



Cloud detection algorithm comparison and validation for operational Landsat data products



Steve Foga^{a,*}, Pat L. Scaramuzza^{a,1}, Song Guo^b, Zhe Zhu^{c,d}, Ronald D. Dilley Jr.^b, Tim Beckmann^a, Gail L. Schmidt^a, John L. Dwyer^e, M. Joseph Hughes^f, Brady Laue^a

^a Stinger Gaffarian Technologies, contractor to the U.S. Geological Survey Earth Resources Observation and Science Center, 47914 252nd Street, Sioux Falls, SD 57198-0001, USA

^b Innovate Inc., contractor to the U.S. Geological Survey Earth Resources Observation and Science Center, 47914 252nd Street, Sioux Falls, SD 57198-0001, USA

^c ASRC Federal Inuteq, contractor to the U.S. Geological Survey Earth Resources Observation and Science Center, 47914 252nd Street, Sioux Falls, SD 57198-0001, USA

^d Department of Geosciences, MS 1053, Science Building 125, Texas Tech University, Lubbock, TX 79409, USA

^e U.S. Geological Survey Earth Resources Observation and Science Center, 47914 252nd Street, Sioux Falls, SD 57198-0001, USA

^f College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 104 CEOAS Administration Building, Corvallis, OR 97331-5503, USA

ARTICLE INFO

Article history:

Received 28 June 2016

Received in revised form 8 March 2017

Accepted 25 March 2017

Available online 12 April 2017

Keywords:

Landsat

CFMask

Cloud detection

Cloud validation masks

Biome sampling

Data products

ABSTRACT

Clouds are a pervasive and unavoidable issue in satellite-borne optical imagery. Accurate, well-documented, and automated cloud detection algorithms are necessary to effectively leverage large collections of remotely sensed data. The Landsat project is uniquely suited for comparative validation of cloud assessment algorithms because the modular architecture of the Landsat ground system allows for quick evaluation of new code, and because Landsat has the most comprehensive manual truth masks of any current satellite data archive. Currently, the Landsat Level-1 Product Generation System (LPGS) uses separate algorithms for determining clouds, cirrus clouds, and snow and/or ice probability on a per-pixel basis. With more bands onboard the Landsat 8 Operational Land Imager (OLI)/Thermal Infrared Sensor (TIRS) satellite, and a greater number of cloud masking algorithms, the U.S. Geological Survey (USGS) is replacing the current cloud masking workflow with a more robust algorithm that is capable of working across multiple Landsat sensors with minimal modification. Because of the inherent error from stray light and intermittent data availability of TIRS, these algorithms need to operate both with and without thermal data. In this study, we created a workflow to evaluate cloud and cloud shadow masking algorithms using cloud validation masks manually derived from both Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 8 OLI/TIRS data. We created a new validation dataset consisting of 96 Landsat 8 scenes, representing different biomes and proportions of cloud cover. We evaluated algorithm performance by overall accuracy, omission error, and commission error for both cloud and cloud shadow. We found that CFMask, C code based on the Function of Mask (Fmask) algorithm, and its confidence bands have the best overall accuracy among the many algorithms tested using our validation data. The Artificial Thermal-Automated Cloud Cover Algorithm (AT-ACCA) is the most accurate nonthermal-based algorithm. We give preference to CFMask for operational cloud and cloud shadow detection, as it is derived from a priori knowledge of physical phenomena and is operable without geographic restriction, making it useful for current and future land imaging missions without having to be retrained in a machine-learning environment.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Cloud cover is a pervasive, dynamic, and often unavoidable feature in space-borne remotely sensed optical imagery. Estimates show the mean annual global cloud cover over land surfaces is approximately 35% in the mid-latitudes (Ju and Roy, 2008) and between 58%

(Rossow and Schiffer, 1999) and 66% (Zhang et al., 2004) over land surfaces worldwide. The effects of clouds in remotely sensed data are well documented in research pertaining to Landsat, an ongoing satellite remote sensing program operational since 1972 and designed to examine Earth's terrestrial environment (Goward et al., 2001; Irons et al., 2012). Multiple approaches have been used in past studies to identify Landsat pixels contaminated by clouds and/or cloud shadows. Some approaches relied upon spectral tests and decision trees incorporating many or all of Landsat's bands (e.g., Helmer et al., 2009; Hollingsworth et al., 1996; Irish, 2000; Irish et al., 2006; Jin et al., 2013; Reeves et al., 2006; Roy et al., 2010; Scaramuzza et al., 2012; Vermote and Saleous, 2007;

* Corresponding author.

E-mail address: steven.foga.ctr@usgs.gov (S. Foga).

¹ Both authors contributed equally to this work.

Vermote et al., 2016; Zhu and Woodcock, 2012). Temporally adjacent Landsat acquisitions can provide contrasting land cover information to help identify contaminated pixels that statistically deviate in certain wavelengths (Hagolle et al., 2010; Jin et al., 2013; Wang et al., 1999; Yang et al., 2003), among or between tasseled-cap components (Hughes and Hayes, 2014; Kennedy et al., 2007; Kennedy et al., 2010), or from expected ground target conditions, such as water (Sheng et al., 2016) or forest (Helmer et al., 2009; Helmer et al., 2012; Huang et al., 2010). Statistical deviations cannot only create new masks, but can be used to find omitted cloud pixels from an incumbent cloud mask (Frantz et al., 2015; Selkowitz and Forster, 2015; Zhu and Woodcock, 2014a, 2014b). Clouds have also been successfully identified through the harmonization of remote sensing datasets, such as imagery from the Moderate Resolution Spectroradiometer (MODIS) and Landsat (Oreopoulos et al., 2011; Roy et al., 2008). Artificial Neural Networks (ANN) have been trained to identify cloud or cloud-like objects by their texture (Lee et al., 1990) or multispectral reference data for clouds (Hughes and Hayes, 2014; Scaramuzza et al., 2012) and cloud shadows (Hughes and Hayes, 2014), though the resulting nodes of either technique do not necessarily have bearing to the physical qualities of clouds.

With so many different cloud detection algorithms available in the literature, it is extremely important to compare the performances of all those approaches and provide guidance on which algorithm to use for the remote sensing community interested in using Landsat data. The U.S. Geological Survey (USGS), which manages the ground stations and data processing operations for the Landsat mission (Irons et al., 2012), also needs to select the algorithm that works best across Landsat sensors and across different parts of the world.

Therefore, in this study we assessed the accuracy of multiple cloud masking algorithms to determine the best globally applicable algorithm to be used in future Landsat quality assurance data products. To accomplish this, we used 278 unique cloud validation masks to compare the accuracy and error of 13 cloud masking algorithms and 5 cloud shadow masking algorithms over the entire globe. We tested incumbent algorithms against newer experimental algorithms that have never before been validated. To utilize some of these algorithms, we first translated them from their scientific implementation to an open-source programming language that could be operated and maintained in a production environment. For validation masks, we used already existing cloud masks developed for Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 8 Operational Land Imager (OLI) instruments. We then developed our own global, semi-random scene selection methodology, selected 96 Landsat 8 OLI/Thermal Infrared Sensor (TIRS) scenes, stratified by both biome and cloud cover, and derived cloud and cloud shadow masks from each scene to use in this validation study.

2. Existing cloud validation masks

To perform algorithm validation, we derived our own unique cloud validation masks from Landsat 8 OLI/TIRS imagery and used two already existing masking datasets (Table 1; Fig. 1). The “L7 Irish” validation scenes were originally selected by Irish et al. (2006), digitized by

Scaramuzza et al. (2012), and consisted of 207 masks. In this study, we used only the 102 “validation” masks, where the remaining masks are designated for algorithm training. These masks consist of cloud and thin cloud classes. All 207 of the “L7 Irish” masks are currently provided online at no cost (USGS, 2016a).

The “L8 SPARCS” dataset was developed to validate the Landsat 8 extension of Hughes and Hayes’ Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS) algorithm (Hughes and Hayes, 2014). Each of the 80 sub-scenes is a 3-km by 3-km subset (1000×1000 30-m samples) of a Landsat 8 OLI image. Landsat 8 OLI images derived from Worldwide Reference System-2 (WRS-2) path/rows were selected randomly for each of the 14 World Wildlife Fund terrestrial Major Habitat Types (MHT) that occur in each of the seven Biogeographical Realms (Olson et al., 2001). Since not every MHT occurs in each Realm, this process provided 64 locations. An additional 16 locations were then selected randomly from all terrestrial scenes. Individual images were selected from all available Landsat 8 OLI imagery acquired through the end of 2015 to provide a mix of cloud forms and vegetation phenology over the dataset. Pixels in each sub-scene were then hand-labeled (i.e., without using spectral thresholding or pre-classification) as cloud, cloud shadow, snow/ice, water, flooded, or clear-sky by a single interpreter. For the validation used here, the snow/ice, water, and flooded classes were relabeled as clear-sky. Six sub-scenes were labeled by an independent interpreter with 96% agreement, implying that at least 4% of pixels in the dataset are ambiguous. The “L8 SPARCS” dataset is currently provided online at no cost (USGS, 2016b).

3. New validation mask creation

3.1. Initial sample selection methods

Our sampling methodology was largely based upon the Landsat 7 ETM+ cloud cover assessment (CCA) methodology presented by Irish et al. (2006), which was originally developed to perform training and validation for the Automated CCA (ACCA) algorithm. While our study did not require a training dataset, as all of our tested algorithms already existed, we purposely digitized an even number of cloud masks so others could use half the dataset for training, and the other half for validation. In our dataset, clouds were not digitized for validation of any one particular CCA algorithm; instead, they were digitized with the categories of fill, clear, thin cloud, cloud, and cloud shadow. We digitized cloud shadows that were clearly identifiable throughout a scene; oftentimes, however, we passed up shadow digitization on scenes that had low illumination or high relief areas that subsequently cast many shadows.

To select our target scenes, we devised a unique method to randomly generate a list of globally representative Landsat scenes. Our idealized scenario was to have a random selection process (i.e., scenes picked by a random number generator) to minimize human-induced bias. This was not achievable due to several unavoidable flaws in the selection process. We initially tried a random sampling based upon latitudinal zones, used by Irish et al. (2006), which resulted in heavy clustering. To resolve this, we weighted each latitudinal zone by percent landmass. This method

Table 1
Cloud validation masks and their respective characteristics.

Name	# Scenes	Cell size	Grid dims.	Sensor	Files	Cld.	Cld. sh.	Thin cld.	Snow-Ice	Water	Fld.
L7 Irish (I)	102	30 m	Whole scene	ETM+ (SLC on)	Level-1 product, metadata, validation mask	Yes	Some	Yes	no	no	no
L8 SPARCS (S)	80	30 m	1000×1000 pixels	OLI	Level-1 product, QA band, validation mask, preview image, metadata	Yes	Yes	No	Yes	Yes	Yes
L8 Biome (B)	96	30 m	Whole scene	OLI	Level-1 product, metadata, validation mask	Yes	Some	Yes	No	No	No

Cld. is cloud; Cld. sh. is cloud shadow; Fld. is flooded.

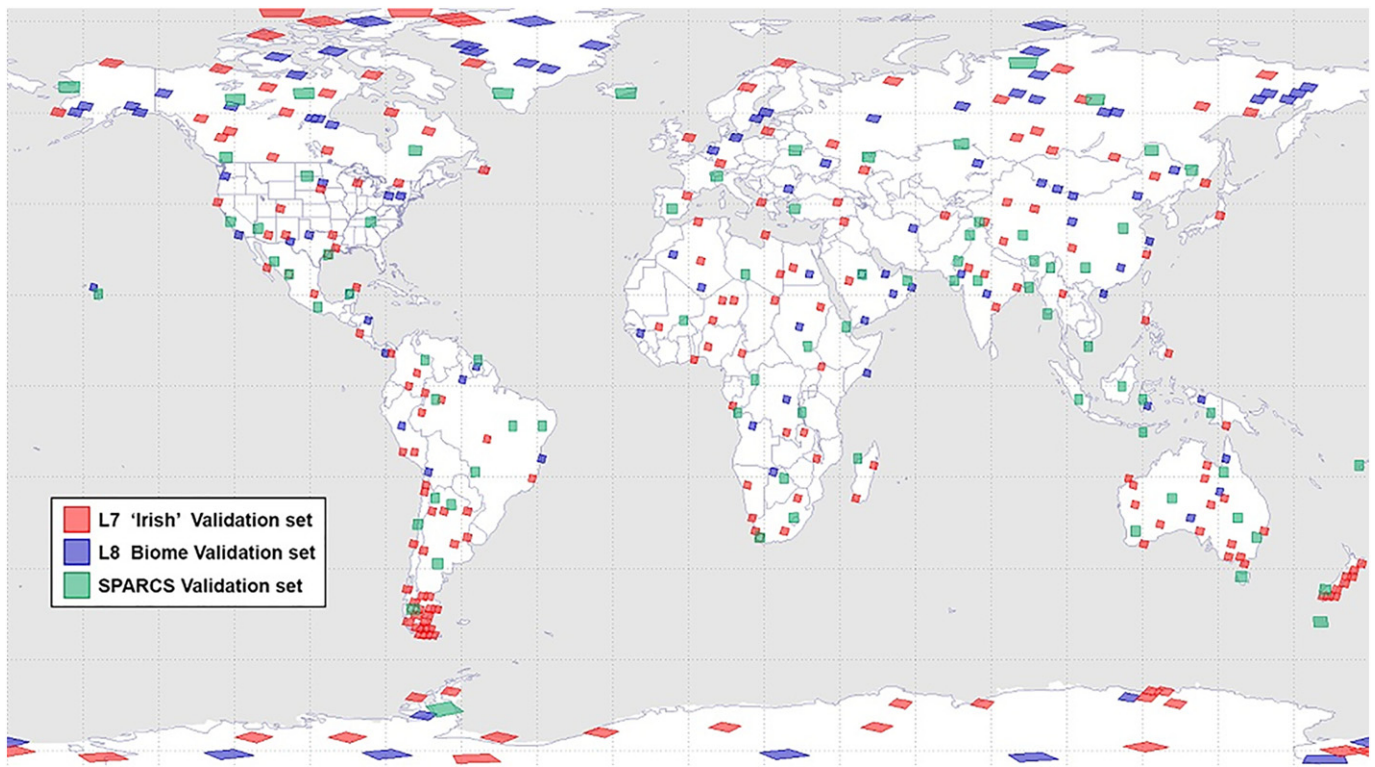


Fig. 1. Global distribution of Landsat 8 Cloud Cover Assessment (CCA) ("L8 Biome") scenes, Landsat 8 Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS) subset tiles (shown here as the entire scene's footprint), and Landsat 7 CCA ("L7 Irish") scenes. A total of 278 scene footprints are shown.

reduced clustering, but ultimately was unusable due to the consistent overestimation of the Landsat Level-1 CCA score. Thus, we revised the methodology altogether, which is detailed below.

3.2. Semi-random sampling by biome

We eliminated the use of latitudinal zones and replaced them with biome-based stratification. By selecting scenes by biome, we reduced the heterogeneity of land cover types within a given latitude zone, particularly in the mid-latitudes, and focused upon performance over certain land cover types, which may be of more interest to the greater scientific community than latitude-based criteria. We used biomes from the International Geosphere Biosphere Programme (IGBP; Loveland et al. (2000)). We selected this dataset based upon its unique inclusion of an Urban biome. The other classes within this dataset are Barren, Forest, Shrubland, Grass/Cropland, Snow/Ice, Wetlands, and Water.

We classified each land-based WRS-2 path/row by its percent composition within each biome. The path/row was then classified by plurality participation into a single biome; then 12 scenes were randomly selected from each biome-filtered path/row list. Random path/row sampling still induced some undesirable clustering in several regions. We decided to sacrifice some of the random selection process in favor of creating a spatially heterogeneous dataset. We discarded approximately 25% of the original path/rows and randomly re-selected new ones, which reduced visibly apparent clustering.

From the final set of path/rows, we manually selected individual Landsat scenes based upon the approximate amount of clouds visible within each scene, while criteria such as seasonality were ignored to maintain simplicity. The scene selections were performed manually by an analyst due to the inaccurate Level-1 CCA score (also noted

by Jones, 2015). We used the following approximate cloud cover bins:

- Clear (<35%)
- Mid-cloud ($\geq 35\%$ and $\leq 65\%$)
- Cloudy (>65%)

The cloud cover bins were determined around the global cloud cover average over land quoted between 35% (Ju and Roy, 2008) and 66% (Zhang et al., 2004). Some data consumers may find scenes > 65% unusable for certain large-scale applications; however, we contend that being able to reductively find any noncloud-contaminated pixels may still be useful in smaller-scale applications, or could be later composited with other uncontaminated acquisitions.

We selected four scenes for each cloud cover bin and biome pair, bringing the total number of validation scenes to 96 (Fig. 2).

3.3. Digitization process

To create the masks, we digitized the Landsat Level-1 multispectral data into thematic rasters. The following digitization steps were performed by a single analyst, as differences due to analyst interpretations in previous cloud masking studies resulted in an average error of about 7% (Scaramuzza et al., 2012). Scenes were digitized using Adobe® Photoshop® software, using a Wacom digitizer display to expedite the classification of cloud pixels. Band compositing was used to better contrast clouds from ground features. Individual bands were kept at their native 16-bit resolution and composited to 48-bit (3-band) images. The typical band combinations used for digitizing clouds are described in Table 2. Band 1 (coastal aerosol) was never used.

Clouds that were not immediately obvious were determined foremost by their adjacent shadow, otherwise by their shape and texture. Clouds generally appeared homogeneous in most Landsat bands, and

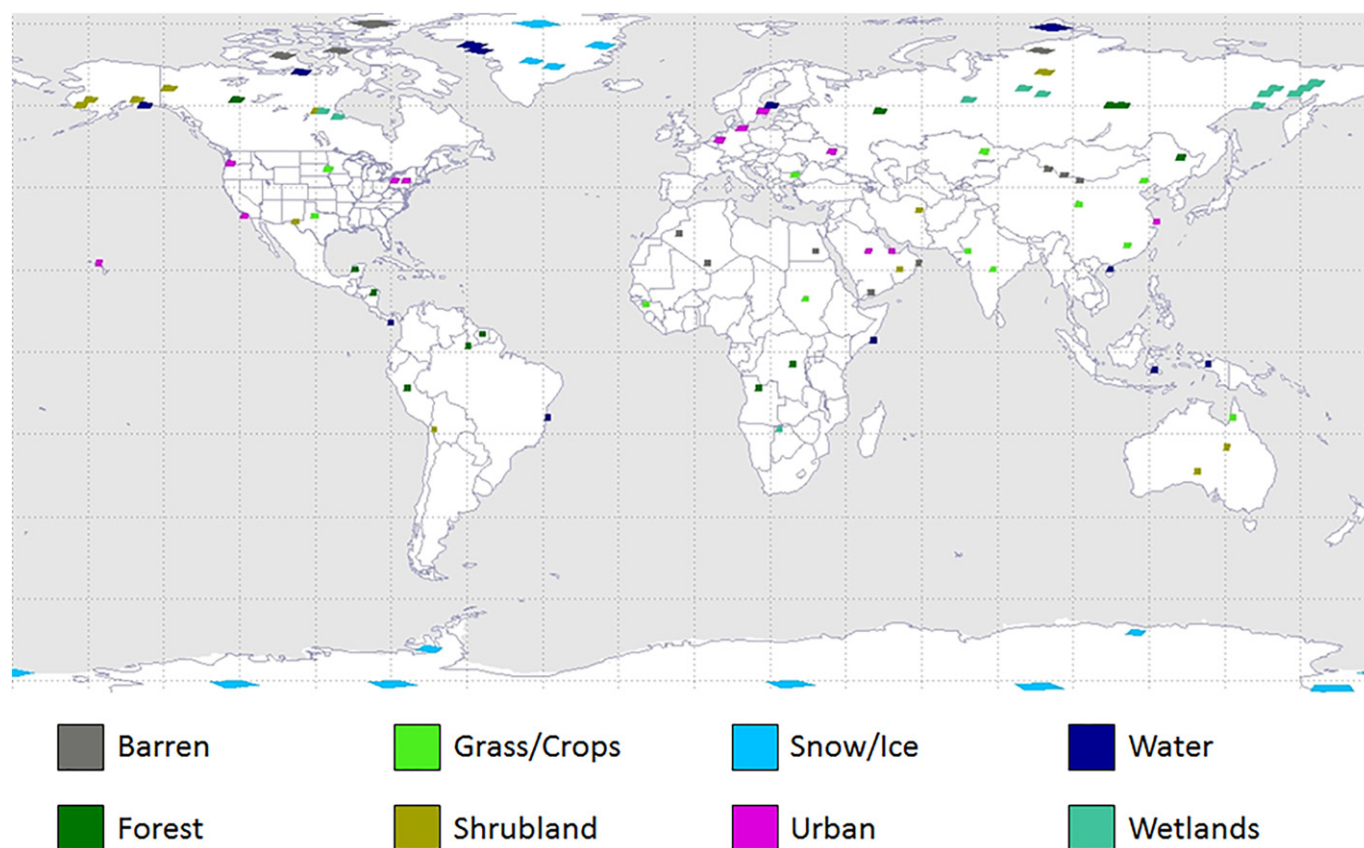


Fig. 2. Global distribution of the 96 unique Landsat 8 Cloud Cover Assessment (CCA) scenes, sorted by International Geosphere-Biosphere Programme (IGBP) biome. Twelve scenes were selected for each of the eight biomes.

were typically an irregular shape in relation to ground features. With these visual criteria, we used the following workflow when creating the cloud masks:

1. Fill pixels around the scene edges were initially subtracted from the scene.
2. Bands were composited to contrast cloud and noncloud features, based upon the conditions listed in Table 2.
3. A contrast stretch was applied until clouds could easily be differentiated from the underlying land and/or water. Clouds ideally appeared both bright and opaque relative to other features. Shadows were generally visible adjacent to clouds, but this was not always true with thinner or higher-altitude clouds. Shadows also varied based upon solar zenith angle and/or ambiguity of cloud shadows versus topographic shadows.
4. Clouds were initially selected using the “magic wand” tool, which selects adjacent pixels based upon a pre-defined homogeneity threshold. The selected clouds were then stored as their own layer and subtracted from the scene. Sometimes the “select by color” tool was more appropriate and faster than the “magic wand” tool.
5. Brighter clouds were removed from the scene, giving thinner clouds greater contrast. The scene's contrast was again stretched, and thin

clouds were selected with the “magic wand” tool. The selected thin clouds were stored in their own layer.

6. The “magic wand” tool often left gaps in the selection of thin clouds. This was solved by expanding all thin cloud edges by one pixel, then contracting them by one pixel, filling any gaps. Sometimes gap filling was skipped, as it was not always necessary.
7. The scene was manually examined and false detections were removed. Generally, these were bright targets in areas such as urban landscapes, snow-covered mountains, and rivers.
8. From the resulting cloud and thin cloud objects, cloud shadows were digitized if easily identified throughout the scene.

Note that a quantitative threshold does not exist to distinguish thin cloud and transparent features such as haze or aerosols, making thin cloud identification inherently subjective to any analyst.

An example of one of the scenes used and the accompanying digitized cloud mask is illustrated in Fig. 3. The “L8 Biome” dataset is currently provided online at no cost (USGS, 2016c).

4. Algorithms

All of our available algorithms were tested against the three unique cloud cover datasets (Table 1; Fig. 1). All of these algorithms

Table 2
Landsat 8 bands and band combinations used to digitize clouds and cloud shadows for the “L8 Biome” dataset.

Target feature	Local conditions	Band(s)	Wavelength(s)
Clouds	Normal	5,3,2 (R,G,B)	Near infrared, green, blue
Clouds	Snow cover	6,3,2 (R,G,B)	Shortwave infrared 1, green, blue
Clouds	Foggy	6,5,3 (R,G,B)	Shortwave infrared 1, near infrared, green
Any (supplementary)	High solar zenith; heavy cloud cover; ice sheets	7; 9; 10; 11 (Grayscale)	Shortwave infrared 2; Cirrus; Thermal 1; Thermal 2

Use of one band combination over another is ultimately subjective and dependent upon local scene conditions.

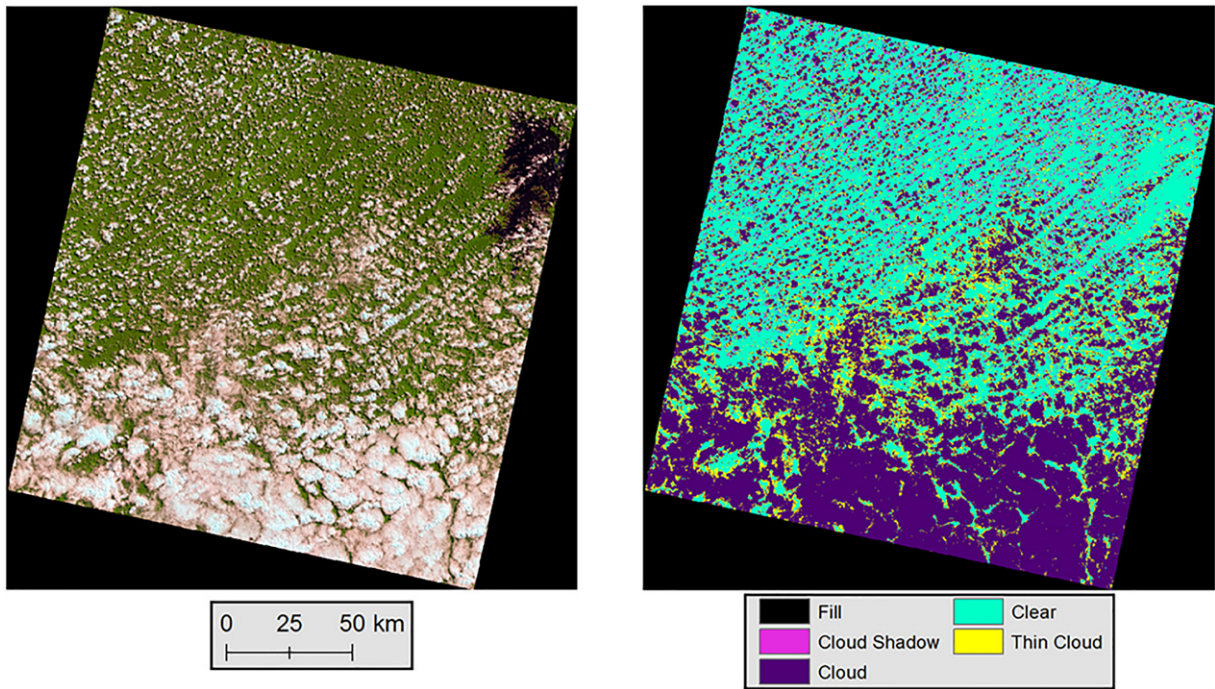


Fig. 3. Left: Landsat 8 Operational Land Imager (OLI) scene used for cloud and cloud shadow mask digitization, acquired over WRS-2 Path 229, Row 57 on 21 May 2014, displayed as a false color composite (bands 6, 5, and 4, respectively). Right: The final “L8 Biome” cloud mask product.

were implemented in our processing infrastructure, and many have been documented in previous literature (Section 4.1–4.6). Here we give special attention to the CFMask algorithm, as it has not been previously documented, but is in use as a quality assurance band provided with higher level Landsat data products, such as surface reflectance. Before this study, a variety of cloud mask algorithms besides Fmask were considered to be a part of either standard Landsat Level-1 or higher level product development. An initial set of algorithms were selected for evaluation based upon documentation in the literature and the ability to have sample cloud masks generated:

- Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) cloud algorithm (Vermote and Saleous, 2007)
- Function of Mask (Fmask) algorithm (Zhu and Woodcock, 2012)
- Spectral Similarity Group (SSG) algorithm (Jin et al., 2013)
- Histogram Fitting for Mapping (HFM) algorithm (Helmer et al., 2009, 2012)
- Web Enabled Landsat Data (WELD) Cloud Cover Assessment (CCA) algorithm (Roy et al., 2010)
- Landscape Fire and Resource Management Planning Tools (LANDFIRE) CCA (Reeves et al., 2006)
- Landsat 7 Automated Cloud Cover Assessment (ACCA) (Irish, 2000; Irish et al., 2006).

The algorithms were evaluated based upon their readiness of implementation, which we defined as being characterized and at least partially validated, being automated (i.e., minimization of sliding parameters), and having documented accuracy assessment. LEDAPS CCA, LANDFIRE CCA, and Fmask fit the majority of those initial specifications. Fmask was ultimately selected because of its active development by the creator, adoption from the academic community, physically meaningful attributes, ongoing development to utilize the Landsat 8 OLI cirrus band, development to support Sentinel-2 bands (Zhu et al., 2015) and a reported high cloud accuracy of 96.41% (Zhu and Woodcock, 2012).

4.1. C implementation of Function of Mask (CFMask)

The C programming language implementation of Function of Mask, known as CFMask, is a translation from Function of Mask, known as Fmask (Zhu and Woodcock, 2012; Zhu et al., 2015). Fmask's source code was translated to C at the USGS Earth Resources Observation and Science (EROS) Center for use in a production environment, though it was initially implemented to support the development of higher level Landsat data products.

4.1.1. Translation of Fmask to C

The original CFMask code was based upon Fmask version 3.2 and has since been modified and enhanced with guidance from both the original author and the scientific community. Fmask has since been upgraded to version 3.3 after enhancements were made to CFMask. Fmask was written in MATLAB and translated to C, ultimately improving algorithm efficiency and run time, which decreased by approximately 90%. To ensure consistency in the translation, we compared the similarity of Fmask and CFMask's output using three test scenes of varying cloud cover and climate conditions (Path 16, Row 41 acquired August 19, 1990, and August 22, 1991; Path 10, Row 54 acquired March 12, 2011). Excluding fill pixels, the disagreeing pixels between the three Fmask and CFMask scenes is 0.44%, indicating that the translation was successful. To ensure CFMask's accuracy and performance were comparable to other incumbent algorithms, it was tested against ACCA (used in Landsat Level-1 processing) and LEDAPS CCA (used in Landsat 4–7 surface reflectance processing) using manually derived truth masks; the results are shown in Table 3. In Zhu and Woodcock (2012) and Zhu et al. (2015),

Table 3
Results from the preliminary three-scene evaluation of the CFMask algorithm.

Name	% Correct	% False	% Omission	% Commission
ACCA	79.90	4.30	7.80	2.30
LEDAPS CCA	85.70	14.30	5.81	28.97
CFMask	90.97	9.03	4.98	16.02

the accuracy assessments were made without any buffering of the cloud mask, while the CFMask results here used the default Fmask parameters that buffered clouds by three pixels in eight connected directions, a process known as *dilation*. Therefore, the reported CFMask overall accuracy will be slightly lower; the overinclusion, or commission error, will be noticeably higher; the underinclusion, or omission error, will be lower than what is reported by Zhu and Woodcock (2012) and Zhu et al. (2015).

4.1.2. CFMask and CFMask cloud confidence bands

We designed CFMask to output two products: the high-confidence cloud bits that are dilated and presented in the final CFMask product (same as Fmask), and the cloud confidence bits derived from the cloud probability code. Cloud dilation fills any “clear” pixels that are surrounded by clouds, reducing uncertainty concerning thin cloud or adjacent-to-cloud pixels. Dilation reduces ambiguity of a cloud-contaminated pixel at the cost of potentially high commission error; this is ideal for land use/land cover change assessments analyzing large spatiotemporal stacks of data, where cloud cover can adversely alter trending analysis. Cloud shadows are derived from the dilated cloud objects (Zhu and Woodcock, 2012). Because shadows are derived from the high-confidence cloud pixels, there is no confidence score associated with cloud shadow.

The cloud confidence band does not use dilation; it provides a “low,” “medium,” and “high” confidence estimation of cloud contamination for every nonfill pixel. Low confidence is generated when the potential cloud score (a unitless value) < 12.5 ; medium confidence is > 12.5 , but < 22.5 ; and high confidence is > 22.5 (Zhu, 2013). These confidence intervals are inserted into the algorithm at the potential cloud layer, which is described in Zhu and Woodcock (2012), Eq. (1).

4.1.3. CFMask with cirrus (CFMask T-Cirrus)

CFMask has also been designed to run using the cirrus band to enhance the potential cloud layer and probability of cloud over land and water for Landsat 8 imagery, using thresholds described in Zhu et al. (2015), Eq. (1). Here we implement the cirrus band to potentially reduce omission error.

4.1.4. Nonthermal CFMask (CFMask NT) and nonthermal CFMask with cirrus (CFMask NT-Cirrus)

Nonthermal Fmask was designed to potentially support the European Space Agency (ESA) Sentinel-2, a platform without thermal detectors (Drusch et al., 2012), and to support potential failure of the TIRS instrument onboard Landsat 8. TIRS has known issues pertaining to the instrument's onboard electronics controlling the scene select mechanism (SSM; USGS, 2016d), as well as issues with stray light. The stray light issues are more prevalent in TIRS band 11, preventing the accurate retrieval of land surface temperature through a split-window atmospheric correction technique (Montanaro et al., 2014). The thermal band is necessary to accurately predict cloud location relative to clear pixel (surface) brightness temperature, and to estimate cloud height, which is essential to determining cloud shadow.

To test nonthermal CFMask's accuracy, we made two different prototype CFMask processing flows and applied them to Landsat 8 data. The first prototype, CFMask NT, removed any thermal threshold tests from the original CFMask code. The second prototype, CFMask NT-Cirrus, used the first prototype, but replaced the cloud probability code over water and over land with cirrus cloud probability tests, exactly as described in Zhu et al. (2015).

4.2. See5

The See5 CCA algorithm was developed from RuleQuest's C5.0 machine-learning regression software package. Based upon a training dataset, See5 is designed to classify pixels and assign confidence levels

based upon specific spectral characteristics. The See5 model contains 245 final branches in its decision tree and was trained with 7 dimensions in its dataset (Landsat 7 ETM+ bands 1 through 5, band 7, and solar elevation angle) (Scaramuzza et al., 2012). In Landsat 8 Level-1 processing, See5 is always run and used to generate the cloud score.

4.3. Automated Cloud Cover Assessment (ACCA)

The ACCA algorithm was originally designed by Hollingsworth et al. (1996) and serves as the Landsat TM and ETM+ cloud detection algorithm (Arvidson et al., 2006), which is used for providing a scene-wide cloud score that is used as searchable metadata for the archive (Irish, 2000). ACCA identifies and refines cloud objects using several passes, which include a per-pixel decision tree and statistical aggregation of the decision tree's output to reduce ambiguity from the first pass.

Landsat's OLI spectral bandpasses were refined from the Landsat TM/ETM+ design to enhance vegetation detection and largely reduce water vapor absorption (Irons et al., 2012), but did not contain thermal infrared bands. Therefore, the ACCA implementation was updated for Landsat 8 to detect clouds without thermal data (Scaramuzza et al., 2012; Sections 4.4 and 4.5). The algorithm implemented for Landsat 8 ACCA is an expansion of the first pass used in the Landsat TM and ETM+ ACCA (USGS, 2016e). ACCA is the default CCA in Landsats 4–7 Level-1 processing, and is applied in Landsat 8 Level-1 production if thermal data are present, and is weighted against See5.

4.4. Artificial Thermal-Automated Cloud Cover Algorithm (AT-ACCA)

The expansion of ACCA also included the creation of an algorithm to be run without thermal data. This is accomplished by creating an artificial thermal (AT) band, thus renaming the algorithm to AT-ACCA. The AT band was derived using a regression model and replaces the thermal band in every instance of ACCA's first pass. In this study we used the “Expanded AT-ACCA” algorithm, which utilizes a neural network designed in RuleQuest's Cubist software to clean up ambiguous pixels. In Landsat 8 Level-1 processing, AT-ACCA is used when no thermal data are available. The validation assessment performed on the Expanded AT-ACCA routine, using the same “L7 Irish” scenes as used here, placed it as 89.7% accurate (Scaramuzza et al., 2012).

4.5. Fixed Temperature-Automated Cloud Cover Algorithm (FT-ACCA)

Fixed Temperature (FT) ACCA uses an approximation of the global annual mean temperature (288 K) as the brightness temperature. FT-ACCA was designed to assess the effects of no thermal data in the ACCA algorithm and will never be considered for implementation in a production environment. Here we implement it to simply illustrate the performance impact of losing thermal data in ACCA. FT-ACCA's previous validation assessment placed it at 68.7% accurate (Scaramuzza et al., 2012).

4.6. Landsat 8 Surface Reflectance Code (LaSRC) Cloud Cover Algorithm (CCA)

The Landsat 8 Surface Reflectance Code (LaSRC) is a provisional surface reflectance retrieval algorithm that utilizes the improved radiometry, higher signal-to-noise ratio, and newer bands aboard Landsat 8, specifically the coastal aerosol and cirrus bands. A cloud masking algorithm was developed in LaSRC as a cursory check for clouds and cirrus clouds before surface reflectance is retrieved (Vermote et al., 2016). LaSRC uses the blue, red, cirrus, and TIRS 1 bands to create a cloud mask. The TIRS 1 band is used to estimate cloud height and a potential cloud shadow, and the green, red and shortwave infrared bands use fixed thresholds to determine cloud shadow presence. Here we use results from LaSRC version 0.2.0.

Table 4

Criteria used in CCA validation masks against the output from the CCA algorithms to classify cloud pixels as correct, false, or ambiguous (“Ambig”).

Manual mask	Algorithm correct if:	Algorithm false if:	Algorithm Ambig if:
Thick	Thick	Clear or shadow	Ambig or medium
Thin	Thick or Ambig or Medium	Clear or shadow	–
Clear	Clear or Shadow	Thick	Ambig or medium
Shadow	Clear or Shadow	Thick	Ambig or medium

Note that “medium” confidence denotes a thin or ambiguous cloud, and not all algorithms determine ambiguity.

5. Results

5.1. All validation masks

Here we examine all three cloud masks – “L7 Irish,” “L8 SPARCS,” and “L8 Biome” – to establish a baseline accuracy assessment using all available cloud and cloud shadow validation data. An assessment of individual biomes using the “L8 Biome” cloud and cloud shadow dataset is covered in Section 5.2.

5.1.1. Accuracy assessment (clouds)

Each cloud masking algorithm was applied to the “L8 Biome,” “L8 SPARCS,” and “L7 Irish” datasets. Many algorithms tested here produce slightly different classifications, and have to be connected with the corresponding validation masks accordingly. CFMask does not assign bits that would qualify as being ambiguous (“Ambig”), but the CFMask confidence band assigns “medium” confidence, which we considered ambiguous for “thick” cloud, but correct for “thin” clouds. Thin clouds can be correct for both “high” and “medium” confidence. “Low” confidence is correct only if the pixel is cloud shadow or clear. The conditions for true, false, and ambiguous assignments are detailed in Table 4.

Percent correct for cloud is calculated using Eq. (1):

$$\%Correct = \frac{thick_as_thick + thin_as_thick + clear_as_clear + shadow_as_shadow}{npix} \quad (1)$$

where *thick_as_thick*, *thin_as_thick*, *clear_as_clear*, and *shadow_as_shadow* (where applicable) are the number of pixels correctly identified, and *npix* is the number of valid pixels in the scene. Omission error is calculated using Eq. (2):

$$\%Omission = \frac{total_cloud - cloud_as_cloud}{total_cloud} \times 100 \quad (2)$$

where *cloud_as_cloud* is the number of pixels correctly identified as cloud, and *total_cloud* is the total number of pixels correctly and incorrectly flagged as cloud. Commission error is calculated using Eq. (3):

$$\%Commission = \frac{clear_as_cloud + shadow_as_cloud}{total_clear + total_shadow} \times 100 \quad (3)$$

where *clear_as_cloud* and *shadow_as_cloud* are the number of pixels incorrectly flagged as cloud, and *total_clear* and *total_shadow* are the total number of pixels correctly and incorrectly flagged as clear or shadow.

The results for all algorithms and validation masks are shown in Table 5. Overall, CFMask (89.3%) and the CFMask cloud confidence band (90.5%) were the most accurate. Predictably, the lack of dilation in the confidence band slightly raises the overall omission error (+0.9%) but drops the commission error (−2.0%). The addition of the cirrus band threshold to CFMask T-Cirrus reduced accuracy (−4.2%) and raised commission error (+11.1%), but also lowered omission error (−1.8%). The associated confidence bits for CFMask T-Cirrus recover some accuracy (+0.9%) and decreased commission (−1.9%) but increased omission (+6.4%) error. ACCA (83.8% overall) had a less strict rule set, resulting in more ambiguity (+9.8%) than the thermal-based CFMask algorithms. See5 (85.8% overall) was more conservative, as it tended to omit more cloud pixels (+14.8%) than ACCA or CFMask.

Of the nonthermal algorithms (AT-ACCA, CFMask NT series, CFMask NT-Cirrus series, FT-ACCA and See5), AT-ACCA had the best overall accuracy (87.5%) and lowest percent of ambiguous pixels (1.3%) of any nonthermal algorithm. AT-ACCA had a distinct advantage over the other listed algorithms, as it was trained to clean up ambiguous ACCA pixels without the use of the thermal band. When thermal was removed from CFMask, the omission error increased by as much as 38% over the original CFMask algorithm (+34.6% overall) and its confidence band (+41.0% overall). The CFMask NT algorithms' overall performance improved with the implementation of the cirrus band, bringing CFMask up +4.8% (83.9% overall) and the CFMask confidence band up +5.2% (84.8% overall). Contrary to the results here, Zhu et al. (2015) found adding cirrus detection to nonthermal Fmask (simulated Sentinel-2

Table 5

Results of tested cloud algorithms and the respective masks used to perform each test, where “I” is “L7 Irish,” “S” is “L8 SPARCS,” and “B” is “L8 Biome.”

Name	Masks used	% Correct	% False	% Ambig.	% Omission	% Commission
ACCA	I,S,B	83.83	6.39	9.78	6.66	5.85
AT-ACCA	I,S,B	87.52	11.21	1.27	12.42	9.83
CFMask	I,S,B	89.29	10.71	0.00	2.72	12.02
CFMask cloud conf.	I,S,B	90.45	8.16	1.38	3.65	10.06
CFMask T-Cirrus	S,B	85.14	14.86	0.00	0.94	23.05
CFMask T-Cirrus conf.	S,B	86.05	12.42	1.53	7.35	21.19
CFMask NT	I,S,B	79.15	20.85	0.00	34.56	8.38
CFMask NT cloud conf.	I,S,B	79.60	14.70	5.69	41.02	8.07
CFMask NT-Cirrus	S,B	83.90	16.10	0.00	5.41	22.67
CFMask NT-Cirrus conf.	S,B	84.81	12.08	3.11	12.48	20.92
FT-ACCA	I,S,B	74.20	5.84	19.96	8.07	3.75
LaSRC CCA	S,B	73.07	19.03	7.91	4.70	23.90
See5	I,S,B	85.79	11.72	2.50	14.83	5.71

Note that CFMask NT-Cirrus series, CFMask T-Cirrus series, and LaSRC CCA algorithms could not be run with the “L7 Irish” dataset, as they are Landsat ETM+ derived, therefore lacking the cirrus band.

Table 6
Results of tested cloud shadow algorithms and the respective masks used to perform each test.

Name	Masks used	Total Correct	% False	% Omission	% Commission
CFMask	I,S,B	96.17	3.83	13.39	1.32
CFMask NT	I,S,B	96.12	3.88	19.66	3.16
CFMask NT-Cirrus	S,B	95.58	4.42	21.74	3.31
CFMask T-Cirrus	S,B	96.04	3.96	19.20	2.94
LaSRC	S,B	93.85	6.15	26.06	4.77

images) was more accurate than the thermal-based Fmask algorithm run over a similar Landsat 4–7 scene. However, they used a preliminary accuracy analysis with point-based classifications, while we use per-pixel masks covering entire scenes (or subset tiles in the case of “L8 SPARCS”).

The overall least accurate algorithms are LaSRC CCA (73.1%) and FT-ACCA (74.2%). While FT-ACCA has a relatively low percentage of false pixels (5.8%), it has the highest percentage of ambiguous pixels (20.0%), indicating that a single thermal threshold value is not meaningful enough for the algorithm to decisively select a cloud versus noncloud pixel. LaSRC CCA has routinely high commission error (23.9%) and was the algorithm with the overall lowest accuracy.

5.1.2. Accuracy assessment (cloud shadows)

Not all algorithms used for the above CCA tests are capable of detecting cloud shadow. Only five algorithms were compared with the validation masks, and the results are reported in Table 6. Percent correct for cloud shadow is calculated using Eq. (4):

$$\%Correct = \frac{shadow_as_shadow + ((total_thick + total_thin + total_clear) - (thick_as_shadow + thin_as_shadow + clear_as_shadow))}{npix} \quad (4)$$

where *shadow_as_shadow* is the number of pixels correctly identified as shadow; *total_thick*, *total_thin*, and *total_clear* are the total number of pixels correctly and incorrectly flagged as thick/thin cloud or clear; *thick_as_shadow*, *thin_as_shadow*, and *clear_as_shadow* are the number of pixels incorrectly flagged as shadow; and *npix* is the number of valid pixels in the scene. Omission error is calculated using Eq. (5):

$$\%Omission = \frac{total_shadow - shadow_as_shadow}{total_shadow} \times 100 \quad (5)$$

where *total_shadow* is the total number of pixels correctly and incorrectly flagged as shadow. Commission error is calculated using Eq. (6):

$$\%Commission = \frac{thick_as_shadow + thin_as_shadow + clear_as_shadow}{total_thick + total_thin + total_clear} \times 100 \quad (6)$$

All tested algorithms have a high success rate of detecting cloud shadows (between 93 and 96%) relative to their cloud detection accuracy. While commission rates are low (between 1 and 4%), omission rates are higher (between 13 and 26%), notably when cirrus cloud test(s) are implemented and/or when thermal test(s) are removed. Some accuracy issues can be attributed to cloud accuracy, as all tested algorithms estimate cloud shadow position partially using the cloud objects. Similar to the cloud assessment results, LaSRC CCA performed the worst in terms of accuracy and error among all tested algorithms, while performance was similar among CFMask-based algorithms, though tests lacking the thermal band performed worse than those including the thermal band.

5.2. “L8 Biome” masks

5.2.1. Dataset creation

The “L8 Biome” dataset exhibited very little qualitative clustering (Fig. 2). Unlike the “L7 Irish” dataset, the “L8 Biome” dataset did not have extensive clustering in the Austral regions (i.e., Chile and New Zealand), but did not have the uniform stratification of the “L8 SPARCS” dataset (Fig. 1). There was some visible spatial clustering in the Wetlands biome, specifically in northeastern Siberia. Globally, many wetlands are not as comparable in size as in this region, making clustering more likely to occur.

5.2.2. Scene selection

The design of the “L8 Biome” CCA scene selection process allowed for an alternate view of the accuracy and strengths of each cloud masking algorithm in terms of the path/row's primary biome and each scene's overall cloud cover. We only assigned one biome per scene, though there were nearly always two or more biomes in a scene. Thus, our curated dataset ignores potential influence from secondary and/or tertiary biomes. One example of this issue was with the cluster of Wetlands in Siberia (Fig. 2), as this area can be completely snow and/or ice covered during winter seasons, causing this biome to be potentially misrepresented in this dataset. Likewise, our Water biome scenes exist adjacent to all other biomes, so a mixture of scenes are represented within our sample Water scenes, such as Barren and Snow/Ice in Greenland, Urban and Forest in Brazil, and Barren in Somalia. Similarly, the Urban biome is never a majority component within a path/row, so Urban scenes include a mix of biomes surrounding the urban area.

5.2.3. Accuracy assessment (clouds)

Due to the relatively poor performance of the CFMask NT series, LaSRC CCA, and FT-ACCA (Table 5), they were excluded from the biome-based analysis. Accuracy of the remaining algorithms was evaluated by both individual biome and individual cloud bin (Table 7). Clouds were often confused with Snow/Ice, but See5 did so less frequently. A reduction in See5's Snow/Ice commission came with a reduction of accuracy over heavily cloud-covered scenes and/or heavily vegetated regions, but performed the best over clear scenes. CFMask and its confidence band performed similarly across all biomes and cloud cover bins, but the confidence band was more accurate than the CFMask band in every biome except Barren. The CFMask NT-Cirrus conf. outperformed CFMask NT-Cirrus in every biome except Shrubland. CFMask T-Cirrus and CFMask T-Cirrus conf. outperformed the nonthermal CFMask NT-Cirrus and CFMask NT-Cirrus conf. in every biome except Shrubland, Snow/Ice, and Urban. The CFMask T-Cirrus algorithms outperformed all other algorithms over “Cloudy” scenes by at least 3%. Thermal-based CFMask outperformed ACCA (the incumbent Level-1 algorithm) in every scenario except Water biomes and sometimes Snow/Ice. See5 was routinely less accurate than thermal-based CFMask in every biome, except Wetlands and Snow/Ice. Most cloud algorithms' errors increased as the number of clouds increased within the scene. Where some algorithms appear to have high overall correctness, some often flagged pixels as ambiguous; this initially appeared to increase accuracy, but in practice reduced the number of correctly identified cloud pixels. Ambiguity remained high with the more overall inaccurate

Table 7

Percent correct, percent ambiguous, omission error and commission error cloud pixels, sorted by both biome and cloud cover type.

Name		Barren	Forest	Grass/Crops	Shrubland	Snow/Ice	Urban	Water	Wetlands	Clear	MidClouds	Cloudy
ACCA	% Correct	86	87.48	83.05	86.6	67.93	86.61	92.23	86.34	84.03	86.45	82.88
	% Ambig.	13.38	12.27	9.26	19.33	1.56	12.78	8.8	14.16	9.37	7.49	17.47
	% Omission	2.06	6.2	9.13	2.13	0.68	3.82	1.33	0.76	4.01	1.17	4.62
	% Commission	7.54	0.32	2.66	4.67	29.44	1.68	10.67	8.94	7.18	4.26	13.29
AT-ACCA	% Correct	88.94	89.68	89.23	92.48	68.3	93.4	95.05	93.83	90.63	87.1	88.59
	% Ambig.	2.2	0.72	0.68	1.83	0.35	0.87	1.03	1.32	0.52	1.06	1.79
	% Omission	8.07	8.65	10.7	4.36	1.55	11.49	3.23	1.69	10.25	3.13	5.27
	% Commission	12	6.48	3.76	15.64	30.54	6.98	20.28	14.57	8.25	9.21	23.88
cfmask	% Correct	92.39	93.83	94.07	87.3	64.09	92.05	86.36	92.38	92.91	86.79	85.74
	% Ambig.	0	0	0	0	0	0	0	0	0	0	0
	% Omission	0.57	1.07	1.97	3.27	13.41	3.39	8.39	0.09	6.66	0.22	4.42
	% Commission	14.72	25.49	6.36	8.5	24.31	13.94	26.99	26.32	5.98	20.21	28.99
cfmask_conf	% Correct	84.62	96.29	94.51	87.19	68.23	93.95	88.05	93.68	93.77	86.84	88.69
	% Ambig.	1.14	1.99	1.34	1.62	6.67	1.36	2.55	1.4	1.11	2.12	3.29
	% Omission	1.15	3.39	4.48	4.87	14.35	6.17	9.8	0.35	10.8	0.69	4.71
	% Commission	15.55	24.19	5.1	8.4	26.46	9.88	24.59	24.48	5.18	15.71	31.03
cfmask_nt_cirrus	% Correct	76.64	88.21	89.71	91.38	48.42	92.8	91.17	81.15	82.89	87.14	81.1
	% Ambig.	0	0	0	0	0	0	0	0	0	0	0
	% Omission	5.95	1.78	10.37	2.07	0.76	4.06	5.19	9.58	3.18	5.6	6.3
	% Commission	16.46	29.23	8.63	17.73	63.53	14.37	35.75	24.41	17.33	27.29	33.22
cfmask_nt_cirrus_conf	% Correct	84.39	90.84	89.26	92.61	50.56	94.87	93.39	82.38	84.01	86.94	83.37
	% Ambig.	5.03	5	3.66	3.44	3.48	2.46	2.05	9.67	2.86	5.09	5.08
	% Omission	7.9	21.04	17.13	4.55	25.06	10	8.19	10.72	10.72	15.14	13.36
	% Commission	17.02	26.01	8.46	18.24	65.7	9.21	34.11	26.85	16.73	22.58	37.79
cfmask_t_cirrus	% Correct	90.84	92.08	90.28	88.22	45.07	91.39	92.46	86.83	85.04	90.12	80.18
	% Ambig.	0	0	0	0	0	0	0	0	0	0	0
	% Omission	0.23	1.98	0.72	2.44	5.7	3.63	1.85	0.06	15.3	32.3	32.6
	% Commission	20.79	25.32	11.58	13.03	63.47	15.34	33.4	33.5	15.3	32.3	32.6
cfmask_t_cirrus_conf	% Correct	91.22	93.51	90.74	89.56	43.58	93.4	94.37	87.66	85.35	90.02	82.58
	% Ambig.	0.83	3.24	0.99	1.9	3.77	1.2	1.64	3.3	2.11	1.86	2.3
	% Omission	0.47	22.21	1.78	3.28	25.09	6.32	3.46	0.21	10.86	6.96	5.12
	% Commission	20.55	24.07	10.21	12.04	67.48	11.16	32.03	32.01	14.9	27.18	35.51
See5	% Correct	78.75	78.69	73	81.63	82.19	83.91	81.14	89.93	94.68	66.95	84.96
	% Ambig.	1.57	4.44	2.77	3.27	5.12	4.01	2.62	3.2	2.4	3.31	4.41
	% Omission	13.21	18.74	27.37	24.44	32.13	15.99	17.28	8.07	15.42	13.55	30
	% Commission	9.78	3.63	0.71	11.46	9.46	6.51	21.89	10.49	1.95	8.07	17.69

algorithms, ACCA and CFMask NT-Cirrus, where an average of 5–10% of pixels was flagged as ambiguous.

Errors of omission and commission are also shown in terms of biome and cloud bin (Table 8). Overall, commission was the greatest issue in Snow/Ice regions, where spectral signatures of clouds in the visible bands were similar to Snow/Ice. In the Snow/Ice biome, See5 was conservative, but also inaccurate. By cloud bin, the greater the amount of clouds in the scene, the greater the chance an algorithm has to make commission errors.

5.2.4. Accuracy assessment (cloud shadows)

As in Section 5.2.3, the cloud shadow results are analyzed by error in both biome and cloud bin categories (Table 8). Unlike in the cloud

results, LaSRC was compared for cloud shadow due to the low number of algorithms to detect cloud shadow. Ambiguous was not an applicable category, as none of the algorithms contain the ability to assign confidence to cloud shadow. Cloud shadow accuracy often correlated with cloud accuracy (“% Correct”); however, the addition of the cirrus test to both thermal and nonthermal-based CFMask variants decreased overall cloud shadow accuracy in Snow/Ice and was within about 1.5% of Urban. However, the cirrus tests improved cloud shadow accuracy over CFMask NT in all other biomes. Scenes with “MidCloud” cover were typically the least accurate, as these conditions warrant the most cloud shadows. Omission error exceeded commission error in every biome and cloud cover type, never exceeding 6% error. Omission error was typically greatest in CFMask NT, likely due to missing thermal

Table 8

Percent correct, omission error, and commission error of cloud shadow pixels, sorted by both biome and cloud cover type.

Name		Barren	Forest	Grass/Crops	Shrubland	Snow/Ice	Urban	Water	Wetlands	Clear	MidClouds	Cloudy
cfmask_nt	% Correct	97.41	93.99	97.57	95.44	94.54	95.99	95.82	96.73	97.66	97.39	92.85
	% Omission	14.81	25.42	49.74	20.23	39.42	16.22	19.19	11.38	37.59	10.30	24.86
	% Commission	1.7	4.21	0.9	4.14	2.04	3.51	3.8	2.97	1.3	2.17	5.26
cfmask_t_cirrus	% Correct	96.79	94.04	97.51	95.56	94.29	97.08	96.52	96.86	97.61	97.01	93.71
	% Omission	10.41	24.88	30.17	11.01	38.11	14.35	15.66	4.95	27.52	7.21	20.39
	% Commission	2.5	4.07	1.57	4.25	2.24	2.56	3.17	3	1.52	2.65	4.6
LaSRC	% Correct	95.83	91.83	90.74	97.76	90.00	96.69	93.93	97.30	94.32	94.09	94.56
	% Omission	18.12	34.96	46.94	17.23	38.37	16.27	21.21	7.42	37.55	11.47	25.04
	% Commission	1.79	5.04	7.05	1.9	9.13	2.81	5.51	2.17	4.28	4.89	3.95
cfmask	% Correct	96.48	94.67	97.41	95.54	96.29	97.01	96.63	97.30	98.19	97.26	93.96
	% Omission	10.97	27.10	31.28	8.73	35.99	14.22	15.50	4.86	28.46	7.77	18.04
	% Commission	2.78	3.31	1.68	4.42	1.96	2.62	3.06	2.57	0.95	2.39	5.02
cfmask_nt_cirrus	% Correct	95.44	93.83	96.76	95.39	94.59	96.74	96.24	96.04	97.33	97.10	92.53
	% Omission	10.00	25.50	38.62	11.96	37.66	15.47	19.06	10.56	29.90	9.00	23.53
	% Commission	3.9	4.32	1.93	4.39	1.96	2.77	3.39	3.69	1.75	2.53	5.61

data to provide accurate cloud heights, and general inaccuracy of the cloud objects used to draw the potential shadow's footprint.

6. Discussion

6.1. Algorithm suitability

As illustrated in the results, some cloud and cloud shadow algorithms have strengths or weaknesses in each biome and cloud cover type. To simplify the results, we use a formula to derive a metric to estimate algorithm-environment suitability:

$$\%Correct - (\%Omission + \%Commission + \%Ambiguous) \quad (7)$$

Here we apply Eq. (7) to the results derived from our tested cloud algorithms applied to our “L8 Biome” validation masks. The same is applied to the cloud shadow results, except with the “% Ambiguous” term removed, as our tested shadow algorithms only produce a positive or negative result:

$$\%Correct - (\%Omission + \%Commission) \quad (8)$$

The results strongly favor the CFMask heritage algorithms, though detecting cloud pixels without thermal data is best with AT-ACCA, and areas of high commission for thermal-based CFMask algorithms (Snow/Ice and Water) are better performed by ACCA (Table 9). In almost every instance of CFMask being used in non-thermal applications, the addition of the cirrus band nearly always produces the best results, except over Forest and Wetlands. For cloud shadows, the CFMask algorithms unanimously performed the best.

6.2. Challenges and future development

6.2.1. Known issues

The USGS implementation of CFMask has been provided as one of the quality assurance bands included with the provisional Landsat surface reflectance products, and science users have acknowledged several weaknesses of the algorithm. CFMask has higher cloud commission rates over active fires because smoke resembles the shape and spectral response of thin clouds, but bears a different thermal signal, making it more difficult to differentiate from land. Smoke is also not generally imageable in the shortwave infrared bands (e.g., Schroeder et al., 2015), which many of our tested cloud algorithms use in one or more spectral tests. Sun glint on water bodies at lower latitudes, foam along coastlines, terrain shadow and snow and/or ice are other features that frequently contribute to commission error. Ultimately, region-based algorithm modifications may be necessary to circumvent some of these issues.

Initial identification of ground targets known for high commission error for both cloud and cloud shadow, such as water, could be performed and then cleaned up within a certain distance of the feature (Hughes and Hayes, 2014). Terrain shadow cleanup can also be accomplished using a digital elevation model (DEM) and solar elevation information (e.g., Jin et al., 2013). For Snow/Ice commission, additional tests could include an increase in the Normalized Difference Snow Index (NDSI) threshold (Hall et al., 1998), verification of cloud objects with corresponding cloud shadows (Choi and Bindenschadler, 2004; Jin et al., 2013), or temporal composites of summer season pixels and a decision tree to remove incorrectly flagged cloud (Selkowitz and Forster, 2015) and/or cloud shadow pixels.

Multiple acquisitions, possibly from one or more satellites, could potentially be used to verify cloud and cloud shadow masking; however, this method is vulnerable to interference from phenomena such as weather, seasonality, natural disaster, anthropomorphic modification, and/or pixels marked incorrectly as “clear” in the composite or “truth”

Table 9
The most suitable cloud (top) and cloud shadow (bottom) algorithm for the listed biome or cloud cover. Suitability was determined using data from Tables 7 and 8 applied to Eqs. (1–3) for cloud (top) and Eqs. (4–6) for cloud shadow (bottom).

	Barren	Forest	Grass/Crops	Shrubland	Snow/Ice	Urban	Water	Wetlands	Clear	MidClouds	Cloudy
Cloud	Best overall	CFMask	AT-ACCA	CFMask	CFMask	CFMask c	ACCA	AT-ACCA	CFMask	CFMask NT-Cc	CFMask NT-Cc
	Best thermal	CFMask	ACCA	CFMask	ACCA	CFMask c	ACCA	Sec5	CFMask	CFMask c	CFMask NT-Cc
	Best nonthermal	AT-ACCA	AT-ACCA	CFMask NT-C	AT-ACCA	CFMask NT-C	AT-ACCA	AT-ACCA	AT-ACCA	CFMask NT-Cc	CFMask NT-C
	Best overall	CFMask T-C	CFMask T-C	CFMask	CFMask	CFMask	CFMask	CFMask	CFMask	CFMask	CFMask T-C
Cloud shadow	Best overall	CFMask T-C	CFMask T-C	CFMask	CFMask	CFMask	CFMask	CFMask	CFMask	CFMask	CFMask T-C
	Best Thermal	CFMask NT-C	CFMask NT	CFMask NT-C	CFMask NT-C	CFMask NT-C	CFMask NT-C	CFMask NT	CFMask NT-C	CFMask NT-C	CFMask NT-C
	Best nonthermal	CFMask NT-C	CFMask NT	CFMask NT-C	CFMask NT-C	CFMask NT-C	CFMask NT-C	CFMask NT	CFMask NT-C	CFMask NT-C	CFMask NT-C
	Best overall	CFMask NT-C	CFMask NT	CFMask NT-C	CFMask NT-C	CFMask NT-C	CFMask NT-C	CFMask NT	CFMask NT-C	CFMask NT-C	CFMask NT-C

C = Cirrus; c = confidence; T = thermal; NT = nonthermal.

scene. This compositing scheme would ideally have to perform with and without thermal data and be functional across all biomes, solar angles and altitude conditions to be deemed a sufficient companion to CFMask. Furthermore, an operational, archive-wide application of such a compositing algorithm would be computationally expensive relative to the per-scene cloud detection routines already in place. A data cube environment where Landsat scenes are dissolved into a single, continuous grid (e.g., Lewis et al., 2016) would be more suitable for such an application.

6.2.2. Cirrus band implementation

The cirrus band is a viable solution to operating a cloud and cloud shadow masking algorithm without the availability of thermal bands; however, it does not solve for omission of low altitude clouds, which subsequently contribute to cloud shadow omission. While we overcame the initial implementation hurdles, several caveats of the Landsat 8 OLI cirrus band must be acknowledged. The cirrus band's onboard signal-to-noise ratio (SNR), calculated using typical spectral radiance levels for the cirrus band's center wavelength (1373 nm), has the highest coherent noise of any Landsat 8 OLI band (160), which is influenced by crosstalk from the shortwave infrared 1 band (Morfitt et al., 2015). However, the cirrus band's SNR remained stable during the span of the Morfitt et al. (2015) study – from Landsat 8's launch in February 2013 to 2015. The cirrus band is the most reliable method for mitigating any issues with Landsat 8 TIRS and cloud masking algorithms (e.g., Zhu et al., 2015); furthermore, it is necessary for cloud masking algorithms for the ESA's Sentinel-2 MultiSpectral Instrument (MSI), as both Sentinel-2a and Sentinel-2b lack thermal detectors (Drusch et al., 2012). In our implementation, the cirrus data helped increase the accuracy where thermal was excluded from the algorithm. However, when thermal and cirrus are used simultaneously, the accuracy decreased about 4–5% from the cirrus band, introducing commission error, particularly over the Snow/Ice biome. Any implementation of the cirrus band for cloud shadow only increased commission and omission error.

6.2.3. Trained algorithm development

The masks derived in this study were designed for potential use in training a neural network to detect clouds. There is always potential to overfit training data, in this case to a specific cloud type, shape, or underlying biome conditions. One solution to overfitting could be to “regularize” the training dataset by intentionally inserting noise (Bishop, 1995). We observed marginally improved accuracy of AT-ACCA when used on Landsat 8 OLI-based cloud masks (“L8 SPARCS” at 89.8%; “L8 Biome” at 88.8%) over the Landsat 7 ETM+ cloud masks (“L7 Irish” at 86.4%), as the AT routine was originally trained with part of the “L7 Irish” dataset (Scaramuzza et al., 2012). The approximately 8–16 times lower SNR of the Landsat 7 ETM+ bands (Morfitt et al., 2015) is likely sufficient enough to regularize the AT-ACCA for use on Landsat 8 OLI data. Therefore, this methodology should be considered for future datasets whose target dataset contains greater noise than the training dataset; Bishop (1995) outlines how to systematically introduce noise into a dataset to prevent overfitting.

Although AT-ACCA was relatively accurate both here and in Scaramuzza et al. (2012), the future applicability of an algorithm trained with machine-learning is rather limited: if bandwidths are sharpened and bands are added and/or removed from future instrument suites, a new series of training and validation masks must be derived to train and validate the new algorithm. This inevitably introduces more uncertainty and subjectivity to the final product. This may also be true for algorithms designed with inherent knowledge and/or physical characteristics, such as CFMask, which require some degree of biased information. However, the ability to alter an algorithm on a test-by-test basis using a priori knowledge of a cloud or cloud shadow may be more desirable to maintain an already accurate algorithm for multiple sensors. This would be accomplished by altering the tests with knowledge of the

spectral response relative to the bands in the sensor, allowing the algorithm to perform similarly to its heritage algorithm, thus maintaining consistency between past, current, and future quality assurance information for all Landsat data products. Here, replacing AT-ACCA with its next best performing algorithm, CFMask NT Cirrus conf., comes at a trade-off in overall accuracy (–2.7%) and more so in Snow/Ice and Water biomes.

7. Conclusion

In this study, we used pre-existing cloud validation masks derived from Landsat 7 ETM+ (“L7 Irish”) and Landsat 8 OLI data (“L8 SPARCS”), and combined them with our own newly derived Landsat 8 OLI masks (“L8 Biome”) to validate 13 cloud (Tables 5 and 7) and 5 cloud shadow (Tables 6 and 8) masking algorithms, using a total of 278 validation masks (Table 1; Fig. 1). We ultimately selected our new validation scenes by biome classification, and manually digitized them to represent cloud, thin cloud, clear, and some cloud shadow classes. Together, these datasets represent the most comprehensive cloud assessment truth masks available for any satellite data archive.

We validated algorithms that already existed in both experimental and operational environments (Section 4), though we chose to implement Fmask (Zhu and Woodcock, 2012) in an operational environment, where we renamed it CFMask and created variants of it to produce cloud confidence bits, incorporate a cirrus test, remove the thermal tests from the algorithm, and add cirrus band tests to recover some accuracy from the lack of thermal data (Table 5).

Our validation results found CFMask's confidence band was the overall most accurate among the thermally based algorithms; AT-ACCA was the overall most accurate nonthermal-based algorithm (Table 5). The addition of a cirrus band threshold to CFMask not only decreased accuracy but also decreased omission error. When a cirrus test was applied to the confidence band (CFMask T-Cirrus conf.), it decreased overall accuracy and introduced more omission and commission error. However, the cirrus band recovers cloud accuracy when thermal data are not present. Using the “L8 Biome” data, we determined most cloud algorithms have high commission error over Snow/Ice and sometimes Water biomes. Many algorithms that expressed low errors of omission and commission often produced ambiguous results. Cloud shadow results are relatively similar, but CFMask shows the best overall performance (Table 7). All cloud shadow algorithms yield low commission but high omission error, which is partially dictated by the cloud object from which the shadows are projected. For certain biome or cloud cover types, CFMask-heritage algorithms were not always the best option; we derived a simple metric to provide the most suitable algorithm for each pairing (Table 9).

While some shortcomings of an algorithm may be improved by iteratively refining one or more of its spectral tests, maintaining a global, unbiased scope of each algorithm is one of the key components of the USGS's mission to provide quality assurance information for Landsat data products. The known issues of the TIRS instrument aboard Landsat 8 (Montanaro et al., 2014) and the need to support Sentinel-2 data products necessitate algorithms both with and without thermal data to operationally generate cloud masks. Therefore, the CFMask series of algorithms are judged superior to all other cloud assessment algorithms for Landsat TM, ETM+, and OLI/TIRS products. The CFMask thermal algorithm is recommended for normal operational use, and the non-thermal variant is recommended when thermal data are not available.

8. Acknowledgments

Work performed under USGS contract G15PC00012. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank Joshua

Picotte, MacKenzie Friedrichs, Dr. Christopher Barnes, and our anonymous reviewers for their helpful comments to improve the manuscript.

References

- Arvidson, T., Goward, S., Gasch, J., Williams, D., 2006. Landsat-7 long-term acquisition plan. *Photogramm. Eng. Remote Sens.* 72 (10):1137–1146. <http://dx.doi.org/10.14358/PERS.72.10.1137>.
- Bishop, C.M., 1995. Training with noise is equivalent to Tikhonov regularization. *Neural Comput.* 7 (1):108–116. <http://dx.doi.org/10.1162/neco.1995.7.1.108>.
- Choi, H., Bindschadler, R., 2004. Cloud detection in Landsat imagery of ice sheets using shadow matching technique and automatic normalized difference snow index threshold value decision. *Remote Sens. Environ.* 91 (2):237–242. <http://dx.doi.org/10.1016/j.rse.2004.03.007>.
- Drusch, M., del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120:25–36. <http://dx.doi.org/10.1016/j.rse.2011.11.026>.
- Frantz, D., Röder, A., Udelhoven, T., Schmidt, M., 2015. Enhancing the detectability of clouds and their shadows in multitemporal dryland Landsat imagery: extending Fmask. *IEEE Geosci. Remote Sens. Lett.* 12 (6):1242–1246. <http://dx.doi.org/10.1109/LGRS.2015.2390673>.
- Goward, S.N., Masek, J.G., Williams, D.L., Irons, J.R., Thompson, R., 2001. The Landsat 7 mission: terrestrial research and applications for the 21st century. *Remote Sens. Environ.* 78 (1):3–12. [http://dx.doi.org/10.1016/S0034-4257\(01\)00262-0](http://dx.doi.org/10.1016/S0034-4257(01)00262-0).
- Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENUS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* 114 (8):1747–1755. <http://dx.doi.org/10.1016/j.rse.2010.03.002>.
- Hall, D.K., Foster, J.L., Verbyla, D.L., Klein, A.G., Benson, C.S., 1998. Assessment of snow-cover mapping accuracy in a variety of vegetation-cover densities in central Alaska. *Remote Sens. Environ.* 66 (2):129–137. [http://dx.doi.org/10.1016/S0034-4257\(98\)00051-0](http://dx.doi.org/10.1016/S0034-4257(98)00051-0).
- Helmer, E.H., Lefsky, M.A., Roberts, D.A., 2009. Biomass accumulation rates of Amazonian secondary forest and biomass of old-growth forests from Landsat time series and the Geoscience Laser Altimeter System. *J. Appl. Remote Sens.* 3 (1):033505. <http://dx.doi.org/10.1117/1.3082116>.
- Helmer, E.H., Ruzicki, T.S., Benner, J., Voggeser, S.M., Scobie, B.P., Park, C., Fanning, D.W., Rammarine, S., 2012. Detailed maps of tropical forest types are within reach: forest tree communities for Trinidad and Tobago mapped with multiseason Landsat and multiseason fine-resolution imagery. *For. Ecol. Manag.* 279:147–166. <http://dx.doi.org/10.1016/j.foreco.2012.05.016>.
- Hollingsworth, B.V., Chen, L., Reichenbach, S.E., Irish, R.R., 1996. Automated cloud cover assessment for Landsat TM images. In: *SPIE's 1996 International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics and Photonics:170–179. <http://dx.doi.org/10.1117/12.258064>.
- Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., Vogelmann, J.E., 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sens. Environ.* 114 (1):183–198. <http://dx.doi.org/10.1016/j.rse.2009.08.017>.
- Hughes, M.J., Hayes, D.J., 2014. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sens.* 6 (6):4907–4926. <http://dx.doi.org/10.3390/rs6064907>.
- Irish, R.R., 2000. Landsat 7 automatic cloud cover assessment. In: *AeroSense 2000*. International Society for Optics and Photonics:348–355. <http://dx.doi.org/10.1117/12.410358>.
- Irish, R.R., Barker, J.L., Goward, S.N., Arvidson, T., 2006. Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogramm. Eng. Remote Sens.* 72 (10):1179–1188. <http://dx.doi.org/10.14358/PERS.72.10.1179>.
- Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: the Landsat data continuity mission. *Remote Sens. Environ.* 122:11–21. <http://dx.doi.org/10.1016/j.rse.2011.08.026>.
- Jin, S., Homer, C., Yang, L., Xian, G., Fry, J., Danielson, P., Townsend, P.A., 2013. Automated cloud and shadow detection and filling using two-date Landsat imagery in the USA. *Int. J. Remote Sens.* 34 (5):1540–1560. <http://dx.doi.org/10.1080/01431161.2012.720045>.
- Jones, J.W., 2015. Efficient wetland surface water detection and monitoring via Landsat: comparison with in situ data from the Everglades Depth Estimation Network. *Remote Sens.* 7 (9):12503–12538. <http://dx.doi.org/10.3390/rs70912503>.
- Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sens. Environ.* 112 (3):1196–1211. <http://dx.doi.org/10.1016/j.rse.2007.08.011>.
- Kennedy, R.E., Cohen, W.B., Schroeder, T.A., 2007. Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sens. Environ.* 110 (3):370–386. <http://dx.doi.org/10.1016/j.rse.2007.03.010>.
- Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sens. Environ.* 114 (12):2897–2910. <http://dx.doi.org/10.1016/j.rse.2010.07.008>.
- Lee, J., Weger, R.C., Sengupta, S.K., Welch, R.M., 1990. A neural network approach to cloud classification. *Geoscience and Remote Sensing, IEEE Transactions* 28 (5):846–855. <http://dx.doi.org/10.1109/36.58972>.
- Lewis, A., Lymburner, L., Purss, M.B., Brooke, B., Evans, B., Ip, A., Oliver, S., 2016. Rapid, high-resolution detection of environmental change over continental scales from satellite data—the Earth Observation Data Cube. *International Journal of Digital Earth* 9 (1):106–111. <http://dx.doi.org/10.1080/17538947.2015.1111952>.
- Loveland, T., Reed, B., Brown, J., Ohlen, D., Zhu, Z., Yang, L., Merchant, J., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* 21 (6–7):1303–1330. <http://dx.doi.org/10.1080/014311600210191>.
- Montanaro, M., Gerace, A., Lunsford, A., Reuter, D., 2014. Stray light artifacts in imagery from the Landsat 8 Thermal Infrared Sensor. *Remote Sens.* 6 (11):10435–10456. <http://dx.doi.org/10.3390/rs6110435>.
- Morfit, R., Barsi, J., Levy, R., Markham, B., Micijevic, E., Ong, L., Scaramuzza, P., Vanderwerff, K., 2015. Landsat-8 Operational Land Imager (OLI) radiometric performance on-orbit. *Remote Sens.* 7 (2):2208–2237. <http://dx.doi.org/10.3390/rs70202208>.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., et al., 2001. Terrestrial ecoregions of the world: a new map of life on Earth a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* 51 (11):933–938. [http://dx.doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](http://dx.doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2).
- Oreopoulos, L., Wilson, M.J., Várnai, T., 2011. Implementation on Landsat data of a simple cloud-mask algorithm developed for MODIS land bands. *Geoscience and Remote Sensing Letters, IEEE* 8 (4):597–601. <http://dx.doi.org/10.1109/LGRS.2010.2095409>.
- Reeves, M.C., Kost, J.R., Ryan, K.C., 2006. *Fuels Products of the LANDFIRE Project*. USDA Forest Service Proceedings RMRS-P-41.
- Rossow, W.B., Schiffer, R.A., 1999. Advances in understanding clouds from ISCCP. *Bull. Am. Meteorol. Soc.* 80 (11):2261–2287. [http://dx.doi.org/10.1175/1520-0477\(1999\)080<2261:AIUCFI>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2).
- Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sens. Environ.* 112 (6):3112–3130. <http://dx.doi.org/10.1016/j.rse.2008.03.009>.
- Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, 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 (1):35–49. <http://dx.doi.org/10.1016/j.rse.2009.08.011>.
- Scaramuzza, P.L., Bouchard, M.A., Dwyer, J.L., 2012. Development of the Landsat data continuity mission cloud-cover assessment algorithms. *Geoscience and Remote Sensing, IEEE Transactions* 50 (4):1140–1154. <http://dx.doi.org/10.1109/TGRS.2011.2164087>.
- Schroeder, W., Oliva, P., Giglio, L., Quayle, B., Lorenz, E., Morelli, F., 2015. Active fire detection using Landsat-8/OLI data. *Remote Sens. Environ.* 185:210–220. <http://dx.doi.org/10.1016/j.rse.2015.08.032>.
- Selkowitz, D.J., Forster, R.R., 2015. An automated approach for mapping persistent ice and snow cover over high latitude regions. *Remote Sens.* 8 (1):16. <http://dx.doi.org/10.3390/rs8010016>.
- Sheng, Y., Song, C., Wang, J., Lyons, E.A., Knox, B.R., Cox, J.S., Gao, F., 2016. Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. *Remote Sens. Environ.* 185:129–141. <http://dx.doi.org/10.1016/j.rse.2015.12.041>.
- U.S. Geological Survey, 2016a. L7 Irish Cloud Validation Masks. U.S. Geological Survey data release. [10.5066/F7XD0ZWC](http://dx.doi.org/10.5066/F7XD0ZWC) (accessed 28 February 2017).
- U.S. Geological Survey, 2016b. L8 SPARCS Cloud Validation Masks. U.S. Geological Survey data release. [10.5066/F7FB5146](http://dx.doi.org/10.5066/F7FB5146) (accessed 28 February 2017).
- U.S. Geological Survey, 2016c. L8 Biome Cloud Validation Masks. U.S. Geological Survey, data release <http://dx.doi.org/10.5066/F7251GDH> (accessed 28 February 2017).
- U.S. Geological Survey, 2016d. Landsat 8 OLI and TIRS Calibration Notices. U.S. Geological Survey Landsat Missions web page, at: <https://landsat.usgs.gov/landsat-8-18-operational-land-imager-oli-and-thermal-infrared-sensor-tirs> (accessed 28 February 2017).
- U.S. Geological Survey, 2016e. Landsat 8 (L8) Data Users Handbook. U.S. Geological Survey Landsat Missions web page at: <https://landsat.usgs.gov/landsat-8-18-data-users-handbook> (accessed 28 February 2017).
- Vermote, E., Saleous, N., 2007. *LEDAPS Surface Reflectance Product Description*. University of Maryland Department of Geography, College Park.
- 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. <http://dx.doi.org/10.1016/j.rse.2016.04.008>.
- Wang, B., Ono, A., Muramatsu, K., 1999. Automated detection and removal of clouds and their shadows from Landsat TM images. *IEICE Trans. Inf. Syst.* E82-D (2), 453–460.
- Yang, L., Huang, C., Homer, C.G., Wylie, B.K., Coan, M.J., 2003. An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Can. J. Remote Sens.* 29 (2):230–240. <http://dx.doi.org/10.5589/m02-098>.
- Zhang, Y., Rossow, W.B., Lacis, A.A., Oinas, V., Mishchenko, M.I., 2004. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: refinements of the radiative transfer model and the input data. *J. Geophys. Res. Atmos.* 109 (D19). <http://dx.doi.org/10.1029/2003JD004457>.
- Zhu, Z., 2013. *Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data*. Boston University (Ph.D. thesis).
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118:83–94. <http://dx.doi.org/10.1016/j.rse.2011.10.028>.
- Zhu, Z., Woodcock, C.E., 2014a. 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. <http://dx.doi.org/10.1016/j.rse.2014.06.012>.
- Zhu, Z., Woodcock, C.E., 2014b. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* 144:152–171. <http://dx.doi.org/10.1016/j.rse.2014.01.011>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015. 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. <http://dx.doi.org/10.1016/j.rse.2014.12.014>.