Remote Sensing of Environment 159 (2015) 269–277

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images

Zhe Zhu *, Shixiong Wang, Curtis E. Woodcock

Center for Remote Sensing, Department of Earth and Environment, Boston University, 685 Commonwealth Avenue, Boston, MA, 02215, USA

ARTICLE INFO

Article history:
Received 1 September 2014
Received in revised form 4 December 2014
Accepted 14 December 2014
Available online 9 January 2015

Keywords:
Fmask
Cloud
Cloud shadow
Snow
Landsat
Sentinel
Cirrus

ABSTRACT

Identification of clouds, cloud shadows and snow in optical images is often a necessary step toward their use. Recently a new program (named Fmask) designed to accomplish these tasks was introduced for use with images from Landsats 4–7 (Zhu & Woodcock, 2012). In this paper, there are the following: (1) improvements in the Fmask algorithm for Landsats 4–7; (2) a new version for use with Landsat 8 that takes advantage of the new cirrus band; and (3) a prototype algorithm for Sentinel 2 images. Though Sentinel 2 images do not have a thermal band to help with cloud detection, the new cirrus band is found to be useful for detecting clouds, especially for thin cirrus clouds. By adding a new cirrus cloud probability and removing the steps that use the thermal band, the Sentinel 2 scenario achieves significantly better results than the Landsats 4–7 scenario for all 7 images tested. For Landsat 8, almost all the Fmask algorithm components are the same as for Landsats 4–7, except a new cirrus cloud probability is calculated using the new cirrus band, which improves detection of thin cirrus clouds. Landsat 8 results are better than the Sentinel 2 scenario, with 6 out of 7 test images showing higher accuracies.

1. Introduction

Clouds, cloud shadows, and snow significantly influence the spectral bands of optical sensors (Dozier, 1989; Irish, Barker, Goward, & Arvidson, 2006; Zhu & Woodcock, 2012). Their presence can cause serious problems for a variety of remote sensing activities, including: image compositing (Roy et al., 2010); correction for atmosphere effects (Vermote, El Saleous, & Justice, 2002); calculation of vegetation indices (Huete et al., 2002); classification of land cover (Zhang, Guindon, & Cihlar, 2002); and most importantly in change detection (Zhu & Woodcock, 2014a). Therefore, it is important to detect clouds, cloud shadows, and snow in satellite images, and screen them accurately before any kind of remote sensing activity is performed.

However, accurate cloud, cloud shadow, and snow detection for satellite images is quite challenging. It is difficult to separate clouds from other clear-sky observations based on the reflectance of spectral bands, as there are many kinds of clouds and each kind may have different spectral characteristics (Platnick et al., 2003). In particular, optically thin clouds are difficult to detect as they may have spectral signatures similar to the Earth surface underneath (Zhang et al., 2002; Zhu & Woodcock, 2012). Cloud shadow detection is also difficult, as there are many kinds of dark surfaces that have spectral signatures similar to cloud shadows, such as topographic shadows, wetlands, and water. Snow detection is somewhat easier due to its distinctive spectral signature. The widely used Normalized Difference Snow Index (NDSI) is helpful in snow detection, as snow pixels are usually high in NDSI values, while pixels without snow tend to have much lower values (Dozier, 1989).

The 40 + years of Landsat data provide one of the most valuable datasets available for mapping and monitoring the Earth surface (Coppin & Bauer, 1994; Kennedy et al., 2014). The availability of user ready image products has proven to be critical for applications uptake of Landsat images. Opening of the archive (Woodcock et al., 2008) combined with analysis ready products has led to a massive increase in usage. Automated cloud screening has made it possible for user to spend less time on data preparation and more time on value-added analyses, especially with regards to image compositing (Griffiths, van der Linden, Kuemmerle, & Hostert, 2013; Roy et al., 2010; White et al., 2014) and time series application (Huang, Goward, et al., 2010; Kennedy, Yang, & Cohen, 2010; Zhu & Woodcock, 2014b). Lots of cloud detection algorithms have been developed for the Thematic Mapper (TM) carried by Landsat 4 and Landsat 5, and the Enhanced Thematic Mapper Plus (ETM+) carried by Landsat 7 (Huang, Thomas, et al., 2010; Irish et al., 2006; Roy et al., 2010; Zhu & Woodcock, 2012). As Landsat TM and ETM+ have very similar spectral bands (Table 1), most cloud detection algorithms tend to treat them as the same sensor (hereafter Landsat TM and ETM+ will be called "Landsats 4–7" in this paper for simplicity). Besides cloud detection, some of these algorithms can also detect cloud shadow and snow in Landsats 4–7 images. For example, Huang, Thomas, et al., 2010 can also provide cloud shadow mask and Oreopoulos et al. (2011) can identify snow. The

* Corresponding author. Tel.: +1 617 233 6031.
E-mail address: zhuzhe@bu.edu (Z. Zhu).

http://dx.doi.org/10.1016/j.rse.2014.12.014
0034-4257/

Table 1
Landsat TM, ETM+, OLI/TIRS, and Sentinel 2 spectral bands. The spectral bands used in the new Fmask algorithm are highlighted in bold letters.

TM bands (μm)	*ETM + bands (μm)*	*OLI/TIRS bands (μm)*	*Sentinel 2 bands (μm)*
Band 1 (0.45–0.52)	Band 1 (0.45–0.515)	Band 1 (0.43–0.45)	Band 1 (0.433–0.453)
Band 2 (0.52–0.60)	Band 2 (0.525–0.605)	Band 2 (0.45–0.51)	Band 2 (0.458–0.523)
Band 3 (0.63–0.69)	Band 3 (0.63–0.69)	Band 3 (0.53–0.59)	Band 3 (0.543–0.578)
Band 4 (0.76–0.90)	Band 4 (0.75–0.90)	Band 4 (0.64–0.67)	Band 4 (0.650–0.680)
Band 5 (1.55–1.75)	Band 5 (1.55–1.75)	Band 5 (0.85–0.88)	Band 5 (0.698–0.713)
Band 6 (10.40–12.50)	Band 6 (10.40–12.50)	Band 6 (1.57–1.65)	Band 6 (0.733–0.748)
Band 7 (2.08–2.35)	Band 7 (2.09–2.35)	Band 7 (2.11–2.29)	Band 7 (0.765–0.785)
	Band 8 (0.52–0.90)	Band 8 (0.50–0.68)	Band 8 (0.785–0.900)
		Band 9 (1.36–1.38)	Band 8a (0.855–0.875)
		Band 10 (10.60–11.19)	Band 9 (0.930–0.950)
		Band 11 (11.50–12.51)	Band 10 (1.365–1.385)
			Band 11 (1.565–1.655)
			Band 12 (2.100–2.280)

newly developed Fmask algorithm is capable of detecting cloud, cloud shadow, and snow at the same time (Zhu & Woodcock, 2012).

Recently, Landsat 8 has been launched with two sensors: the Operational Landsat Imager (OLI) and the Thermal Infrared Sensor (TIRS) (Irons, Dwyer, & Barsi, 2012; Roy et al., 2014). This new satellite has all the spectral bands of Landsat TM and ETM + and it also has some new bands (Table 1). The new Short Wave Infrared (SWIR) band (Band 9) (1.36–1.38 μm) is especially helpful for detecting high altitude clouds, such as cirrus clouds (hereafter this new SWIR band will be called "cirrus band" in this paper for simplicity). Because of the strong water absorption at this specific band, clouds at high altitudes will have small above-cloud two-way water vapor path lengths, and this will make the cirrus band relatively bright (Gao, Goetz, & Wiscombe, 1993). On the other hand, for low altitude clouds and Earth surfaces in most environments, the two-way water vapor path length is usually quite large, and this will make the cirrus band generally dark. However, bright surfaces in dry environments or at high altitudes (e.g., polar regions or high mountains) can also have relatively high reflectance in the cirrus band and can be misidentified as clouds (Wilson & Oreopoulos, 2013). To reduce this kind of commission error, a relatively large threshold of the cirrus band is required, but this will also exclude some of the thin cirrus clouds. Wilson and Oreopoulos (2013) proposed a new cloud detection algorithm for Landsat 8 images. This algorithm is able to provide reasonably good cloud and snow detection for Landsat 8 data based on many spectral tests that are originally developed for cloud detection in Moderate Resolution Imaging Spectroradiometer (MODIS) images by Luo, Trishchenko, & Khlopenkov, 2008. In this study, a threshold of 0.0113 in Top Of Atmosphere (TOA) reflectance of the cirrus band is suggested for detecting cirrus clouds. This threshold is able to find most of the cirrus clouds, but will also persistently misclassify bright surfaces in dry environments or at high altitudes with cirrus band TOA reflectances higher than this threshold. In addition, the current Level 1 Landsat 8 images include a Quality Assessment (QA) band that also provides cirrus confidence information. Due to the time constraints for releasing the Landsat 8 products, the current cirrus detection in the QA band is based on a simple assumption that if cirrus band reflectance is larger than 0.02, this pixel is labeled as cirrus cloud. This threshold may miss many cirrus clouds that have TOA reflectance less than 0.02 and also may suffer from false detection for bright surfaces in dry environments or at high elevations with TOA reflectance higher than 0.02. Moreover, both of the two algorithms do not have the ability to provide cloud shadow masks for Landsat 8 images.

In the near future, two Sentinel 2 satellites will be launched, with Sentinel 2A planned in 2014 and Sentinel 2B to be launched 18 months later (Berger, Moreno, Johannessen, Levelt, & Hanssen, 2012). The sensors on Sentinel 2 have 13 spectral bands from the visible to the SWIR,

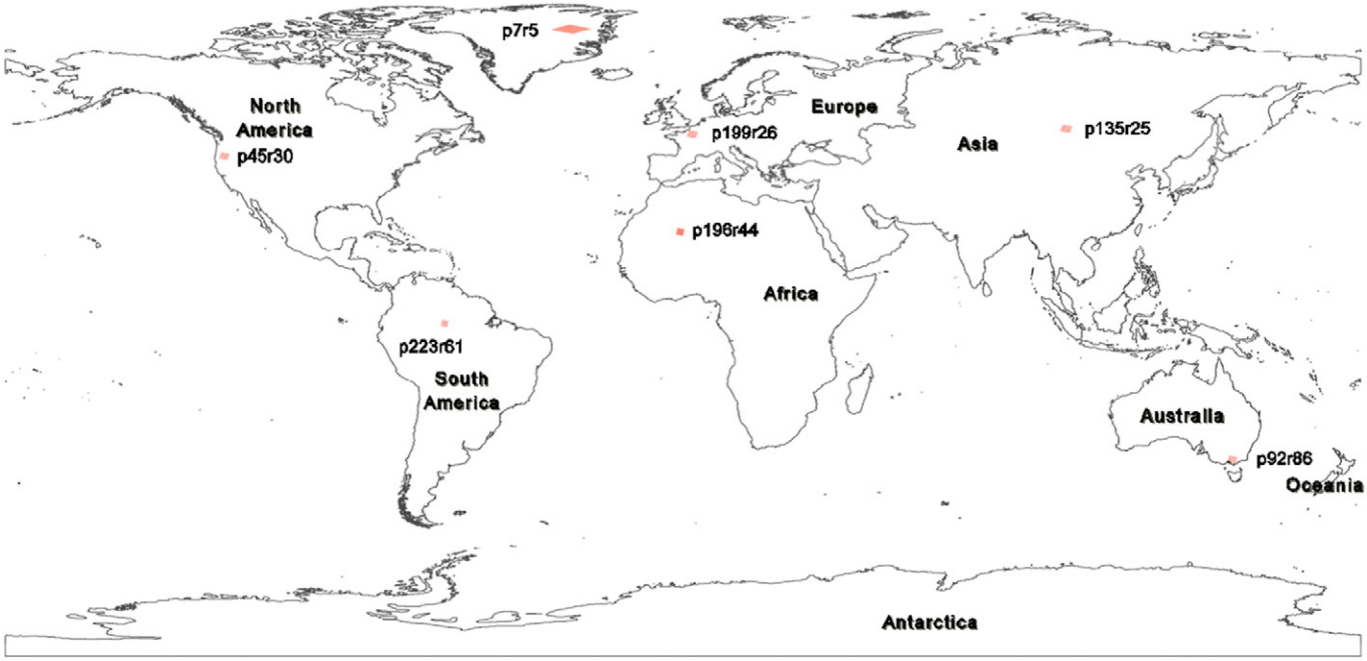

Fig. 1. To represent the wide variety of environments and different cloud types, 7 Landsat images are selected. The red color shows the coverage of each Landsat scene.

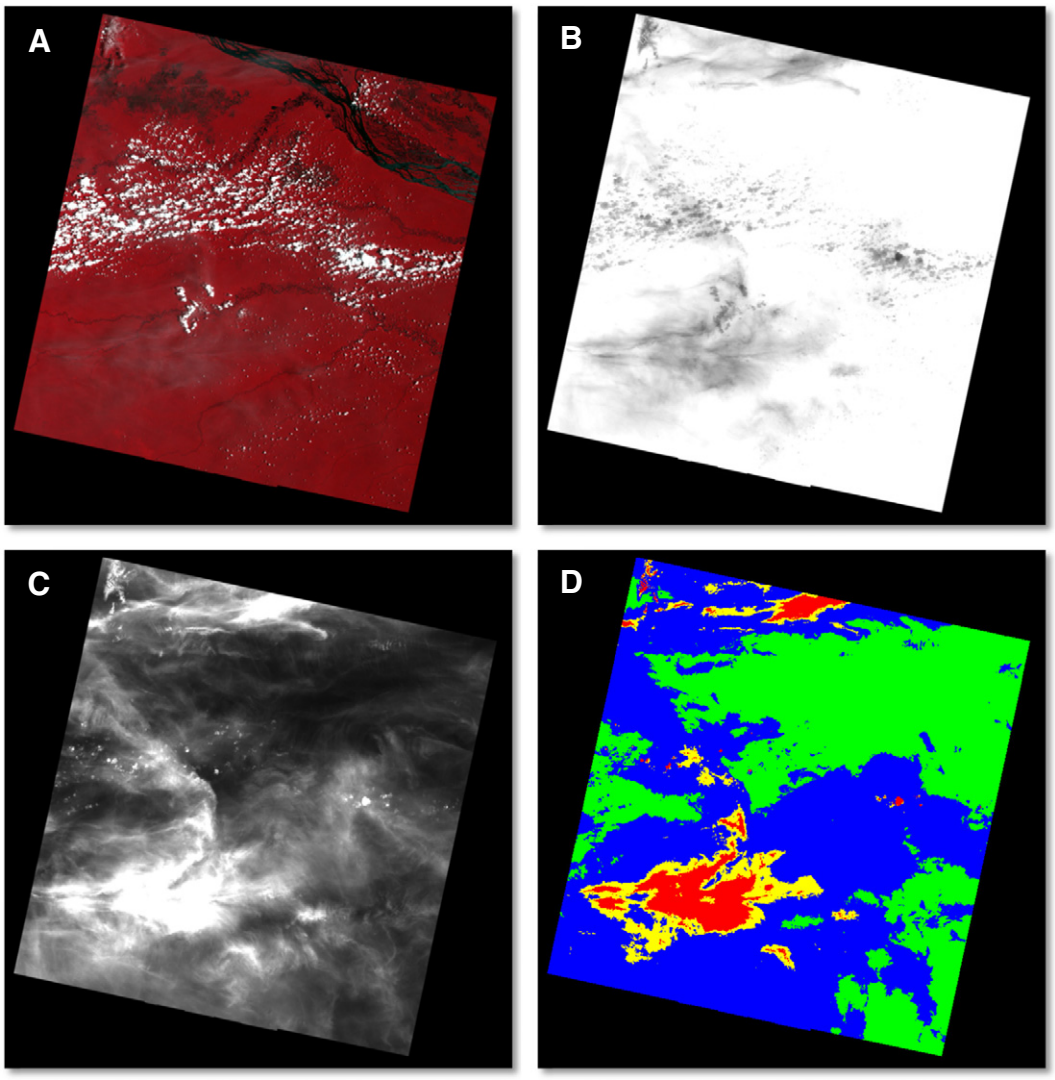

Fig. 2. Landsat 8 image located at Path 233/Row 61. Fig. 2A shows the false color composite image (NIR, red, and green bands). Fig. 2B shows the thermal band (Band 10) and Fig. 2C shows the cirrus band (Band 9). Fig. 2D shows a density slice of the cirrus band TOA reflectance (0 < green < 0.01; 0.01 < blue < 0.03; 0.03 < yellow < 0.04; 0.04 < red < 1). Note that for the thin cirrus clouds with TOA reflectance between 0.01 and 0.03 (blue color in Fig. 2D) are almost invisible in the false color composite image (Fig. 2A).

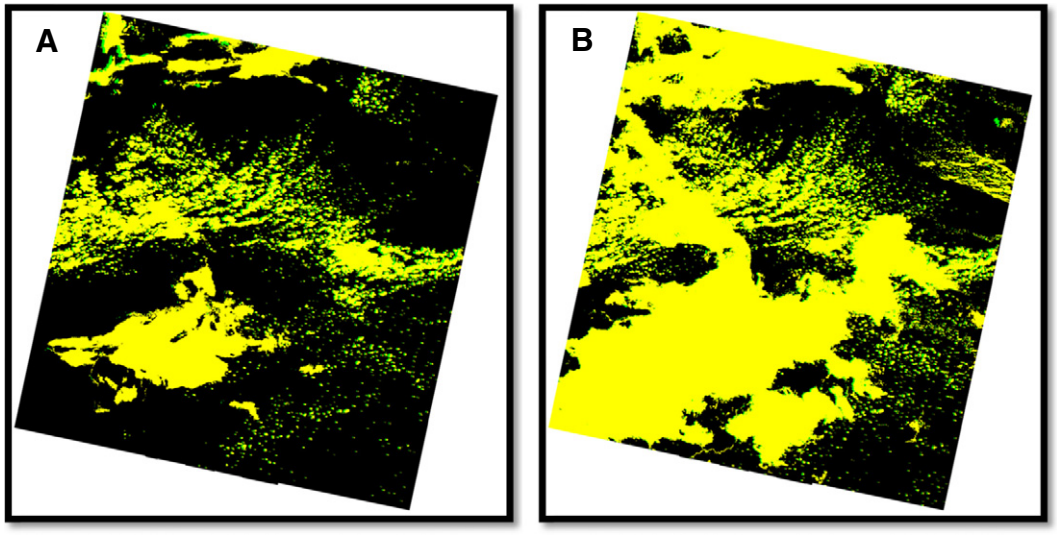

Fig. 3. Fmask results for the Landsats 4–7 (Fig. 3A) and the Sentinel 2 (Fig. 3B) scenarios at Path 233/Row 61 (clouds in yellow, cloud shadows in green, and clear-sky in black).

including 6 that are similar to the Landsat TM and ETM+ bands (Table 1) (Drusch et al., 2012). It is important to design an automated cloud, cloud shadow, and snow detection algorithm for Sentinel 2 before it is launched. Compared to Landsat TM and ETM+, Sentinel 2 has the advantage of a cirrus band, but does not have a thermal band, which is also an important band for cloud detection. Most of the cloud detection algorithms are heavily dependent on the thermal band, as cloud pixels are much colder than clear-sky pixels (Huang, Thomas, et al., 2010; Irish et al., 2006; Zhu & Woodcock, 2012). Moreover, most of the cloud shadow detection algorithms also need the thermal band to estimate cloud height (Huang, Thomas, et al., 2010; Khlopenkov & Trishchenko, 2007; Masek et al., 2006; Zhu & Woodcock, 2012). Therefore, it would be quite challenge to design a good cloud and cloud shadow detection algorithm for Sentinel 2 images.

The Fmask algorithm was originally developed for masking cloud, cloud shadow, and snow for Landsats 4–7. By using an object-based cloud and cloud shadow matching algorithm, it is capable of providing cloud, cloud shadow, and snow masks for each individual image. This algorithm has been widely used and has been integrated into the Landsat surface reflectance Climate Data Record (CDR) provided by U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (Maiersperger et al., 2013; Masek et al., 2006). As Fmask is a relatively new algorithm, we are constantly improving it based on the large amount of feedback from users. This paper first introduces the major improvements made to the Fmask algorithm. Then, the Fmask algorithm is expanded to the Landsat 8 scenario by taking advantage of the cirrus band. Last, a version of Fmask is developed that could be used with Sentinel 2 images based on the bands available on Landsat 8. So for the Sentinel 2 scenario, optical bands and the cirrus band are used. Most of the Fmask algorithm designed for Sentinel 2 and Landsat 8 remains the same as the original Landsats 4–7 version. Changes occur only where the spectral bands of the two new sensors differ from the Landsats 4–7 scenario. There are three major questions considered in this paper: 1) How to better use the new cirrus band on Landsat 8 and Sentinel 2 for better cloud detection?; 2) Can we still detect clouds and their shadows accurately when there is no thermal band available (the Sentinel 2 scenario)?; and 3) Which scenario (Landsats 4–7, Sentinel 2, or Landsat 8) has the best results in cloud detection?

2. Algorithm improvements and expansions

2.1. Algorithm improvements

There are many improvements made for this version of the Fmask algorithm (3.2. version) compared to the original Fmask algorithm (1.6. version) published in 2012. This new version is able to achieve better results and at the same time maintaining similar computation efficiency as the previous version. The improvements are briefly introduced in the following six aspects.

Cloud Detection. This improvement reduces commission error for clear-sky pixels that have a high cloud probability (calculated by Fmask). In the original Fmask algorithm, a pixel is labeled cloudy directly if its cloud probability is higher than 99%, without taking any spectral tests (Zhu & Woodcock, 2012). However, this criterion tends to overestimate clouds, as there are clear-sky pixels that are bright, white, and cold and as a result have high cloud probability values (e.g., high altitude snow). In order to prevent this kind of commission error, this criterion is removed from the new Fmask algorithm.

Cloud Detection over Water. The original Fmask algorithm uses a fixed threshold (cloud probability of 50%) for cloud detection over water (Zhu & Woodcock, 2012). This threshold works well for most water bodies but may not work well for places where the water pixels are both cold (compared to the land pixels) and bright (e.g., turbid water or water containing high concentration of sediments). In the new Fmask algorithm, the threshold used for cloud detection in water areas is a dynamic threshold derived by adding the upper level (82.5 percentile) of cloud probability of all clear water pixels with a constant of 0.2 (this approach mirrors the dynamic threshold used for detecting clouds in land areas). This statistically derived threshold is able to provide better cloud detection over water areas.

Potential Shadow Detection. Based on the assumption that cloud shadows are usually dark in the Near Infrared (NIR) band, the original Fmask algorithm uses the difference between the NIR band and the flood-fill transformation (Soille, 1999; Soille, Vogt, & Colombo, 2003) of the same NIR band to extract potential cloud shadows (Zhu & Woodcock, 2012). This method works well for most places, but when there are many dark objects in the NIR band, it will label too many pixels as potential cloud shadow, which may cause problems for the step that matches clouds and cloud shadows. As cloud shadows are also dark in the SWIR band, in the new Fmask algorithm we use both the NIR band and one of the SWIR bands (centered at 1.65 μm) to extract potential cloud shadows. Therefore, in this new Fmask algorithm, the flood-fill transformation is used for both the NIR and a SWIR band. The pixel is labeled as "potential cloud shadow" if the flood-fill transformed image minus the original image is larger than 0.02 in both NIR and SWIR bands. This improvement reduces the number of pixels identified as "potential cloud shadow" and produces better cloud shadow detection results.

Cloud Shadow Detection. The Fmask algorithm matches clouds with their shadows based on similarity measurements (Zhu & Woodcock, 2012). This algorithm iterates cloud height from a minimum possible height to a maximum possible height and computes the similarity between cloud and cloud shadow for different cloud heights. In the

Table 2
The percent disagreement for Fmask results between the Landsats 4–7 and the Sentinel 2 scenarios at Path 223/Row 61.

Sentinel 2 / Landsats 4–7	Clear land	Clear water	Cloud shadow	Snow/ice	Cloud
Clear land	**45.13%**	0.00%	2.48%	0.00%	28.35%
Clear water	0.00%	**0.84%**	0.04%	0.00%	0.11%
Cloud shadow	0.14%	0.01%	**1.17%**	0.00%	1.88%
Snow/ice	0.00%	0.00%	0.00%	**0.00%**	0.00%
Cloud	0.00%	0.00%	0.01%	0.00%	**19.84%**

Table 3
Confusion matrix for assessing Fmask accuracies for the Landsats 4–7 scenario at Path 223/Row 61.

Landsats 4–7 scenario					
Ground truth (pixels)					
Class	Clear land	Clear water	Cloud shadow	Cloud	User
Clear land	**6**	4	4	176	3.16%
Clear water	0	**0**	0	0	N/A
Cloud shadow	0	0	**0**	10	0.00%
Cloud	0	0	0	**0**	N/A
Producer	100.00%	0.00%	0.00%	0.00%	Overall = **3.00%**

Table 4
Confusion matrix for assessing Fmask accuracies for the Sentinel 2 scenario at Path 233/Row 61.

Sentinel 2 scenario					
Ground truth (pixels)					
Class	Clear land	Clear water	Cloud shadow	Cloud	User
Clear land	**0**	0	0	0	N/A
Clear water	0	**0**	0	0	N/A
Cloud shadow	2	0	**4**	8	28.57%
Cloud	4	4	0	**178**	95.70%
Producer	0.00%	0.00%	100.00%	95.70%	Overall = **91.00%**

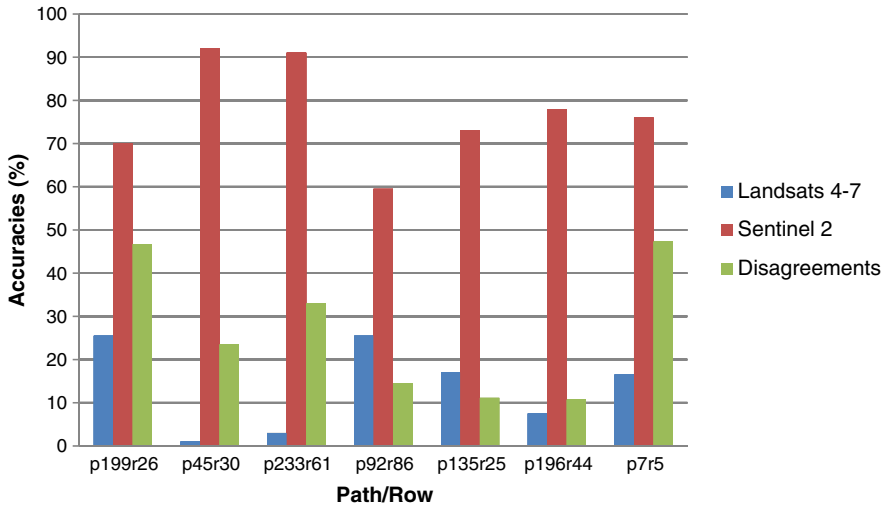

Fig. 4. Fmask results accuracies for all 7 locations between the Landsats 4–7 and the Sentinel 2 scenarios. The blue bars are the overall accuracies for the Landsats 4–7 scenario and the red bars are the overall accuracies for the Sentinel 2 scenario. The green bars show the percent disagreement for Fmask results between the Landsats 4–7 and the Sentinel 2 scenarios.

original Fmask algorithm, the testing of cloud heights continues if the similarity is increasing or does not decrease to 98% of the maximum measured similarity; otherwise, the search for a cloud height stops and the cloud shadow is matched at the maximum similarity. However, sometimes the iteration may stop earlier than it should, as before the similarity gets to the maximum similarity; there may be local maxima that are 2% larger than the similarity measured for neighboring cloud heights. In the improved Fmask algorithm, the cloud and cloud shadow match will not stop unless the similarity value decreases to 95% of the maximum similarity value. This simple change improves matching of clouds and cloud shadows by preventing errors caused by premature stops in the search for an optimum cloud height.

Snow Detection I. The original Fmask algorithm uses a temperature threshold to minimize false detection of snow on warm and bright nonsnow surfaces, in which a threshold of 277 K was chosen (Zhu & Woodcock, 2012). This threshold has been previously applied to the MODIS snow cover algorithm (Hall et al., 2001). Recent results show that some warm snow pixels can also have brightness temperatures higher than 277 K, and in Collection 5 of MODIS snow cover algorithm, a higher temperature threshold of 283 K is used (Hall, 2012). Therefore, in the new Fmask algorithm, if there is thermal band available, the higher temperature threshold of 283 K is chosen for snow detection

Snow Detection II. The pixels surrounding clouds and cloud shadows are masked out in the Fmask algorithm because many of these pixels may still be influenced by the thin edges of clouds and their shadows (Zhu & Woodcock, 2012). Based on many users' feedback of the Fmask algorithm, the pixels surrounding snow can also be partially influenced by snow and they may cause problems for remote sensing activities. Therefore, in the new Fmask algorithm, we dilate the pixels surrounding snow in 8-connected directions for a user-specified distance (can be any positive integer based on the user's preference). The pixels removed from this dilation are included in the snow mask in Fmask results.

2.2. Algorithm expansions

The original Fmask algorithm can only provide cloud, cloud shadow, and snow masks for images from Landsats 4–7. The new Fmask algorithm is also capable of providing cloud, cloud shadow, and snow mask for Landsat 8 and Sentinel 2 images. The cirrus band included on the OLI sensor and in Sentinel 2 provides an opportunity for better detecting thin cirrus clouds. As Sentinel 2 images do not have a thermal band, the new Fmask algorithm designed for Sentinel 2 images is different from the algorithm designed for Landsat 8 images. As Landsat 8 has two thermal bands, we will use the one with shorter wavelength (Band

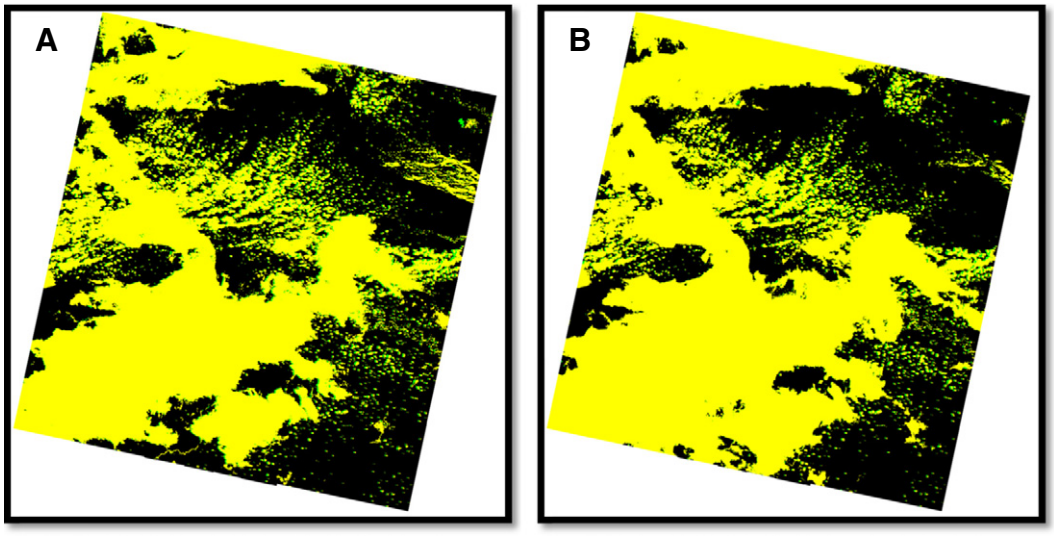

Fig. 5. Fmask results for the Sentinel 2 (Fig. 5A) and the Landsat 8 (Fig. 5B) scenarios at Path 233/Row 61 (clouds in yellow, cloud shadows in green, and clear-sky in black).

10) due to less atmosphere influences and the recent report of "ghosting" issues found in the other thermal band (Schott et al., 2014).

2.2.1. Algorithm for Landsat 8

For Landsat 8 images, most of the Fmask algorithm is exactly the same as it is for Landsats 4–7 images, except for places where the cirrus band is applied. The cirrus band is used in the first pass of generating the potential cloud layer, in which a cirrus cloud test is applied for better identification of Potential Cloud Pixels (PCP). This cirrus cloud test is quite simple: if the TOA reflectance of the cirrus band is larger than 0.01, this pixel is labeled as a PCP. Note that the threshold used here is much lower than the threshold used by the MODIS cloud detection algorithm (Ackerman et al., 2010). In the MODIS cloud products, land pixels with cirrus band TOA reflectance less than 0.03 are considered clear. However, there are many cirrus clouds with TOA reflectance in the cirrus band less than 0.03. One reason for the MODIS cloud algorithm to use this high threshold is the cross-talking problem observed in the cirrus band, which is caused by energy leakage from another SWIR band centered 11 μm (Guenther, Xiong, Salomonson, Barnes, & Young, 2002). This cross-talk forces the MODIS cloud detection algorithm to use a higher threshold to reduce false positive errors in detecting cirrus clouds. For Landsat 8 images, there is no cross-talk problem observed and a smaller threshold is possible for improved detection of thin cirrus clouds with TOA reflectance less than 0.03. The threshold of 0.01 is derived by a recent study that finds this threshold corresponds to the lowest 1% of simulated cloudy-sky reflectance (Wilson & Oreopoulos, 2013).

The cirrus band is also applied to calculate the cloud probability. As the cirrus band will only have high TOA reflectance for clouds at high altitudes (mostly cirrus clouds), the new Fmask algorithm uses the value of the TOA reflectance of the cirrus band to calculate a cirrus cloud probability and adds this new probability to the original cloud probability calculated by the original Fmask algorithm. By combining the old cloud probability that identifies both thin and thick clouds and the new cirrus band probability that identifies mostly cirrus clouds, this new cloud probability is capable of providing better cloud probabilities, especially for places with lots of thin cirrus clouds (see Section 3 for details).

The cirrus cloud probability is calculated based on the magnitude of cirrus band TOA reflectance. The cirrus cloud probability increases linearly with the TOA reflectance of the cirrus band and gets to 1 when the TOA reflectance equals to 0.04 (Ackerman et al., 2010). If the cirrus band TOA reflectance is larger than 0.04, the cirrus cloud probability will continue to increase linearly with the cirrus band TOA reflectance (Eq. 1).

$$\textit{Cirrus cloud probability} = \textit{cirrus band}/0.04 \quad (1)$$

With the help of the cirrus band, this new version of the Fmask algorithm is able to detect both thick and thin clouds; including some extremely thin cirrus clouds in Landsat 8 images (see Section 3 for details). Based on the cloud and cloud shadow approach, this new Fmask algorithm is also capable of detecting cloud shadows casted by some extremely thin cirrus clouds.

2.2.2. Algorithm for Sentinel 2 images

Except for the thermal band, Sentinel 2 images have all the spectral bands of Landsats 4–8. The lack of a thermal band precludes the use of several of the tests that are part of Fmask for the Landsat satellites. However, with the extra cirrus band, the same cirrus test used for Landsat 8 can be employed for Sentinel 2 data for the first pass in generating the potential cloud layer. Therefore, the new cloud probability for the Sentinel 2 scenario is calculated as in Eq. (2) or Eq. (3). The brightness probability and the variability probability are calculated the same as the original Fmask algorithm and the cirrus cloud probability is calculated the same as the new Fmask algorithm designed for Landsat 8 images (Eq. 1). If it is a water pixel, the new cloud probability is a combination of brightness probability and cirrus cloud probability (Eq. 2). If it is a land pixel, the new cloud probability is a combination of whiteness probability and cirrus cloud probability (Eq. 3).

$$\textit{New cloud probability}\ (\textit{water}) = \textit{brightness probability} + \textit{cirrus cloud probability} \quad (2)$$

$$\textit{New cloud probability}\ (\textit{land}) = \textit{variability probability} + \textit{cirrus cloud probability} \quad (3)$$

For cloud shadow detection, the thermal band is used in two places in the original Fmask algorithm: to predict cloud height range and to build a three-dimensional cloud object. In the new Fmask algorithm designed for Sentinel 2, the predicted cloud height range is fixed between 200 m and 1,200 m and each cloud object is treated as a flat plate instead of three-dimensional cloud object for predicting cloud shadow locations. For snow detection, the thermal band originally used for a temperature screen to reduce false positive errors is no longer applied to Sentinel 2 images.

Though there is no thermal band available in Sentinel 2 images, with the extra cirrus band, this new version of the Fmask algorithm can still capture many kinds of clouds, including clouds that are extremely thin and match shadows for them (see Section 3 for details). For cloud shadow detection, if a cloud is thick, the matched shadows may have a different shape than the actual cloud shadow. For snow detection, the lack of temperature screen may lead to more commission errors, but will also reduce omission errors for some warm snow pixels.

Table 5
The percent disagreement for Fmask results between the Sentinel 2 and the Landsat 8 scenarios at Path 233/Row 61.

Landsat 8 / Sentinel 2	Clear land	Clear water	Cloud shadow	Snow/ice	Cloud
Clear land	**39.15%**	0.00%	0.59%	0.00%	5.54%
Clear water	0.00%	**0.83%**	0.00%	0.00%	0.01%
Cloud shadow	0.77%	0.03%	**1.80%**	0.00%	1.10%
Snow/ice	0.00%	0.00%	0.00%	**0.00%**	0.00%
Cloud	1.11%	0.08%	0.20%	0.00%	**48.79%**

3. Fmask results and accuracy assessment

As the cloud and cloud shadow detection accuracies for Landsats 4–7 images have already been assessed against a total of 142 reference images (Zhu & Woodcock, 2012), the main focus here is to compare the accuracies computed from the three different scenarios: Landsats 4–7, Landsat 8, and Sentinel 2. A global sample with specific characteristics of clouds and environments are necessary for this purpose. Therefore, we selected 7 Landsat 8 images from different parts of the world (Fig. 1). These images include a variety of environments (including ice caps, snow, desert, urban, forest, grass, and mountains) and different kinds of clouds (including some extremely thin cirrus clouds). As Landsat 8 images have all the spectral bands as Landsats 4–7 images

Table 6
Confusion matrix for assessing Fmask accuracies for the Sentinel 2 scenario at Path 233/Row 61.

Sentinel 2 scenario					
Ground truth (pixels)					
Class	Clear land	Clear water	Cloud shadow	Cloud	User
Clear land	**3**	0	1	169	1.73%
Clear water	0	**0**	0	0	N/A
Cloud shadow	0	0	**1**	22	4.35%
Cloud	0	1	1	**2**	50.00%
Producer	100.00%	0.00%	33.33%	1.04%	Overall = **3.00%**

Table 7
Confusion matrix for assessing Fmask accuracies for the Landsat 8 scenario at Path 233/Row 61.

Landsat 8 scenario					
Ground truth (pixels)					
Class	Clear land	Clear water	Cloud shadow	Cloud	User
Clear land	**0**	0	0	0	N/A
Clear water	0	**1**	0	0	100.00%
Cloud shadow	2	0	**2**	18	9.09%
Cloud	1	0	1	**175**	98.87%
Producer	0.00%	100.00%	66.66%	90.67%	Overall = **89.00%**

and all the spectral bands used by this version of Fmask for Sentinel 2 images (Table 1), we use Landsat 8 images to simulate Landsats 4–7 and Sentinel 2 images as the inputs for the new Fmask algorithm. Though there are differences in band width, quantization, Signal Noise Ratio (SNR), and resolution (with Sentinel 2) between different sensors, which may influence the results, we assume those influences are trivial compared to the impacts of include or exclude certain important spectral bands (the thermal band or the cirrus band). Due to the large number of tables and figures for the 7 locations, we will only show Fmask results for one of the images located in South America (Path 223/Row 61) in the paper. The tables and figures for the other 6 locations are in the supplementary materials (Tables S1-S36 & Figs. S1-S42). Also, summary results for all the tables and figures are presented below.

The South America image is located in the Amazon Basin. Most of the image is covered by primary forests, and there are several rivers in the image. There are a variety of clouds in this image. Some of them are quite easy to detect as they are both bright and white. Some of the clouds are almost invisible in the false color composite (Fig. 1A), but they show low temperature in the thermal band (Fig. 1B) and high reflectance in the cirrus band (Fig. 1C). Most of the thin cirrus clouds are within the blue color (between 0.01 and 0.03) of the density slice of the cirrus band TOA reflectance (Fig. 1D). This indicates that if we use the same threshold (a threshold of 0.03) as the MODIS cloud algorithm, almost all these thin cirrus clouds (blue color in Fig. 1D) in the Landsat image will not be found.

3.1. Comparing Fmask results between the Landsats 4–7 and the Sentinel 2 scenarios

We first compare Fmask results between Landsats 4–7 and Sentinel 2 for all 7 sites. To assess which scenario is better, we only assess the areas where the two results differ. For each image, a total of 200 points are selected based on random sampling within the area where the results of the two scenarios disagree. Manual interpretation is carefully done for all 200 points to determine their real categories.

Fig. 3 shows Fmask results for the Landsats 4–7 and the Sentinel 2 scenarios (clouds are yellow, cloud shadows are green, and clear-sky pixels are black). The Sentinel 2 scenario result (Fig. 3B) detects many more clouds than the results for the Landsats 4–7 scenario (Fig. 3A). Most of the clouds that are detected by the Sentinel 2 scenario but missed by the Landsats 4–7 scenario are thin cirrus clouds with cirrus band TOA reflectance less than 0.03 (Fig. 2D). Note that there are also some commission errors in the river areas in the results for the Sentinel 2 scenario (upper-left corner of Fig. 3B). Table 2 illustrates the percent disagreement for Fmask results between the Landsats 4–7 and the Sentinel 2 scenarios. Large differences are observed in this Table and the largest difference (28.35% of the image) is in the category that the Landsats 4–7 scenario identifies as clear land but the Sentinel 2 scenario detects as cloud. This means the Sentinel 2 scenario identifies more clouds than the Landsats 4–7 scenario, which agrees with the visual assessment between Fig. 3A and B. The confusion matrices show that the Landsats 4–7 scenario (Table 3) misses all cloud pixels (a total of 186 pixels), while the Sentinel 2 scenario (Table 4) has made very few mistakes (8 out of 186 pixels) in this category. The overall accuracy for the places that differ between the two results indicates the Sentinel 2 scenario (91%) is dramatically higher than for the Landsats 4–7 scenario (3%). Fig. 4 shows the accuracies of Fmask results for all 7 sites comparing the Landsats 4–7 and the Sentinel 2 scenarios. The blue bars are the overall accuracies for the Landsats 4–7 scenario and the red bars are the overall accuracies for the Sentinel 2 scenario. The green bars show the percent disagreement for the two Fmask results. All 7 sites show quite large disagreements in Fmask results, especially for the images located at Path 199/Row 26 and Path 7/Row 5. For all 7 sites, Sentinel 2 Fmask results are significantly better than Landsats 4–7 Fmask results, especially for the two images located at Path 45/Row 30 and Path 233/Row 61.

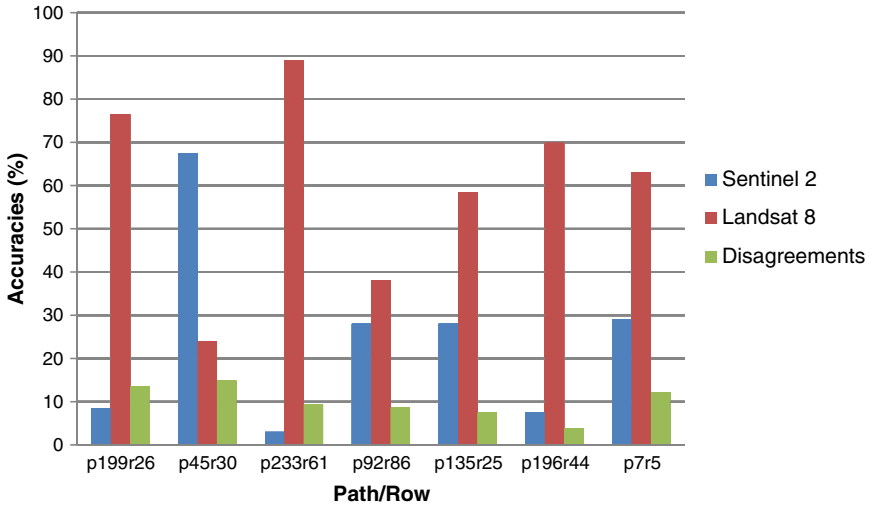

Fig. 6. Fmask results accuracies for all 7 locations between the Sentinel 2 and the Landsat 8 scenarios. The blue bars are the overall accuracies for Sentinel 2 scenario and the red bars are the overall accuracies for the Landsat 8 scenario. The green bars show the percent disagreement for Fmask results between the Sentinel 2 and the Landsat 8 scenarios.

3.2. Comparing Fmask results between the Sentinel 2 and the Landsat 8 scenarios

We compared Fmask results for the Sentinel 2 and the Landsat 8 scenarios for all 7 sites. As for the comparison above, for each site a total of 200 points were selected based on random sampling within the areas where the results of the two scenarios disagree. We carefully interpreted a total of 200 reference pixels to determine their real categories.

Fig. 5 shows that compared to the Fmask results for the Sentinel 2 scenario, Fmask results for the Landsat 8 scenario detect slightly more clouds (most of them are extremely thin cirrus clouds) and at the same time the commission errors in the river areas are greatly reduced. Table 5 illustrates the percent disagreement for Fmask results between the Sentinel 2 and the Landsat 8 scenarios. The difference between the Sentinel 2 and the Landsat 8 scenarios is relatively small, with the largest difference (5.54%) in the category where the Sentinel 2 results identify pixels as clear land that the Landsat 8 scenario detects as cloud. This means the Landsat 8 scenario identifies slightly more clouds than the Sentinel 2 scenario, which agrees with the visual assessment between Fig. 5A and B. The confusion matrices show that the Sentinel 2 scenario (Table 6) misidentifies most of the cloud pixels (191 out of 193 pixels) as clear land pixels, while the Landsat 8 scenario (Table 7) has made very few mistakes (18 out of 193) in this category. The overall accuracy for the Landsat 8 scenario (89%) is much higher than the Sentinel 2 scenario (3%). Fig. 6 shows the accuracies for all 7 sites for the Sentinel 2 and the Landsat 8 scenarios. The difference between the two scenarios is reduced significantly (approximately 10% of the image). Fmask results for the Landsat 8 scenario are better than Fmask results for the Sentinel 2 scenario for 6 of the tested sites. There is only one site in North America (Path 45/Row 30), Fmask result for the Landsat 8 scenario is better than Fmask result for the Sentinel 2 scenario. In this particular location, there are large areas of mountains with large elevation changes (Fig. S3A). As the environment temperature is negatively related with the surface elevation (Hartmann, 1994, Chap. 3), the measure brightness temperatures for high elevation pixels are usually much colder for pixels located at lower elevations (Fig. S3B). These relatively cold pixels can cause problems when the thermal band is used to calculate the temperature probability, as most of them will have high temperature probabilities and some of them may be falsely detected as cloud. For the other 6 sites, the extra thermal band available for the Landsat 8 scenario improves results relative to the Sentinel 2 scenario.

4. Discussion

While the sample of images included in this paper is relatively small, the results demonstrate the relative importance of the cirrus band and the thermal band in cloud detection. The spectral bands used by the new Fmask algorithm for the Landsats 4–7 and the Sentinel 2 scenarios are almost the same, except Sentinel 2 has a cirrus band and Landsats 4–7 has a thermal band. However, the results from the Sentinel 2 and the Landsats 4–7 scenarios are quite different. Large differences are observed for all 7 sites and the Sentinel 2 scenario shows much higher accuracies compared to the Landsats 4–7 scenario. This result indicates the relative importance of the cirrus band compared to the thermal band in designing cloud, cloud shadow, and snow detection algorithms. While it is clearly better to have multispectral optical, thermal and cirrus bands available for cloud detection, the results in this analysis indicate that if you have to choose between a cirrus band and a thermal band for cloud detection, a cirrus band is preferable. While this result may not be repeated in all locations, in general it appears that the cirrus band is extremely effective in identifying clouds.

Though we have tested the new Fmask algorithm for a total of 7 images at different parts of the world with a variety of environments (includes ice caps, desert, snow, urban, forest, grass, and mountain areas), there are still many more places with other kinds of environments that should be tested. In the future, we will collect more images for testing the algorithm and assessing the accuracies for the cloud, cloud shadow, and snow masks for the three different sensors. Moreover, it should be remembered that in this paper we are not using real Sentinel 2 images and we will be able to test the new Fmask algorithm on real Sentinel 2 images directly when they are available in the future.

5. Conclusion

We developed a new Fmask algorithm that improves the previous version of the Fmask algorithm and expands the algorithm to screen cloud, cloud shadow, and snow for Sentinel 2 and Landsat 8 images. This algorithm works well for all three scenarios (Landsats 4–7, Sentinel 2, and Landsat 8). However, due to the different spectral bands each scenario uses, the Fmask results accuracies are quite different. Fmask results from the Sentinel 2 scenario shows much better accuracies than Fmask results from the Landsats 4–7 scenario (much higher accuracy for all 7 sites). The Landsat 8 scenario performs slightly better than the Sentinel 2 scenario for most of the environments (higher accuracy for 6 sites). Considering the already high accuracies achieved by the Landsats 4–7 scenario results (Zhu & Woodcock, 2012), the new Fmask algorithms designed for Sentinel 2 and Landsat 8 are promising.

Acknowledgements

We gratefully acknowledge the supports of USGS Landsat Science Team Program for Better Use of the Landsat Temporal Domain: Monitoring Land Cover Type, Condition, and Change (grant number G11PS00422) and the supports of NASA Earth Science U.S. Participating Investigator Program for Enhancing Compatibility of Sentinel 2 and Landsat Products for Improved Monitoring of the Earth System (grant number NNX11AE18G). We would also like to extend our gratitude to the two anonymous reviewers for their thoughtful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.rse.2014.12.014.

References

Ackerman, S., Strabala, K., Menzel, P., Frey, R., Moeller, C., & Gumley, L. (2010). In MODIS Cloud Mask Team (Ed.), *Discriminating clear-sky from cloud with MODIS algorithm theoretical basis document (MOD35*. Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin.

Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., & Hanssen, R.F. (2012). ESA's sentinel missions in support of Earth system science. *Remote Sensing of Environment*, *120*, 84–90.

Coppin, P.R., & Bauer, M.E. (1994). Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features. *IEEE Transactions on Geoscience and Remote Sensing*, *32*, 918–927.

Dozier, J. (1989). Spectral signature of alpine snow cover from the Landsat Thematic Mapper. *Remote Sensing of Environment*, *28*, 9–22.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., et al. (2012). Sentinel–2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, *120*, 25–36.

Gao, B.C., Goetz, A.F., & Wiscombe, W.J. (1993). Cirrus cloud detection from airborne imaging spectrometer data using the 1.38 μm water vapor band. *Geophysical Research Letters*, *20*(4), 301–304.

Griffiths, P., van der Linden, S., Kuemmerle, T., & Hostert, P. (2013). A pixel-based Landsat compositing algorithm for large area land cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(5), 2088–2101.

Guenther, B., Xiong, X., Salomonson, V.V., Barnes, W.L., & Young, J. (2002). On-orbit performance of the Earth Observing System Moderate Resolution Imaging Spectroradiometer; first year of data. *Remote Sensing of Environment*, *83*(1), 16–30.

Hall, D.K. (2012, June 5). *MODIS snow cover algorithm changes for collection 6.* The MODIS snow & sea ice global mapping project (Online: http://modis-snow-ice.gsfc.nasa.gov/?c=collection6 (accessed March 19, 2014)).

Hall, D.K., Riggs, G.A., Salomonson, V.V., Barton, J.S., Casey, K., Chien, J.Y.L., et al. (2001, September). *Algorithm theoretical basis document (ATBD) for the MODIS snow and sea ice-mapping algorithms*. NASA GSFC.

Hartmann, D.L. (1994). *Global physical climatology* (1st ed.). San Diego: Academic Press.

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 Sensing of Environment*, *114*, 183–198.

Huang, C., Thomas, N., Goward, S.N., Masek, J.G., Zhu, Z., Townshend, J.R., et al. (2010). Automated masking of cloud and cloud shadow for forest change analysis using Landsat images. *International Journal of Remote Sensing*, *31*(20), 5449–5464.

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 Sensing of Environment*, *83*(1), 195–213.

Irish, R.R., Barker, J.L., Goward, S.N., & Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogrammetric Engineering and Remote Sensing*, *72*(10), 1179.

Irons, J.R., Dwyer, J.L., & Barsi, J.A. (2012). The next Landsat satellite: The Landsat data continuity mission. *Remote Sensing of Environment*, *122*, 11–21.

Kennedy, R.E., Andréfouët, S., Cohen, W.B., Gómez, C., Griffiths, P., Hais, M., et al. (2014). Bringing an ecological view of change to Landsat-based remote sensing. *Frontiers in Ecology and the Environment*, *12*(6), 339–346.

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 Sensing of Environment*, *114*(12), 2897–2910.

Khlopenkov, K.V., & Trishchenko, A.P. (2007). SPARC: New cloud, snow, and cloud shadow detection scheme for historical 1-km AVHHR data over Canada. *Journal of Atmospheric and Oceanic Technology*, *24*(3).

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 Sensing of Environment*, *112*(12), 4167–4185.

Maiersperger, T., Scaramuzza, P., Leigh, L., Shrestha, S., Gallo, K., Jenkerson, C., et al. (2013). Characterizing LEDAPS surface reflectance products by comparisons with AERONET, field spectrometer, and MODIS data. *Remote Sensing of Environment*, *136*, 1–13.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., et al. (2006). A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geoscience and Remote Sensing Letters*, *3*(1), 68–72.

Oreopoulos, L., Wilson, M.J., & Várnai, T. (2011). Implementation on Landsat data of a simple cloud-mask algorithm developed for MODIS land bands. *IEEE Geoscience and Remote Sensing Letters*, *8*(4), 597–601.

Platnick, S., King, M.D., Ackerman, S.A., Menzel, W.P., Baum, B.A., Riédi, J.C., et al. (2003). The MODIS cloud products: Algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 459–473.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., et al. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment*, *114*(1), 35–49.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, et al. (2014). Landsat-8: science and product vision for terrestrial global change research. *Remote Sensing of Environment*, *145*, 154–172.

Schott, J.R., Gerace, A., Raqueno, N., Ientilucci, E., Raqueno, R., & Lunsford, A.W. (2014, October). Chasing the TIRS ghosts: calibrating the Landsat 8 thermal bands. *SPIE optical engineering + applications* (pp. 1–20). International Society for Optics and Photonics.

Soille, P. (1999). *Morphological image analysis: principles and applications*. Springer-Verlag, 173–174.

Soille, P., Vogt, J., & Colombo, R. (2003). Carving and adaptive drainage enforcement of grid digital elevation models. *Water Resources Research*, *39*(12), 1–13.

Vermote, E.F., El Saleous, N.Z., & Justice, C.O. (2002). Atmospheric correction of MODIS data in the visible to middle infrared: first results. *Remote Sensing of Environment*, *83*(1), 97–111.

White, J. C., Wulder, M. A., Hobart, G. W., Luther, J. E., Hermosilla, T., Griffiths, P., et al. (2014). Pixel-based image compositing for large-area dense time series applications and science. *Canadian Journal of Remote Sensing*, *40*(3), 192–212.

Wilson, M.J., & Oreopoulos, L. (2013). Enhancing a simple MODIS cloud mask algorithm for the Landsat data continuity mission. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(2), 723–731.

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., et al. (2008). Free access to Landsat imagery. *Science*, *320*(5879), 1011.

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 Sensing of Environment*, *82*(2), 173–187.

Zhu, Z., & Woodcock, C.E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, *118*(15), 83–94.

Zhu, Z., & Woodcock, C.E. (2014a). Automated cloud, cloud shadow, and snow detection based on multitemporal Landsat data: an algorithm designed specifically for land cover change. *Remote Sensing of Environment*, *152*, 217–234.

Zhu, Z., & Woodcock, C.E. (2014b). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, *144*, 152–171.