landsat ETM+ surface reflectance. *IEEE Trans. Geosci. Remote Sens.* 44, 2207–2218. <https://doi.org/10.1109/TGRS.2006.872081>.
- Gong, P., Wang, J., Yu, L., Zhao, Yongchao, Zhao, Yuanxuan, Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Yao, W., Zhang, Han, Zhu, P., Zhao, Z., Zhang, Haiying, Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, Liang, Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, Jin, Chen, Jun, Guo, J., Yu, Liang, Wang, L., Liu, X., Shi, T., Zhu, M., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* 34, 2267–2654. <https://doi.org/10.1080/01431161.2012.748992>.
- Griffiths, P., Kuemmerle, T., Baumann, M., Radeloff, V.C., Abrudan, I.V., Lieskovsky, J., Munteanu, C., Ostapowicz, K., Hostert, P., 2014. Forest disturbances, forest recovery, and changes in forest types across the carpathian ecoregion from 1985 to 2010 based on Landsat image composites. *Remote Sens. Environ.* 151, 72–88. <https://doi.org/10.1016/j.rse.2013.04.022>.
- 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. <https://doi.org/10.1016/j.rse.2018.10.031>.
- Griffiths, P., van der Linden, S., Kuemmerle, T., Hostert, P., 2013. A pixel-based Landsat compositing algorithm for large area land cover mapping. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* 6, 2088–2101. <https://doi.org/10.1109/JSTARS.2012.2228167>.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342, 850–853. <https://doi.org/10.1126/science.1244693>.
- Hansen, M.C., Roy, D.P., Lindquist, E., Adusei, B., Justice, C.O., Altstatt, A., 2008. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sens. Environ.* 112, 2495–2513. <https://doi.org/10.1016/j.rse.2007.11.012>.
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., 2022. Land cover classification in an era of big and open data: optimizing localized implementation and training data selection to improve mapping outcomes. *Remote Sens. Environ.* 268, 112780 <https://doi.org/10.1016/j.rse.2021.112780>.
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Campbell, L.B., 2016. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. *Int. J. Digital Earth* 9, 1035–1054. <https://doi.org/10.1080/17538947.2016.1187673>.
- Higginbottom, T.P., Symeonakis, E., Meyer, H., van der Linden, S., 2018. Mapping fractional woody cover in semi-arid savannas using multi-seasonal composites from Landsat data. *ISPRS J. Photogramm. Remote Sens.* 139, 88–102. <https://doi.org/10.1016/j.isprsjprs.2018.02.010>.
- Holben, B.N., 1986. Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* 7, 1417–1434. <https://doi.org/10.1080/01431168608948945>.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2).
- Jin, S., Dewitz, J., Danielson, P., Granneman, B., Costello, C., Zhu, Z., 2023. National Land Cover Database 2019: a new strategy for creating clean Landsat composite images. *J. Remote Sens.* In press.
- Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., Zhu, Z., Xian, G., Howard, D., 2019. Overall methodology design for the United States National Land Cover Database 2016 products. *Remote Sens.* 11 <https://doi.org/10.3390/rs11242971>.
- Ju, J., Masek, J.G., 2016. The vegetation greenness trend in Canada and US Alaska from 1984–2012 Landsat data. *Remote Sens. Environ.* 176, 1–16. <https://doi.org/10.1016/j.rse.2016.01.001>.
- Kato, A., Thau, D., Hudak, A.T., Meigs, G.W., Moskal, L.M., 2020. Quantifying fire trends in boreal forests with Landsat time series and self-organized criticality. *Remote Sens. Environ.* 237, 111525 <https://doi.org/10.1016/j.rse.2019.111525>.
- Liu, X., Huang, Y., Xu, X., Li, Xuecao, Li, Xia, Ciais, P., Lin, P., Gong, K., Ziegler, A.D., Chen, A., Gong, P., Chen, J., Hu, G., Chen, Y., Wang, S., Wu, Q., Huang, K., Estes, L., Zeng, Z., 2020. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nat. Sustain.* 3, 564–570. <https://doi.org/10.1038/s41893-020-0521-x>.
- Luo, Y., Trishchenko, A.P., Khlopenkov, K.V., 2008. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sens. Environ.* 112, 4167–4185. <https://doi.org/10.1016/j.rse.2008.06.010>.
- Lück, W., Van Niekerk, A., 2016. Evaluation of a rule-based compositing technique for Landsat-5 TM and Landsat-7 ETM+ images. *Int. J. Appl. Earth Obs. Geoinf.* 47, 1–14. <https://doi.org/10.1016/j.jag.2015.11.019>.
- 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. <https://doi.org/10.1109/LGRS.2005.857030>.
- Melaas, E.K., Sulla-Menashe, D., Gray, J.M., Black, T.A., Morin, T.H., Richardson, A.D., Friedl, M.A., 2016. Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat. *Remote Sens. Environ.* 186, 452–464. <https://doi.org/10.1016/j.rse.2016.09.014>.
- Nelson, K.J., Steinwand, D., 2015. A Landsat data tiling and compositing approach optimized for change detection in the conterminous United States. *Photogramm. Eng. Remote Sens.* 81, 573–586. <https://doi.org/10.14358/PERS.81.7.573>.
- Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540, 418–422. <https://doi.org/10.1038/nature20584>.
- Pflugmacher, D., Cohen, W.B., Kennedy, R.E., Yang, Z., 2014. Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sens. Environ.* 151, 124–137. <https://doi.org/10.1016/j.rse.2013.05.033>.
- Potapov, P., Turubanova, S., Hansen, M.C., 2011. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sens. Environ.* 115, 548–561. <https://doi.org/10.1016/j.rse.2010.10.001>.
- Potapov, P.V., Turubanova, S.A., Tyukavina, A., Krylov, A.M., McCarty, J.L., Radeloff, V.C., Hansen, M.C., 2015. Eastern Europe's forest cover dynamics from 1985 to 2012 quantified from the full Landsat archive. *Remote Sens. Environ.* 159, 28–43. <https://doi.org/10.1016/j.rse.2014.11.027>.
- Qian, G., Sural, S., Gu, Y., Pramanik, S., 2004. Similarity between Euclidean and cosine angle distance for nearest neighbor query. In: *Proceedings of the 2004 ACM Symposium on Applied Computing*, pp. 1232–1237.
- Qiu, S., Zhu, Z., He, B., 2019. Fmask 4.0: improved cloud and cloud shadow detection in Landsat 4–8 and Sentinel-2 imagery. *Remote Sens. Environ.* 231, 111205 <https://doi.org/10.1016/j.rse.2019.05.024>.
- Qiu, S., Zhu, Z., Shang, R., Crawford, C.J., 2021. Can landsat 7 preserve its science capability with a drifting orbit? *Sci. Remote Sens.* 4, 100026 <https://doi.org/10.1016/j.srs.2021.100026>.
- Qiu, S., Zhu, Z., Woodcock, C.E., 2020. Cirrus clouds that adversely affect Landsat 8 images: what are they and how to detect them? *Remote Sens. Environ.* 246, 111884 <https://doi.org/10.1016/j.rse.2020.111884>.
- Roberts, D., Mueller, N., McIntyre, A., 2017. High-dimensional pixel composites from earth observation time series. *IEEE Trans. Geosci. Remote Sens.* 55, 6254–6264. <https://doi.org/10.1109/TGRS.2017.2723896>.
- Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalsky, V., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010. Web-enabled Landsat data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sens. Environ.* 114, 35–49. <https://doi.org/10.1016/j.rse.2009.08.011>.
- Roy, D.P., Li, Z., Zhang, H.K., Huang, H., 2020. A conterminous United States analysis of the impact of Landsat 5 orbit drift on the temporal consistency of Landsat 5 thematic mapper data. *Remote Sens. Environ.* 240, 111701 <https://doi.org/10.1016/j.rse.2020.111701>.
- Ruefenacht, B., 2016. Comparison of three Landsat TM compositing methods: a case study using modeled tree canopy cover. *Photogramm. Eng. Remote Sens.* 82, 199–211. <https://doi.org/10.14358/PERS.82.3.199>.
- Ryan, K.C., Opperman, T.S., 2013. LANDFIRE—A national vegetation/fuels data base for use in fuels treatment, restoration, and suppression planning. *For. Ecol. Manag.* 294, 208–216. <https://doi.org/10.1016/j.foreco.2012.11.003>.
- Senf, C., Leitão, P.J., Pflugmacher, D., van der Linden, S., Hostert, P., 2015. Mapping land cover in complex Mediterranean landscapes using Landsat: improved

- classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sens. Environ.* 156, 527–536. <https://doi.org/10.1016/j.rse.2014.10.018>.
- Singh, A., 1989. Review article: digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 10, 989–1003. <https://doi.org/10.1080/01431168908903939>.
- Small, C.G., 1990. A survey of multidimensional medians. *Int.Stat.Rev.Revue Internationale de Statistique* 263–277. <https://doi.org/10.2307/1403809>.
- Struyf, A., Hubert, M., Rousseeuw, P., 1997. Clustering in an object-oriented environment. *J. Stat. Softw.* 1, 1–30. <https://doi.org/10.18637/jss.v001.i04>.
- Suess, S., van der Linden, S., Okujeni, A., Griffiths, P., Leitão, P.J., Schwieder, M., Hostert, P., 2018. Characterizing 32 years of shrub cover dynamics in southern Portugal using annual Landsat composites and machine learning regression modeling. *Remote Sens. Environ.* 219, 353–364. <https://doi.org/10.1016/j.rse.2018.10.004>.
- Thome, K.J., 2001. Absolute radiometric calibration of Landsat 7 ETM+ using the reflectance-based method. *Remote Sens. Environ.* 78, 27–38. [https://doi.org/10.1016/S0034-4257\(01\)00247-4](https://doi.org/10.1016/S0034-4257(01)00247-4).
- Tucker, C.J., Compton, J., Grant, D.M., Dykstra, J.D., 2004. NASA's global orthorectified Landsat data set. *Am. Soc. Photogramm.Remote Sens.* 70, 313–322. <https://doi.org/10.14358/PERS.70.3.313>.
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* 185, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>.
- Vermote, E.F., Saleous, N., 2007. LEDAPS surface reflectance product description. University of Maryland Department of Geography, College Park.
- Vermote, E.F., Vermeulen, A., 1999. Atmospheric correction algorithm: spectral reflectances (MOD09). Algorithm Technical Background Document.
- Vicente-Serrano, S.M., Pérez-Cabello, F., Lasanta, T., 2008. Assessment of radiometric correction techniques in analyzing vegetation variability and change using time series of Landsat images. *Remote Sens. Environ.* 112, 3916–3934. <https://doi.org/10.1016/j.rse.2008.06.011>.
- 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. <https://doi.org/10.1016/j.rse.2010.08.005>.
- White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A., Guindon, L., 2014. Pixel-based image compositing for large-area dense time series applications and science. *Can. J. Remote. Sens.* 40, 192–212. <https://doi.org/10.1080/07038992.2014.945827>.
- Wolfe, R.E., Roy, D.P., Vermote, E., 1998. MODIS land data storage, gridding, and compositing methodology: level 2 grid. *IEEE Trans. Geosci. Remote Sens.* 36, 1324–1338. <https://doi.org/10.1109/36.701082>.
- Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., 2008. Free access to Landsat imagery. *Science* 320, 1011. <https://doi.org/10.1126/science.320.5879.1011a> (1979).
- Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* 122, 2–10. <https://doi.org/10.1016/j.rse.2022.113195>.
- Wulder, M.A., Ortellepp, S.M., White, J.C., Maxwell, S., 2008. Evaluation of Landsat-7 SLC-off image products for forest change detection. *Can. J. Remote. Sens.* 34, 93–99. <https://doi.org/10.5589/m08-020>.
- Wulder, M.A., Roy, D.P., Radeloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M., Healey, S., Zhu, Z., Scambos, T.A., Pahlevan, N., 2022. Fifty years of Landsat science and impacts. *Remote Sens. Environ.* 280, 113195.
- Xu, Y., Yu, L., Zhao, F.R., Cai, X., Zhao, J., Lu, H., Gong, P., 2018. Tracking annual cropland changes from 1984 to 2016 using time-series Landsat images with a change-detection and post-classification approach: experiments from three sites in Africa. *Remote Sens. Environ.* 218, 13–31. <https://doi.org/10.1016/j.rse.2018.09.008>.
- Zhang, Y., Guindon, B., Cihlar, J., 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. *Remote Sens. Environ.* 82, 173–187. [https://doi.org/10.1016/S0034-4257\(02\)00034-2](https://doi.org/10.1016/S0034-4257(02)00034-2).
- Zhang, Y., Woodcock, C.E., Arévalo, P., Olofsson, P., Tang, X., Stanimirova, R., Bullock, E., Tarrío, K.R., Zhu, Z., Friedl, M.A., 2022. A global analysis of the spatial and temporal variability of usable Landsat observations at the pixel scale. *Front. Remote Sens.* 55 <https://doi.org/10.3389/frsen.2022.894618>.
- Zhu, Z., Qiu, S., He, B., Deng, C., 2018. Cloud and cloud shadow detection for Landsat images: the fundamental basis for analyzing Landsat time series. *Remote Sens.Time Ser.Image Process.* 3–24 <https://doi.org/10.1201/9781315166636-10>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015a. Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and sentinel 2 images. *Remote Sens. Environ.* 159, 269–277. <https://doi.org/10.1016/j.rse.2014.12.014>.
- Zhu, Z., Woodcock, C.E., 2014. Automated cloud, cloud shadow, and snow detection in multitemporal landsat data: an algorithm designed specifically for monitoring land cover change. *Remote Sens. Environ.* 152, 217–234. <https://doi.org/10.1016/j.rse.2014.06.012>.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>.
- Zhu, Z., Woodcock, C.E., Holden, C., Yang, Z., 2015b. Generating synthetic Landsat images based on all available Landsat data: predicting Landsat surface reflectance at any given time. *Remote Sens. Environ.* 162, 67–83. <https://doi.org/10.1016/j.rse.2015.02.009>.
- Zhu, Z., Wulder, M.A., Roy, D.P., Woodcock, C.E., Hansen, M.C., Radeloff, V.C., Healey, S.P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.F., Lymburner, L., Pahlevan, N., Scambos, T.A., 2019. Benefits of the free and open Landsat data policy. *Remote Sens. Environ.* 224, 382–385. <https://doi.org/10.1016/j.rse.2019.02.016>.
- Zhu, Z., Zhang, J., Yang, Z., Aljaddani, A.H., Cohen, W.B., Qiu, S., Zhou, C., 2020. Continuous monitoring of land disturbance based on Landsat time series. *Remote Sens. Environ.* 238, 111116 <https://doi.org/10.1016/j.rse.2019.03.009>.