



Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery

Shi Qiu^{a,c}, Zhe Zhu^{b,c,*}, Binbin He^{a,d,**}

^a School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

^b Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 06269, USA

^c Department of Geosciences, Texas Tech University, Lubbock, TX 79409, USA

^d Center for Information Geoscience, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

ARTICLE INFO

Edited by Menghua Wang

Keywords:

Cloud
Cloud shadow
Landsat
Sentinel-2
Auxiliary data
Spectral-contextual
New cloud probability

ABSTRACT

We developed the Function of mask (Fmask) 4.0 algorithm for automated cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 images. Three major innovative improvements were made as follows: (1) integration of auxiliary data, where Global Surface Water Occurrence (GSWO) data was used to improve the separation of land and water, and a global Digital Elevation Model (DEM) was used to normalize thermal and cirrus bands; (2) development of new cloud probabilities, in which a Haze Optimized Transformation (HOT)-based cloud probability was designed to replace temperature probability for Sentinel-2 images, and cloud probabilities were combined and re-calibrated for different sensors against a global reference dataset; and (3) utilization of spectral-contextual features, where a Spectral-Contextual Snow Index (SCSI) was created for better distinguishing snow/ice from clouds in polar regions, and a morphology-based approach was applied to reduce the commission error in bright land surfaces (e.g., urban/built-up and mountain snow/ice). The Fmask 4.0 algorithm showed higher overall accuracies for Landsats 4–8 imagery than the 3.3 version (Zhu et al., 2015) (92.40% versus 90.73% for Landsats 4–7 and 94.59% versus 93.30% for Landsat 8), and much higher overall accuracies for Sentinel-2 imagery than the 2.5.5 version of the Sen2Cor algorithm (Müller-Wilm et al., 2018) (94.30% versus 87.10%).

1. Introduction

Moderate spatial resolution images (10 to 30 m) from Landsats 4–8 and Sentinel-2 provide a great opportunity for monitoring global environmental change (Drusch et al., 2012; Roy et al., 2014; Storey et al., 2016; Wulder et al., 2015). However, these optical images are extensively influenced by clouds and their shadows (Ju and Roy, 2008; Li and Roy, 2017), and their detections are usually the first and most crucial preprocessing step. Although manual interpretation can create accurate cloud and cloud shadow masks, it is very time-consuming and no longer acceptable for processing a large volume of images, particularly since the free and open policy of Landsat and Sentinel-2 data (Turner et al., 2015; Woodcock et al., 2008; Zhu et al., 2019). Therefore, it is important to design fully automated algorithms to detect clouds and their shadows in Landsats 4–8 and Sentinel-2 images.

Based on how many images are used for detecting clouds and cloud shadows, we can categorize the algorithms into two groups: single-date algorithms and multitemporal algorithms (Zhu et al., 2018). The

multitemporal algorithms are particularly popular after the free policy of Landsat and Sentinel-2 (Frantz et al., 2015; Goodwin et al., 2013; Hagolle et al., 2010; Wang et al., 1999; Zhu and Helmer, 2018; Zhu and Woodcock, 2014). As the presence of cloud and cloud shadow will lead to sudden changes in reflectance, the multitemporal algorithms can identify clouds and cloud shadows by comparing the cloudy image with a clear-sky reference image (or a model predicted clear-sky image). Though the multitemporal algorithms are reported to have higher accuracies than the single-date algorithms, they are more complicated to use (require clear-sky reference images or high-density time series data) and may cause problems in monitoring land change (Zhu and Woodcock, 2014).

The single-date algorithms use one satellite image to detect clouds (or cloud shadows), and they can be divided into two categories: physical-rule-based approaches and machine-learning-based approaches (Zhu et al., 2018). The machine-learning-based approaches are straightforward and simple to use. Based on the training samples of clouds (or cloud shadows), we can classify every cloud (or cloud

* Correspondence to: Z. Zhu, Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT 06269, USA.

** Correspondence to: B. He, School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China.

E-mail addresses: zhe@uconn.edu (Z. Zhu), binbinhe@uestc.edu.cn (B. He).

shadow) pixel based on a variety of classifiers, such as decision trees (Hollstein et al., 2016; Potapov et al., 2011; Roy et al., 2010; Scaramuzza et al., 2012), neural networks (Hughes and Hayes, 2014; Zi and Xie, 2018), fuzzy models (Melesse and Jordan, 2002; Shao et al., 2017), and support vector machines (Zhou et al., 2016). This kind of methods require large training samples and may fail in certain unique conditions without enough representative training samples (Huang et al., 2010b; Hughes and Hayes, 2014).

The physical-rule-based approaches detect clouds based on the physical characteristics of clouds, such as “white” (similar reflectance across spectral bands), “bright” (high reflectance), “cold” (low temperature), and “high” (high in elevation). Most of the algorithms are using a constant or adaptive threshold in different spectral bands derived from these physical rules to detect clouds (Choi and Bindschadler, 2004; Frantz et al., 2018; Huang et al., 2010b; Irish, 2000; Irish et al., 2006; Louis et al., 2016; Müller-Wilm et al., 2018; Oishi et al., 2018; Oreopoulos et al., 2011; Qiu et al., 2017; Sun et al., 2017; Vermote and Saleous, 2007; Wilson and Oreopoulos, 2013; Zhu et al., 2015; Zhu and Woodcock, 2012), in which some of them can also detect cloud shadows by projecting their corresponding clouds onto land surfaces based on the relative geometry among the sun, the satellite, and the clouds (Huang et al., 2010b; Qiu et al., 2017; Vermote and Saleous, 2007; Zhu et al., 2015; Zhu and Woodcock, 2012).

Among all these physical-rule-based algorithms, the Function of mask (Fmask) algorithm is a widely used cloud and cloud shadow detection algorithm for both Landsats 4–8 and Sentinel-2 images (Zhu et al., 2015; Zhu and Woodcock, 2012). It detects clouds and cloud shadows based on rules and statistics derived from their physical characteristics. In particular, clouds are identified based on their characteristics of “white”, “bright”, “cold”, and “high”, and cloud shadows are identified by considering their “dark” characteristic and based on the assumption that cloud shadows share similar shape with the corresponding clouds and follow the cloud-shadow projection geometry. This approach (Fmask version 3.3) has been adopted by U.S. Geological Survey (USGS) to provide Quality Assessment (QA) band for all Landsats 4–8 Collection 1 products due to its high overall accuracy (Foga et al., 2017). Though Fmask has been successfully used in various

studies around the world (Claverie and Masek, 2017; Jönsson et al., 2017; Li et al., 2015; Melaas et al., 2013; Pasquarella et al., 2018; Pekel et al., 2016; Yan and Roy, 2014; Zhu et al., 2016), it has issues that cannot be ignored.

1.1. The blindness of the geographic area

Fmask assumes we know nothing about the area covered by the satellite images, which makes it easy to use but also incurs problems. For example, Fmask detects clouds over land and water separately, but land and water were simply separated by a few spectral tests. For optically thick clouds, there is no way to tell what is underneath the clouds based on the same satellite image, unless the land/water mask is available ahead of time. Moreover, the previous version of Fmask uses Brightness Temperature (BT) derived from thermal band to identify clouds in Landsat data, which is problematic in mountainous areas, where BT is often negatively related to elevation (Fig. 1c and d). Also, the cirrus band, used for improving cirrus cloud detection in Fmask for Landsat 8 and Sentinel-2 images, has the similar physical characteristic that it can be influenced by magnitude of elevation and they are positively related (Fig. 1c and e). That is because the cirrus band is a water vapor absorption band, and higher altitude corresponds to shorter water vapor path, which results in higher Top Of Atmosphere (TOA) reflectance (Gao et al., 1993). These impacts will easily cause commission errors of clouds in high mountains (Fig. 1b), which cannot be addressed unless other auxiliary data such as Digital Elevation Model (DEM) is included to correct their impacts. Overall, auxiliary data such as land/water mask and DEM can provide location-based prior-knowledge for better detection of clouds. Recently, the Mountainous Fmask (MFmask) algorithm tried to reduce the effect of elevation on the thermal band based on DEMs (Qiu et al., 2017) but ignored its impact on the cirrus band.

1.2. Lack of adequate cloud probabilities

Fmask calculates spectral variation, temperature, and cirrus cloud probabilities to capture the “white”, “cold”, and “high” characteristics

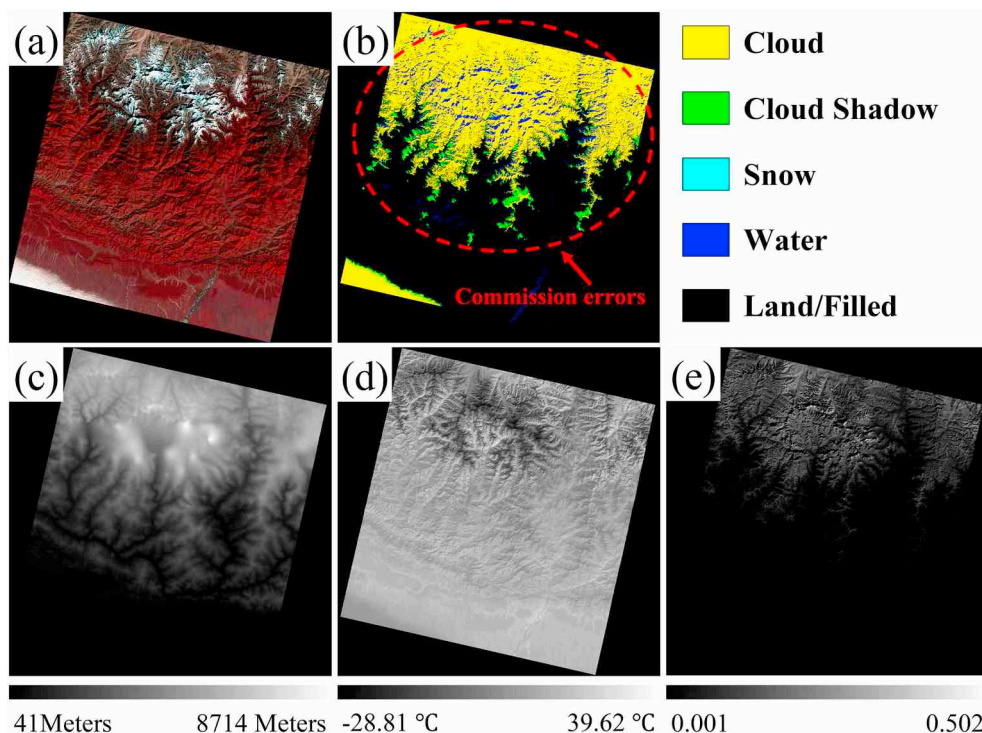


Fig. 1. An example of Fmask 3.3 issue caused by blindness of the geographic area. (a) False color composite image (NIR, red, and green bands). (b) Fmask results. (c) DEM. (d) Brightness temperature. (e) Cirrus band TOA reflectance. The results are from a Landsat 8 image acquired on Jan. 9th, 2018 at Path 140/Row 41. The “clouds” in the red dashed circle are the commission errors from the Fmask 3.3 results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

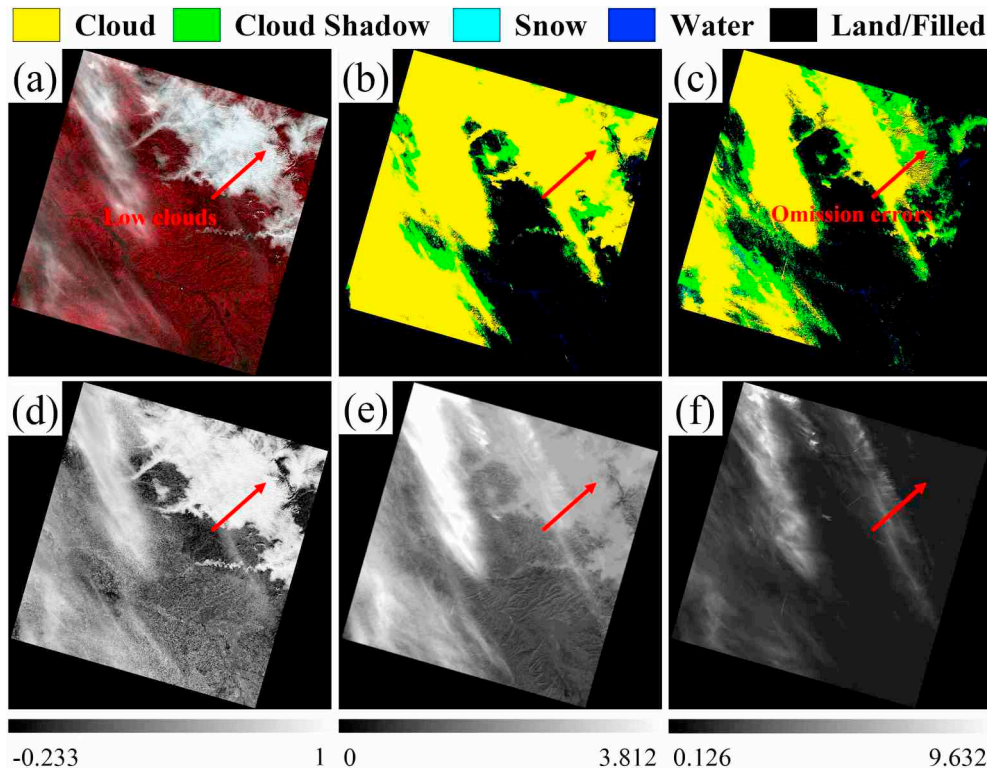


Fig. 2. An example of Fmask 3.3 issue caused inadequate cloud probabilities. (a) False color composite image (NIR, red, and green bands). (b) Fmask results using the thermal band. (c) Fmask results without using the thermal band for the corresponding Sentinel-2 image. (d) Spectral variation probability. (e) Temperature probability. (f) Cirrus cloud probability. The results are from a Landsat 8 image acquired on Jan. 17th, 2016 at Path 199/Row 29. The red arrows point out the location of omission errors that are caused by lack of thermal band. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of clouds over land, respectively (Fig. 2d, e, and f). This works well for Landsat but is problematic for Sentinel-2 (Fig. 2b and c). As Sentinel-2 does not have the thermal band, the lack of temperature probability can cause serious omission errors in cloud detection (Fig. 2c). Besides, Fmask heavily depends on the cirrus band to detect cirrus clouds because of their high TOA reflectance, but certain atmospheric and surface conditions (e.g., aerosol, dry vegetation, snow, and desert) can also present high spatial variability in water vapor and have relatively high cirrus band TOA reflectance, especially in dry environments (Bouffies et al., 1997; Gao et al., 1993; Gao and Kaufman, 2003). Over-dependence on the cirrus cloud probability may result in large commission errors for the certain atmospheric and surface conditions. Moreover, Fmask detects clouds using an image-based cloud probability threshold, which was calibrated only based on Landsat 7 data in the past (Zhu and Woodcock, 2012). It might not be the global optimal threshold for different sensors, particularly when changes are made in calculating cloud probability.

1.3. Confusion with bright and cold surfaces

Bright and cold surfaces (e.g., urban/built-up and snow/ice) can have the same spectral signatures as clouds (Dozier, 1989; Frantz et al., 2018; Oishi et al., 2018), which are easily confused with clouds in cloud detection algorithms. Theoretically, most of clouds can be separated from snow/ice because clouds generally have higher reflectance in Shortwave Infrared (SWIR) band than snow/ice (Choi and Bindshadler, 2004; Hall et al., 1995; Irish, 2000). However, some clouds composed of ice crystals will have very similar SWIR reflectance as snow/ice (Dozier, 1989). The thermal and cirrus bands are also helpful for separating clouds and snow/ice, but for snow and ice at specific locations (e.g., polar regions and high mountains), they may show similar spectral patterns as clouds. Another typical kind of commission error is noticeable in urban/built-up areas. Most urban/built-up can be correctly separated from clouds in Landsat data using the thermal band because they are usually much warmer than clouds (Landsberg, 1981). However, some urban/built-up objects (e.g., cool

roof) can have temperatures as cold as or even colder than clouds (Akbari et al., 2001), which makes them almost impossible to separate from clouds in the spectral domain. This issue is even worse for Sentinel-2 data, as without the thermal band, urban/built-up areas are easily confused with clouds even if they are relatively warmer than the clouds. Those errors can cause severe problems for time series analysis, as if we constantly misclassify snow/ice or urban/built-up as cloud, we will have no or very few clear observations to use for these locations (Fig. 3).

In this study, we will present Fmask 4.0 algorithm mainly by providing solutions for the above three major problems as follows: (1) integrate auxiliary data from global water mask and DEM to improve cloud detection (Section 3.1); (2) create new cloud probabilities and calibrate new global optimal probability thresholds for different sensors (Section 3.2); and (3) combine the spectral and contextual features to reduce false positive errors in cloud detection (Section 3.3).

2. Data

2.1. Landsat and Sentinel-2 data

Major inputs of Fmask 4.0 include TOA reflectance and/or BT from Landsats 4–8 and Sentinel-2 (Table 1). As all the spectral bands in Landsats 4–8 data are provided at a spatial resolution of 30-meter, Fmask 4.0 creates cloud and cloud shadow masks at the same 30-meter spatial resolution. For Sentinel-2 data, three different spatial resolutions of 10-, 20-, and 60-meter are provided for different spectral bands in a tile of $100 \times 100 \text{ km}^2$. To keep a similar cloud and cloud shadow mask spatial resolution as Landsat imagery and considering Sentinel-2's spatial resolution (many spectral bands are at 20-meter spatial resolution), Fmask 4.0 creates cloud and cloud shadow masks at a 20-meter resolution for Sentinel-2. The Sentinel-2 spectral bands with 10- and 60-meter spatial resolutions will be resampled to 20-meter, in which the 10-meter bands are based on a pixel aggregation method and the 60-meter bands are based on a nearest sampling approach. For the original 20-meter bands, we keep them the same.

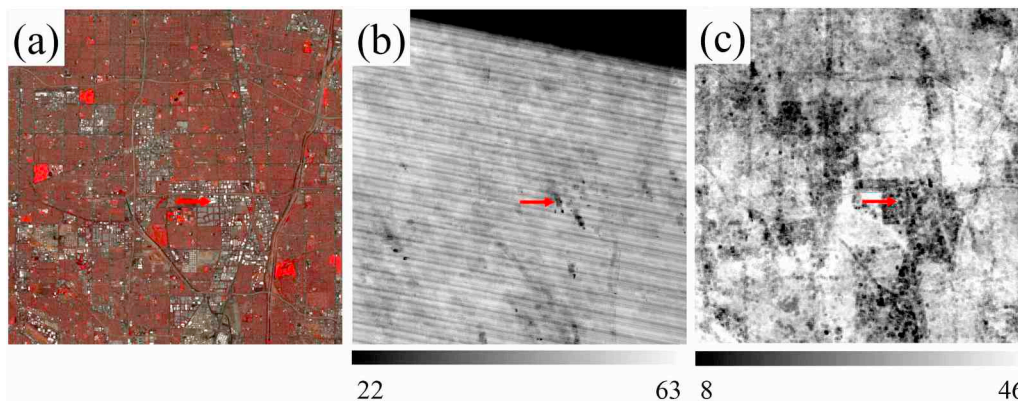


Fig. 3. An example of Fmask 3.3 issue caused by consistent commission errors of clouds from bright and cold surfaces. (a) False color composite image (NIR, red, and green bands) for the Landsat 8 image acquired on Jul. 24th, 2016. (b) Number of total clear observations based on Fmask 3.3 statistics from 91 Landsats 7–8 images. (c) Number of total clear observations based on Fmask 3.3 statistics from 112 Sentinel-2 images. Note the stripes in Fig. 3b are caused by the failure of Landsat 7 scan line corrector after May 31st, 2003. The red arrows indicate the places where urban/built-up pixels are misidentified as clouds consistently, which

greatly reduced the number of available clear observations. The statistics are from all available Landsat images (Path 41/Row 37) and Sentinel-2 images (Tile 11SLT) collected between January 2016 and December 2017 over the Los Angeles area ($18 \times 18 \text{ km}^2$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

The spectral bands of Landsats 4–8 and Sentinel-2 data. Bands used in Fmask 4.0 are highlighted in bold letters. The NIR band in this study indicates Band 4 for Landsats 4–7, Band 5 for Landsat 8, and Band 8a for Sentinel-2 if not otherwise specified.

Band Name	Landsats 4–5 TM Bands (μm)	Landsat 7 ETM+ Bands (μm)	Landsat 8 OLI/TIRS Bands (μm)	Sentinel-2 MSI Bands (μm)
\	\	\	Band 1 (0.435–0.451)	Band 1 (0.433–0.453)
Blue	Band 1 (0.45–0.52)	Band 1 (0.45–0.52)	Band 2 (0.452–0.512)	Band 2 (0.458–0.523)
Green	Band 2 (0.52–0.60)	Band 2 (0.52–0.60)	Band 3 (0.533–0.590)	Band 3 (0.543–0.578)
Red	Band 3 (0.63–0.69)	Band 3 (0.63–0.69)	Band 4 (0.636–0.673)	Band 4 (0.650–0.680)
\	\	\	Band 5 (0.698–0.713)	Band 5 (0.698–0.713)
VRE3	\	\	\	Band 6 (0.733–0.748)
Wide NIR	Band 4 (0.76–0.90)	Band 4 (0.77–0.90)	\	Band 7 (0.765–0.785)
Narrow NIR	\	\	Band 5 (0.851–0.879)	Band 8 (0.785–0.900)
\	\	\	\	Band 8a (0.855–0.875)
Cirrus	\	\	Band 9 (1.363–1.384)	Band 9 (0.930–0.950)
SWIR1	Band 5 (1.55–1.75)	Band 5 (1.55–1.75)	Band 6 (1.566–1.651)	Band 10 (1.365–1.385)
SWIR2	Band 7 (2.08–2.35)	Band 7 (2.09–2.35)	Band 7 (2.107–2.294)	Band 11 (1.565–1.655)
Thermal	Band 6 (10.40–12.50)	Band 6 (10.40–12.50)	Band 10 (10.60–11.19)	Band 12 (2.100–2.280)
\	\	\	Band 11 (11.50–12.51)	\
\	\	Band 8 (0.52–0.90)	Band 8 (0.503–0.676)	\

TM: Thematic Mapper; ETM+: Enhanced Thematic Mapper Plus; OLI/TIRS: Operational Land Imager/Thermal Infrared Sensor; MSI: Multispectral Instrument; VRE: Vegetation Red Edge; NIR: Near Infrared; SWIR: Shortwave Infrared.

2.2. Auxiliary data

Global Surface Water Occurrence (GSWO), DEM, and DEM derivatives (slope and aspect) are the auxiliary data used in Fmask 4.0. The GSWO dataset provides terrestrial water dynamics (the intra- and inter-annual variability and change) over long time periods at a 30-meter resolution based on a 32-year Landsat record (Pekel et al., 2016). The dataset creates water occurrence for each pixel, where 0% indicates permanent land and 100% indicates permanent water. For DEM data, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) DEM Version 2 was used for algorithm development because of its global availability and sufficient accuracy (Nikolakopoulos et al., 2006; Tachikawa et al., 2011). ASTER DEM data are provided with 1 arc-second resolution (approximately 30 m), which can generate the corresponding slope and aspect with the aid of TopoToolbox (Schwanghart and Scherler, 2014).

2.3. Training and validation data

2.3.1. Training data

Manually interpreted cloud and cloud shadow reference masks derived from “L8 Biome” dataset will be used as the fundamentals to develop Fmask 4.0 (Foga et al., 2017). “L8 Biome” dataset includes a total of 96 Landsat 8 images covering different parts of the world and with

different cloud cover conditions. Those images were selected using a biome-based stratified sampling approach, covering urban, barren, forest, shrubland, grass/cropland, snow/ice, wetlands, and water. The masks for each image were produced by multiple visual criteria (e.g., brightness, shape, and texture) with various band combinations.

We examined this dataset carefully and excluded six images with low accuracies, which give us 90 Landsat 8 images (Fig. 4; Table S1). Among the 90 Landsat 8 images, we selected 17 images (mostly in the polar regions) which only contain snow/ice or clouds to build the routine to separate snow/ice and clouds (Fig. 4; Table S1). As Landsat 8 has all Landsats 4–7 bands and most of Sentinel-2 bands used in Fmask 4.0, we used Landsat 8 data to simulate Landsats 4–7 and Sentinel-2 data and train the algorithm based on the same reference dataset derived from Landsat 8. As Sentinel-2 images have different spatial resolution and spatial extent compared with Landsat 8 images, we clipped Landsat 8 images into the same $100 \times 100 \text{ km}^2$ and resampled to the 20-meter resolution. All training data will be used as the basis for development and calibration of Fmask 4.0.

2.3.2. Validation data

To quantitatively assess Fmask 4.0 accuracy for Landsat 7 (representing Landsats 4–7 considering of their very similar spectral bands), Landsat 8, and Sentinel-2 respectively, we randomly selected two locations from each of the eight land cover types of the “L8 Biome”

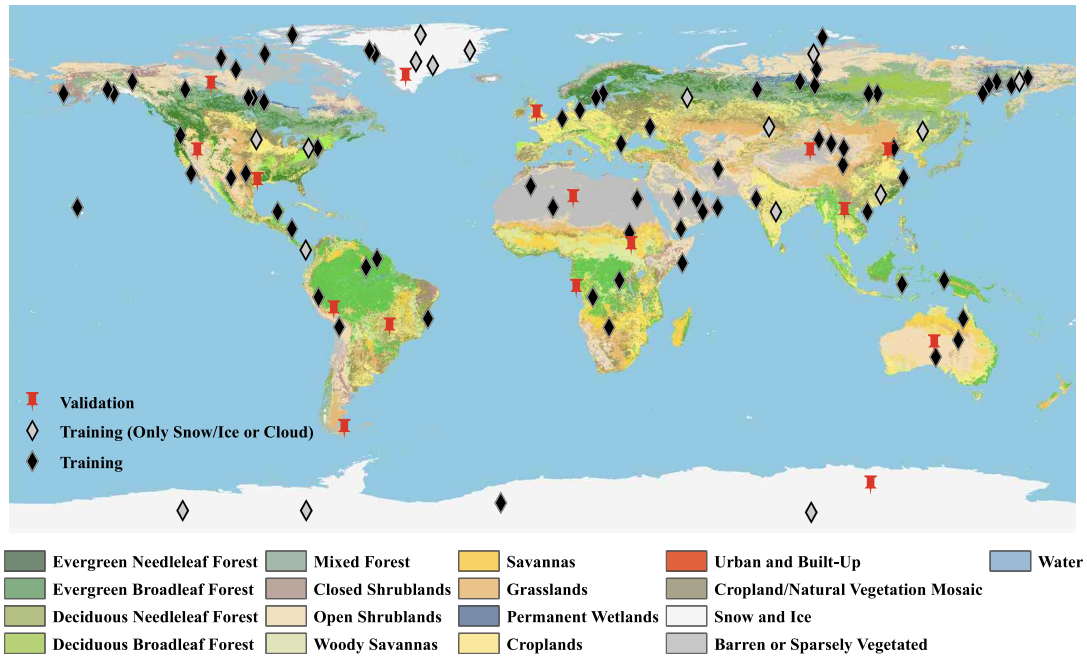


Fig. 4. Global distribution of training and validation data used in this study. The land cover background is derived from Global Land Cover Characterization (GLCC) product (Loveland et al., 2000).

dataset (Foga et al., 2017), and from each location we randomly selected one image for each sensor type. This will give us 16 validation images for each sensor (Fig. 4; Table S2-S4). For each image, a total of 100 pixels were selected based on the simple random sampling. The interpretation was carefully made to determine their true categories (e.g., cloud, cloud shadow, or clear) by looking at the original satellite images. Sometimes it is hard to make a confident interpretation for certain samples, especially for optically thin clouds and their shadows (hard to define their boundaries). In this case, we excluded those samples in our accuracy assessment and thus there are a total of 1553 reference samples for Landsats 4–7 (719 cloud samples, 62 cloud shadow samples, and 772 clear-sky samples), 1552 reference samples for Landsat 8 (686 cloud samples, 72 cloud shadow samples, and 794 clear-sky samples), and 1527 reference samples for Sentinel-2 (724 cloud samples, 56 cloud shadow samples, 747 clear-sky samples).

3. Fmask 4.0

The overall flowchart of Fmask 4.0 is illustrated in Fig. 5. As the previous Fmask algorithms have been well documented (Qiu et al., 2017; Zhu et al., 2015; Zhu and Woodcock, 2012), we mainly focus on the new methodologies used for solving the problems we have mentioned above (Section 1). Note that the Fmask 4.0 algorithm described here is mainly focused on cloud detection, and cloud shadow detection of Fmask 4.0 is directly inherited from MFmask (Qiu et al., 2017).

3.1. Auxiliary data integration

The auxiliary data (e.g., GSWO, DEM, slope, and aspect) are essential for Fmask 4.0. Previously, DEM, slope, and aspect have been applied to improve cloud and cloud shadow detection by normalizing thermal band, removing terrain shadows, and revising cloud shadow shape over slope-side, which are particularly helpful in mountainous areas (Qiu et al., 2017). Besides these enhancements, Fmask 4.0 will include GSWO to better separate land and water (Section 3.1.1). DEM will also be applied to normalize the cirrus band obtained by Landsat 8 and Sentinel-2 to reduce its variations along the elevations (Section

3.1.2). Finally, we will down-scale all auxiliary data and test alternative options of auxiliary data to find a balance between accuracy and efficiency (Section 3.1.3).

3.1.1. Better separation of land and water

Fmask applies different approaches for detecting clouds over land and water; therefore, it is essential to know the underlying surface types (land or water) before cloud detection is applied. Water pixel was initially determined only using Normalized Difference Vegetation Index (NDVI) and Near Infrared (NIR) band (hereafter called “spectrally derived water”); and the remaining pixels were grouped into land pixels (Zhu and Woodcock, 2012). This approach can separate land and water pixels well when they are clear-sky or thin cloud pixels, but will not work for places covered by thick clouds.

Water Test

$$= (NDVI < 0.01 \text{ and Band NIR} < 0.11) \text{ or } (NDVI < 0.1 \text{ and Band NIR} < 0.05) \text{ or } (GSWO > O_{\text{water}} \text{ and snow/ice}) \\ = \text{false} \quad (1)$$

where,

$$O_{\text{water}} = 17.5 \text{ percentile of spectrally derived water pixels' occurrence} - 5\%$$

Fmask 4.0 integrates GSWO data into the water test to better separate land and water by Eq. (1), mainly for places covered by thick clouds. GSWO can provide water occurrence from 0% to 100% for each pixel. Considering water occurrence varies in different areas, we used the spectrally derived water pixels to calculate an image-based water occurrence threshold to extract all water from GSWO data. The 17.5 percentile (lower level) is inherited from the original Fmask algorithm (Zhu and Woodcock, 2012) to exclude commission errors from the spectrally derived water pixels, and 5% can tolerate the uncertainty of the GSWO data. At the same time, snow/ice pixels need to be excluded because GSWO will also include them as water. This approach can separate land and water better even with heavy cloud coverage.

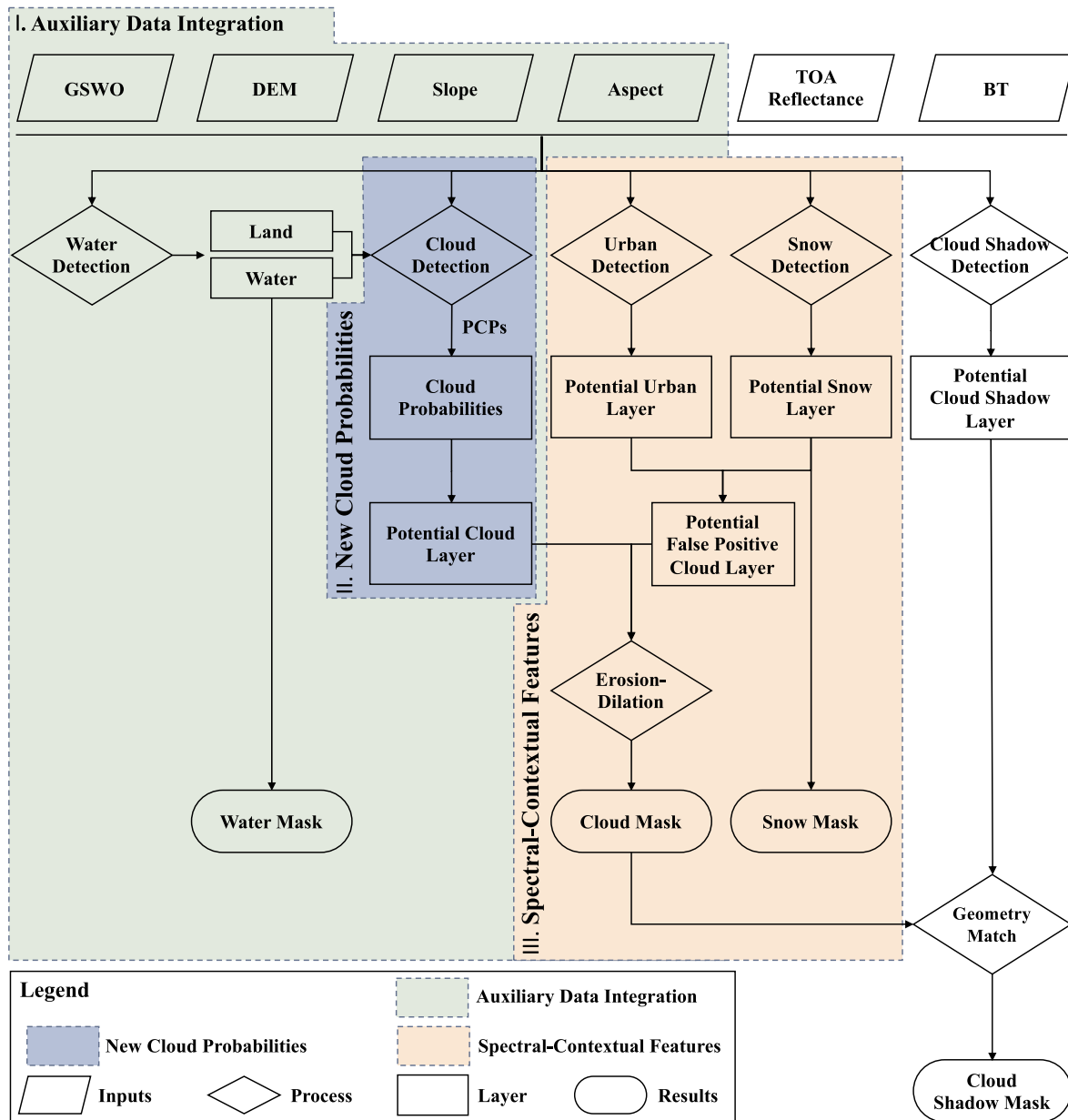


Fig. 5. Flowchart of Fmask 4.0. GSWO: Global Surface Water Occurrence; DEM: Digital Elevation Model; BT: Brightness Temperature; PCPs: Potential Cloud Pixels.

3.1.2. Normalization of cirrus band (for Landsat 8 and Sentinel-2)

Due to the strong water vapor absorption in the lower atmosphere, the cirrus band is particularly useful to detect cirrus clouds that are usually high in elevation (Gao et al., 1993; Gao and Kaufman, 1995). However, elevation variations can also impact on this band as places with high elevations will have higher cirrus band TOA reflectance due to the shorter water vapor path (Fig. 6a).

To reduce the elevation-induced variability of cirrus band TOA, we normalized the cirrus band based on DEM in Fmask 4.0. The normalization is based on dark cirrus band samples extracted from isometric zones for every 100-meter elevation, which are determined by the two percentiles of the clear-sky pixels' cirrus band TOA reflectance in each zone. The clear-sky pixels are non-Potential Cloud Pixels (non-PCPs) identified by several spectral tests in Fmask (Zhu and Woodcock, 2012). Assuming these dark cirrus band samples are only impacted by the different water vapor absorption paths caused by elevation change, we subtracted the dark cirrus band sample's TOA reflectance from each pixel's cirrus band TOA reflectance within each isometric zone. This normalization can greatly reduce the variability of cirrus band caused

by elevation change, making cirrus band TOA reflectance only responsible for the presence of cirrus clouds (Fig. 6b). Note that if the image is full of clouds and therefore does not have enough clear-sky pixels for further analysis, Fmask 4.0 will directly adopt the original Fmask approach without applying this normalization process.

3.1.3. Auxiliary data upscale to improve efficiency

We developed our algorithm based on auxiliary data from 30-meter GSWO and 30-meter ASTER DEM. However, considering GSWO and DEM are mainly used to provide a regional statistic, it is possible that coarser spatial resolution of these auxiliary data may achieve similar accuracies. This is particularly important because of the large computation/storage and extensive efforts involved in manually downloading, mosaicking, projecting, and resampling all auxiliary data to the same resolution and extent of remotely sensed imagery.

Therefore, we did a sensitivity analysis by examining the relationship between auxiliary data spatial resolution and detection accuracy (Fig. 7). When the spatial resolution of the DEM varies from 30 to 1950 m, the average cloud detection overall accuracy decreased very

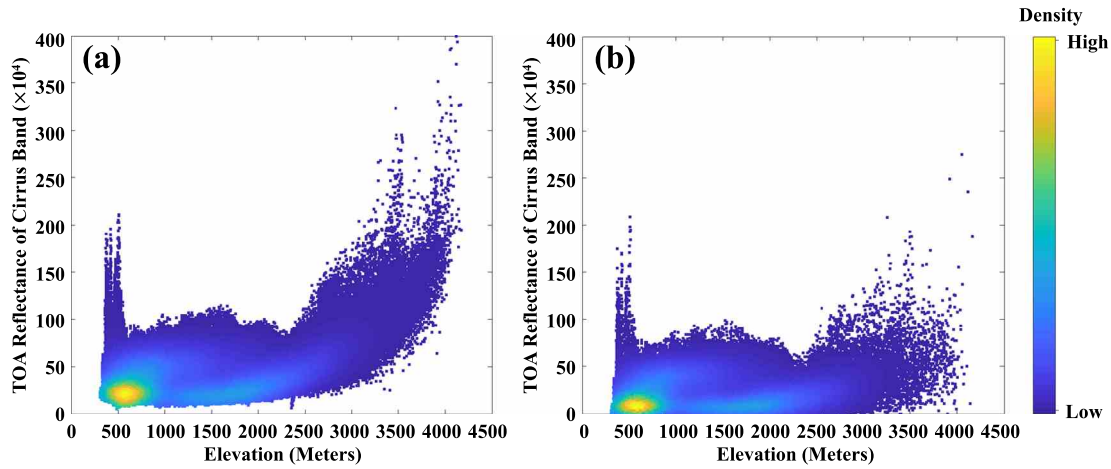


Fig. 6. Normalization of the cirrus band for a Sentinel-2A image acquired on May 7th, 2016 at Tile 29RNQ. (a) Original cirrus band TOA reflectance versus elevation. (b) Normalized cirrus band TOA reflectance versus elevation.

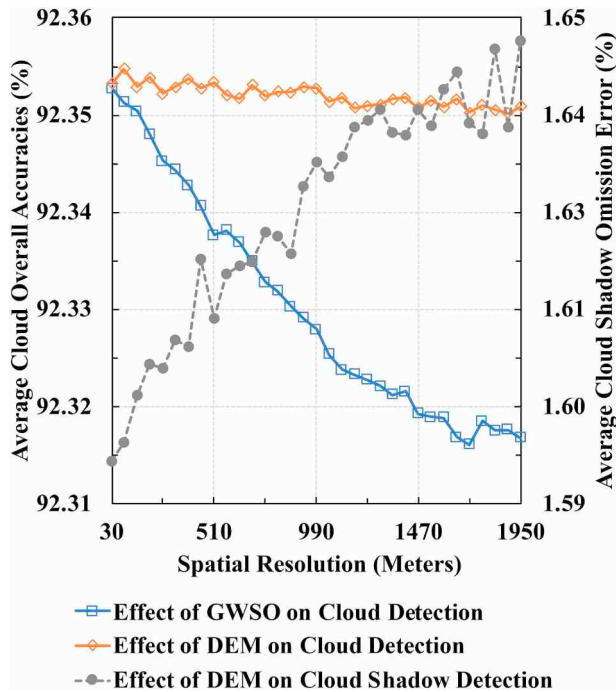


Fig. 7. Effects of the spatial resolution of DEM and GSWO on cloud and cloud shadow detection accuracies based on 90 Landsat 8 images.

slightly (approximately 0.005%), and the average omission error in cloud shadow detection increased marginally ($< 0.05\%$). Note that the cloud detection overall accuracy is calculated as the ratio of the agreement between manual cloud mask and Fmask cloud results to the total number of pixels; and the cloud shadow omission error is calculated as the ratio of the number of omitted cloud shadow pixels to the total number of pixels. We only focused on the omission errors for cloud shadows because the commission errors are generally less of an issue because of the relatively small sizes of cloud shadows (Qiu et al., 2017; Zhu and Woodcock, 2012). Considering the small accuracy differences when the spatial resolution of the DEM changes, we think DEM with 1 km would generally work for our purpose. In this case, we will be able to use the USGS Global 30 Arc-Second Elevation (GTOPO30) data as the default data source for DEM auxiliary data. On the other hand, the responses of different spatial resolutions of GSWO indicate it is important to maintain the high spatial resolution of this variable

(accuracy drops dramatically with the increase of spatial resolution of GSWO). To balance accuracy and efficiency, we recommend resampling the GSWO into 150 m, which can reach a similar detection accuracy and reduce 96% volume of the GSWO auxiliary data.

3.2. New cloud probabilities

Cloud probabilities are the core components of Fmask, and Fmask proposed different probabilities over land and water, respectively. Fmask 4.0 still uses the original approach for water (Zhu and Woodcock, 2012), but mainly improves cloud probabilities over land that are originally consisted of spectral variation, temperature, and/or cirrus cloud probabilities. The previous version of Fmask calculates the spectral variation probability from visible to SWIR bands to capture the “white” feature of cloud, but ignored the contrast between NIR and SWIR bands; therefore, we need to modify the spectral variability probability by adding this unique contrast test to further exclude the commission errors of clouds from the specific surfaces with significant NIR-SWIR contrasts (such as built-up and barren pixels) (Section 3.2.1). Moreover, Fmask calculates the temperature probability based on the thermal band for capturing the “cold” feature of the cloud. However, Sentinel-2 does not have the thermal band. Therefore, we need to invent a new cloud probability to replace this important probability (Section 3.2.2). Additionally, the cirrus cloud probability can be calculated based on the cirrus band from Landsat 8 and Sentinel-2, but the previous Fmask weighted heavily on the cirrus cloud probability, and cloud detection may sometimes be dominated by this probability. Therefore, we will adjust the new cloud probabilities by reducing the weight of cirrus cloud probability (Section 3.2.3). Finally, considering many improvements have been made in Fmask 4.0, the original cloud probability thresholds are also outdated. We will need to calibrate the global optimal cloud probability thresholds for different sensors for Fmask 4.0 (Section 3.2.4).

3.2.1. Spectral variation probability

While land surface has large spectral variability, clouds usually present very consistent spectral reflectance across the optical bands. Besides NDVI, Normalized Difference Snow Index (NDSI), and Whiteness index, Fmask 4.0 adds Normalized Difference Built-up Index (NDBI) to calculate the spectral variation probability over land (Zha et al., 2003). The NDVI and NDSI can be used to capture the spectral variability in NIR/red, and SWIR/green dimensions. The whiteness is calculated to capture the variability in the blue, red, and green bands (Zhu and Woodcock, 2012). By including NDBI, Fmask 4.0 can capture

the spectral variability in NIR/SWIR dimension. This is helpful to exclude built-up and barren in cloud detection because built-up normally shows positive NDBI, but clouds are the opposite (Zha et al., 2003).

3.2.2. HOT-based cloud probability (for Sentinel-2)

As Sentinel-2 does not have the thermal band, the previous version of Fmask only calculates the spectral variation probability and the cirrus cloud probability to capture the “white” and “high” features of cloud (Zhu et al., 2015). For low altitude clouds, however, the cirrus cloud probability will be almost the same as that for the Earth's surface, resulting in large omission errors. To solve this problem, a new Haze Optimized Transformation (HOT)-based cloud probability is calculated to replace the temperature probability for Sentinel-2 data.

HOT was originally proposed by Zhang et al. (2002) because of the high correlation between blue and red bands for most clear-sky land surfaces. It is calculated based on the blue and red TOA reflectance (Eq. (2)), and the original Fmask algorithm uses HOT to help to extract PCPs. Except for very bright land surfaces, such as urban and snow/ice, the higher the HOT values, the more likely they are clouds. Therefore, we normalized the HOT values based on the clear-sky land pixels to convert them into a HOT-based cloud probability ($IHOT$). Assuming the HOT values of clear-sky pixels differ within 0.04, Fmask 4.0 makes this normalization process similar as how the temperature probability is calculated based on the thermal band (Eq. (3)) (Zhu and Woodcock, 2012). Note the 17.5 and 82.5 percentiles are also inherited from the original Fmask and can provide the HOT interval for clear-sky land pixels (Zhu and Woodcock, 2012). This new cloud probability is only used for Sentinel-2 data and is particularly helpful to

capture the low altitude clouds.

$$HOT = \text{Band Blue} - 0.5 \cdot \text{Band Red} - 0.08 \quad (2)$$

$$IHOT = \frac{HOT - (HOT_{low} - 0.04)}{(HOT_{high} + 0.04) - (HOT_{low} - 0.04)} \quad (3)$$

where,

$$(HOT_{low}, HOT_{high}) = (17.5, 82.5) \text{ percentile of clear - sky land pixels' HOT.}$$

3.2.3. Combination of cloud probabilities

Fmask 4.0 calculates final cloud probability by combining different probabilities for Landsats 4–7, Landsat 8, and Sentinel-2 respectively (Table 2). For Landsats 4–7, Fmask 4.0 computes the combined cloud probability for land and water the same as the previous version of Fmask algorithm (Zhu and Woodcock, 2012), besides the thermal band will be normalized based on DEMs (Qiu et al., 2017). For Landsat 8, Fmask 4.0 adds a weighted (0.3) cirrus cloud probability to the same cloud probabilities designed for Landsats 4–7. For Sentinel-2, Fmask 4.0 uses the HOT-based cloud probability instead of the temperature probability for land areas plus another weighted (0.5) cirrus cloud probability. For water areas, the absence of temperature probability has limited effects on the cloud detection accuracy. Therefore, Fmask 4.0 computes the combined cloud probability over water by adding the brightness probability with the same weighted cirrus cloud probability for Sentinel-2 data.

In Fmask 4.0, we reduced the weight of cirrus cloud probability for Landsat 8 and Sentinel-2 due to the fact that some non-cloud objects, such as haze, fog, aerosol, dry vegetation, snow, and desert, can also have high cirrus band TOA reflectance especially in dry environments (Bouffies et al., 1997; Gao et al., 1993; Gao and Kaufman, 2003). As Fig. 8 shows, the full inclusion of the cirrus cloud probability (weighted 100%) increased the cloud probability for cloud pixels considerably (especially for high cirrus clouds) (See right dashed rectangle in Fig. 8b), but also influenced the cloud probability for clear-sky surfaces (See left dashed rectangle in Fig. 8b). Our goal is to separate the clouds and clear-sky surfaces based on the cloud probability layer, and the less the overlap between the distribution of cloud and clear-sky surface in cloud probability layer, the better the separability of the cloud probability. It is obvious that if we give 100% weight in cirrus cloud probability, higher overlap is expected, and less separability is provided in the cloud probability layer. Thus, we adjusted the weight of cirrus cloud probability and chose the one with the minimum overlap using the 90 reference images (Fig. 9). Based on the sensitivity analysis, the optimal weight of cirrus cloud probability is 0.3 for Landsat 8 and 0.5 for Sentinel-2. Note that due to the lack of a thermal band, Sentinel-2 has a higher weight than Landsat 8 (relies on cirrus band more).

Table 2

Computations of cloud probability in Fmask 4.0 for Landsats 4–7, Landsat 8, and Sentinel-2.

Data type	Cloud probability	
	Land	Water
Landsats 4–7	$IVar \cdot lTemp$	$wTemp \cdot wBright$
Landsat 8	$IVar \cdot lTemp + 0.3 \cdot Cir$	$wTemp \cdot wBright + 0.3 \cdot Cir$
Sentinel-2	$IVar \cdot IHOT + 0.5 \cdot Cir$	$wBright + 0.5 \cdot Cir$

Note that $IVar$ presents spectral variability probability for land (Section 3.2.1); $lTemp$ presents temperature probability for land (Qiu et al., 2017); $IHOT$ presents HOT-based cloud probability for land (Section 3.2.2); $wTemp$ presents temperature probability for water (Zhu and Woodcock, 2012); $wBright$ presents brightness probability for water (Zhu and Woodcock, 2012); Cir presents cirrus cloud probability (Zhu et al., 2015).

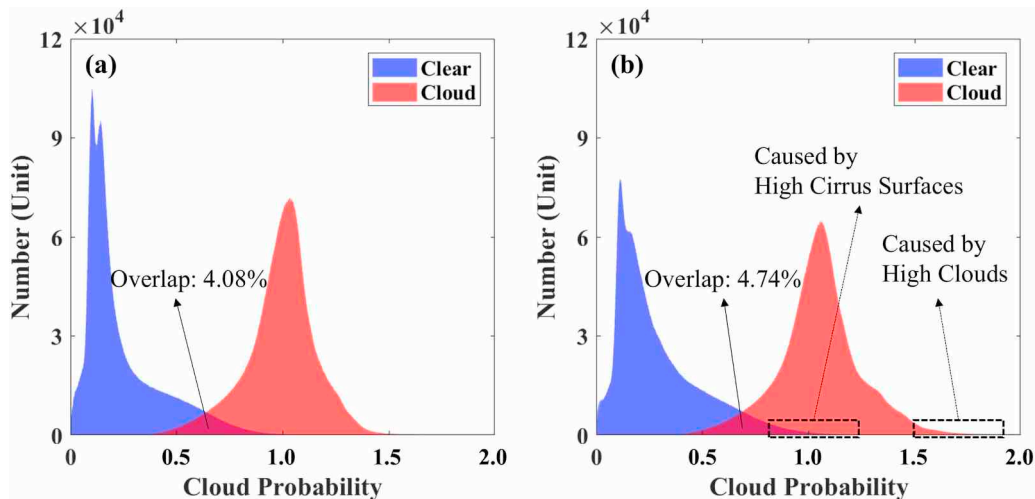


Fig. 8. Distributions of cloud probability without cirrus cloud probability (Fig. 8a) and with 100% cirrus cloud probability (Fig. 8b) based on manually interpreted clouds for a Landsat 8 image acquired on May 21st, 2014 at Path 229/Row 57.

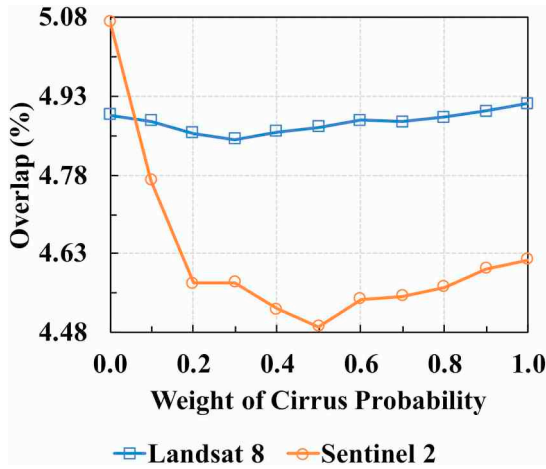


Fig. 9. Selection of the cirrus cloud probability weight for Landsat 8 and Sentinel-2 images based on 90 reference images. Note that the interval is 0.1 and the minimum overlap is achieved at the weight of 0.3 for Landsat 8 and 0.5 for Sentinel-2.

3.2.4. New cloud probability thresholds

The previous version of Fmask extracts clouds from the cloud probability layer using an image-based threshold that is the upper level (82.5 percentile) of clear-sky pixels' probability plus a constant of 0.2 (Zhu and Woodcock, 2012). The 0.2 was originally designed for Landsats 4–7. However, the cloud probability in Fmask 4.0 has already changed by applying BT normalization, cirrus normalization, HOT-based cloud probability, and weighted cirrus cloud probability. Therefore, we re-calibrated cloud probability thresholds for Fmask 4.0 based on the sensitivity analysis against the 90 reference images (Fig. 10). We found the optimal threshold (highest overall accuracy) for Landsats 4–7 is 0.1, for Landsat 8 is 0.175, and for Sentinel-2 is 0.2, and they will replace the constant value of 0.2 in the original Fmask algorithm.

3.3. Spectral-contextual features

Snow/ice and urban/built-up are easily confused with clouds because of their similar spectral signatures. Here, we propose to combine spectral and contextual dimensions to develop a snow/cloud separation index to better determine snow/ice pixels (Section 3.3.1), and apply a morphology method to remove the remaining commission errors induced by urban/built-up and mountain snow/ice (Section 3.3.2).

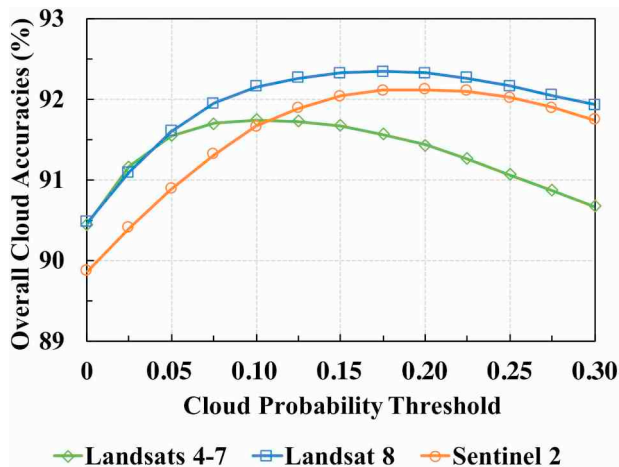


Fig. 10. Selection of optimal thresholds for Landsats 4–7, Landsat 8, and Sentinel-2 images based on 90 reference images.

3.3.1. Separate snow/ice from clouds (in polar regions)

Fmask selects PCPs based on several spectral tests and uses the rest of non-PCPs to calculate cloud probability, in which snow/ice pixels are excluded using a fixed NDSI threshold (0.8) (Zhu and Woodcock, 2012). This process tends to pick most of the snow/ice pixels as PCP since the NDSI threshold to distinguish clouds and snow/ice would be variable for the different conditions (e.g., atmospheric conditions and grain size of snow) (Choi and Bindenschadler, 2004). It will cause the rest of non-PCPs not representative of the land surface characteristics, especially for snow-covered images in polar regions. What is more, the NDSI values of clouds and snow/ice can have a large overlap in these areas, which makes them spectrally inseparable.

Fortunately, clouds and snow/ice are usually quite different in texture, in which places that covered by large areas of snow/ice are smoother than places covered by clouds (Coakley and Bretherton, 1982; Six et al., 2004). As the green band is less sensitive to influences from atmospheric conditions and grain size of snow/ice (Dozier, 1984, 1989), we can compute a Spectral-Contextual Snow Index (SCSI) by integrating the local Standard Deviation (SD) of green band and NDSI (Eq. (4)). SCSI can distinguish homogenous snow/ice from clouds because they usually have higher NDSI and smoother texture than clouds. A pixel would be more likely to be snow/ice when its SCSI is close to 0.

$$SCSI_n = SD_n(\text{Band Green}) \cdot (1 - \text{NDSI}) \quad (4)$$

where, n represents the local window size in kilometers.

To determine the optimal window size of SCSI, we calculated SCSI distributions of clouds and snow/ice for 17 reference images only covered by snow/ice and clouds (Fig. 4; Table S1). We can calculate SCSI distribution of clouds and snow/ice for these reference images based on different window sizes of SCSI, and the best window size should correspond to the least overlap. Fig. 11 shows that the overlap decreases continuously along with the increment of window size. This means that a larger window can result in a better ability to distinguish between clouds and snow/ice, but this also leads to more computation. To balance efficiency and accuracy, we selected a window size of 10 km ($SCSI_{10}$) for Fmask 4.0.

A pixel is classified as snow/ice if its $SCSI_{10}$ is less than 0.0009, and will be used as non-PCPs for subsequent statistical analyses (e.g., cloud probability calculation). The threshold of 0.0009 was determined by a sensitivity analysis, where we let $SCSI_{10}$ vary from 0 to 0.002 (at 0.00005 intervals) and examine how many snow/ice and cloud pixels can be correctly identified based on the same 17 snow/ice and cloud masks (Fig. 12). It is clear that when $SCSI_{10}$ is less than 0.0009, almost

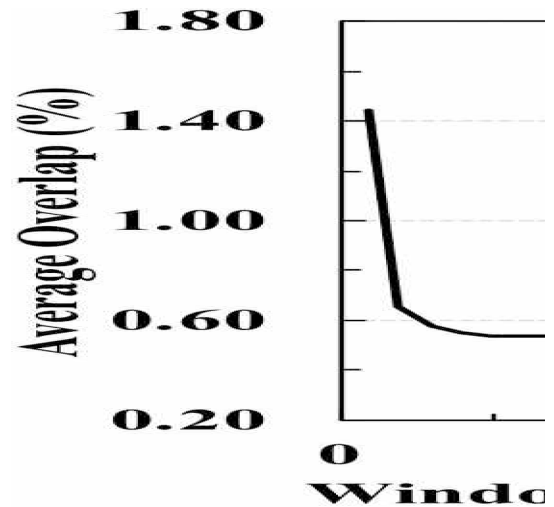


Fig. 11. Window size selection of SCSI based on 17 snow/ice and cloud covered Landsat 8 images. Note that a window size of 10 km (333 pixels for Landsat and 501 pixels for Sentinel-2) appears to be good enough to capture the variation of most clouds. SCSI: Spectral-Contextual Snow Index.

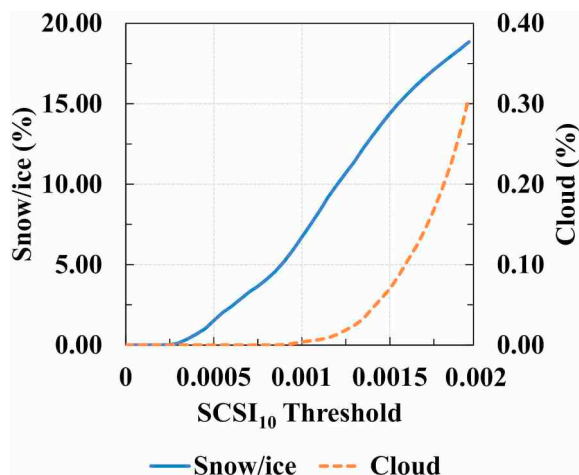


Fig. 12. Selection of $SCSI_{10}$ threshold based on 17 Landsat 8 reference images covered snow/ice and clouds. Note that a threshold of 0.0009 can correctly select approximate 5% snow/ice but almost none clouds.

all pixels are snow, and it is rare to include any cloud pixels. Note that this new $SCSI$ works well for places that covered by large areas of smooth snow/ice (common in polar regions) and may have limited

separability for mountain snow/ice or snow-covered forest areas.

3.3.2. Separate clouds from bright surfaces (urban/built-up and snow/ice)

This section is to reduce false positive errors caused by the remaining bright surfaces, such as urban/built-up and mountain snow/ice, in Fmask cloud detection results. Since urban/built-up and snow/ice have their unique spectral features, we can use simple spectral tests to identify all Potential False Positive Cloud Pixels (PFPCPs) that include these misidentified bright surfaces. At the same time, these bright surfaces normally show isolation, line, rectangle, and sharp shapes, which have an extremely large perimeter-to-area ratio, but clouds are usually the opposite. Based on this spectral-contextual difference, we can use a morphology approach to separate clouds from bright surfaces.

Extraction of potential false positive cloud pixels

Urban/built-up presents a unique spectral response that the reflectance in SWIR band is higher than NIR band (Zha et al., 2003), but clouds are the opposite (Irish, 2000). We can extract urban/built-up pixels by combining $NDBI > 0$ and $NDBI > NDVI$ (Zha et al., 2003), in which $NDVI$ is to exclude the scenarios that the vegetations with decreased leaf water content may have higher reflectance in SWIR than in NIR band (Cibula et al., 1992; Xu, 2007). Later, we used a line enhancement operator for the $NDBI$ layer (line-enhanced $NDBI$) to further highlight the urban/built-up area and depress the non-urban/built-up, as urban/built-up area contains lots of linear structures (e.g., roads)

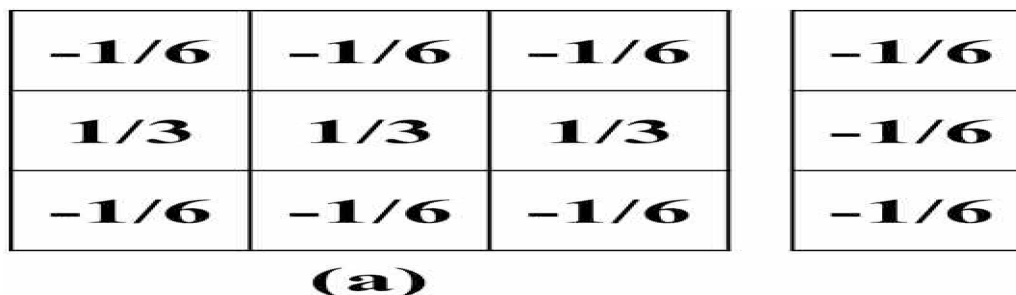


Fig. 13. Four line-enhancement kernels which respond maximally to horizontal, vertical, and oblique ($+45$ and -45 degrees) lines.

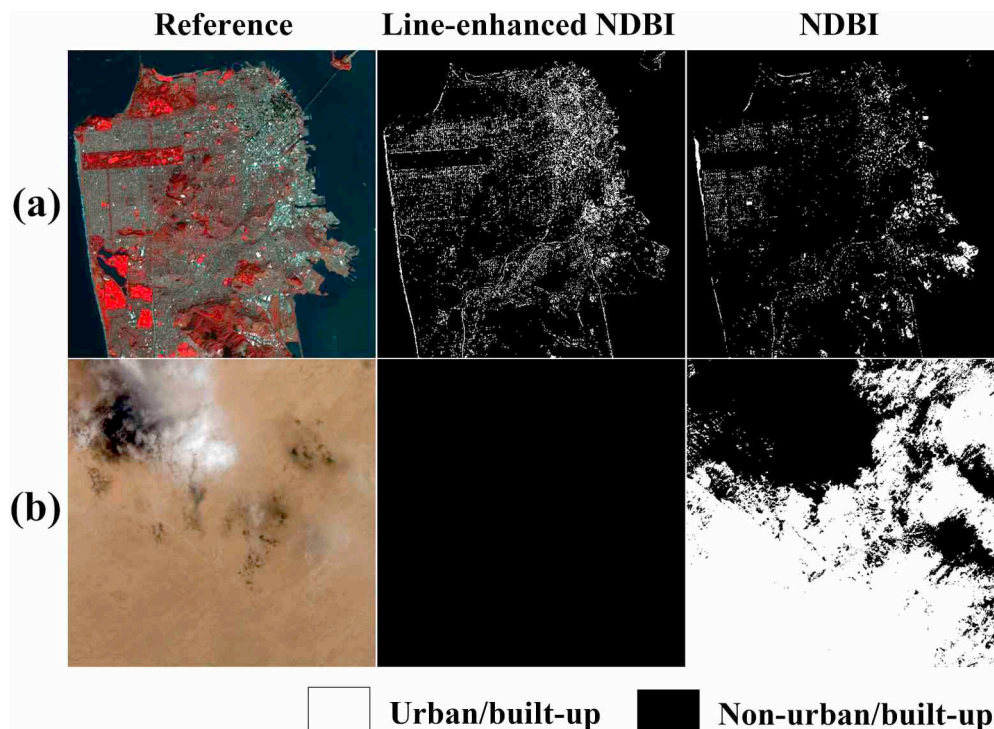


Fig. 14. Comparisons of line-enhanced $NDBI$ and $NDBI$ to extract urban/built-up. (a) Landsat 8 image located in San Francisco, U.S. acquired on Nov. 5th, 2017 at Path 44/Row 34 ($15 \times 15 \text{ km}^2$). (b) Landsat 8 image located in the Nubian Desert, Sudan acquired on Aug. 28th, 2017 at Path 178/Row 46 ($15 \times 15 \text{ km}^2$). The reference images in a and b are shown with NIR, red, and green bands composited. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that usually appear as bright lines on a darker background (e.g., vegetations) (Guindon et al., 2004). This operator applies four templates corresponding to four directions (Fig. 13) to filter each pixel and retains the maximum response. In this way, pixels with large NDBI value, such as desert, will be removed, but urban/built-up features (especial for roads) remained. Fig. 14 illustrates the results for two subsets of Landsat 8 images acquired from urban/built-up and desert areas, indicating the line-enhanced NDBI can better extract the urban/built-up pixels and avoid non-urban/built-up (e.g., desert) at the same time.

The line-enhanced NDBI can capture most of the urban/built-up pixels, but may also include some cloud pixels. As clouds are generally colder than pixels (Landsberg, 1981), we used an image-based threshold in the thermal bands for all Landsats 4–8 images to further exclude these cloud pixels. The threshold can be automatically determined using Otsu's method (Otsu, 1979). For Sentinel-2, we excluded the cloud pixels based on a newly developed Cloud Displacement Index (CDI), which improves cloud detection based on the view angle parallax of three Sentinel-2 bands (Band 7, 8, and 8a) (Frantz et al., 2018). The CDI values of most clouds are less than -0.5 but considering densely packed buildings and slight across-track parallax can also have lower CDI values (Frantz et al., 2018), we decreased the CDI threshold to -0.8 to exclude the remaining cloud pixels for Sentinel-2. As commission errors are also found near urban/built-up areas, we dilated urban/built-up pixels with a window of 500 m in 8 connected directions.

The false positive clouds are also observed in areas of mountain snow/ice. The potential snow/ice pixels can be determined by following the initial Fmask test, in which mountain snow/ice pixels can be further determined when a pixel's slope is larger than 20 degrees, as most mountainous regions have slopes beyond this threshold (Burbank et al., 1996). As commission errors are also found near areas of mountain snow/ice, we dilated the mountain snow/ice pixels with a window of 500 m in 8 connected directions. Snow/ice pixels in flat areas (no dilation) are also included in the final PFPCPs.

A morphology-based elimination process

After identification of all PFPCPs, we will perform a morphology process for them that can efficiently eliminate (or at least reduce) the bright surfaces characterized by large perimeter-to-area ratio. The fundamental operators in morphology are erosion and dilation, which are related directly to object shape (Haralick et al., 1987). As urban/built-up and mountain snow/ice are usually composed of isolated pixels, lines of pixels, and rectangles of pixels, they will be removed by erosion with a disk-shaped structuring element (Fig. 15). This type of erosion generally cannot remove entire clouds because of their low perimeter-to-area ratio shape and relatively large size, but this will change cloud boundaries. To recover the cloud shape, Fmask 4.0 dilates the remaining pixels with the same structure element twice (Fig. 15). A single dilation often may recover the original shape of many “round” clouds, but may not work for clouds with sharp boundary. Therefore, Fmask 4.0 makes another dilation to recover

original cloud shape as much as possible. The dilated pixels cannot exceed the initial extent of clouds as we do not want to increase commission errors. Based on the sensitivity analysis against 90 reference images, the optimal radius of the structuring element is 150 m (5 pixels) for Landsats 4–7 and 90 m (3 pixels) for Landsat 8, and 90 m (5 pixels) for Sentinel-2 (same with Landsat 8) (Fig. 16). Note that the radius used for Landsats 4–7 is larger than Landsat 8 mainly because of the more commission errors caused by the lower cloud probability threshold and the lack of the cirrus band. This morphology-based process can eliminate most of the false positive cloud errors in Landsat data, but less helpful for Sentinel-2 data mainly because the lack of thermal band would result in more commission errors of clouds from the bright surfaces (Fig. 3). Therefore, for Sentinel-2 images, Fmask 4.0 further removes small objects ($< 10,000$ pixels) in PFPCPs based on the minimum CDI value in each object, and if its minimum CDI is larger than -0.5 , it is excluded from the final cloud results.

4. Results

Like we have done previously (Qiu et al., 2017; Zhu et al., 2015; Zhu and Woodcock, 2012), we evaluate our cloud and cloud shadow results by dilating 3 pixels for cloud shadow pixels, but no dilation for cloud pixels.

4.1. Results for Landsats 4–8

We compared the results between Fmask 4.0 and Fmask 3.3. As the visual comparisons in Fig. 17 show, Fmask 4.0 performed better in

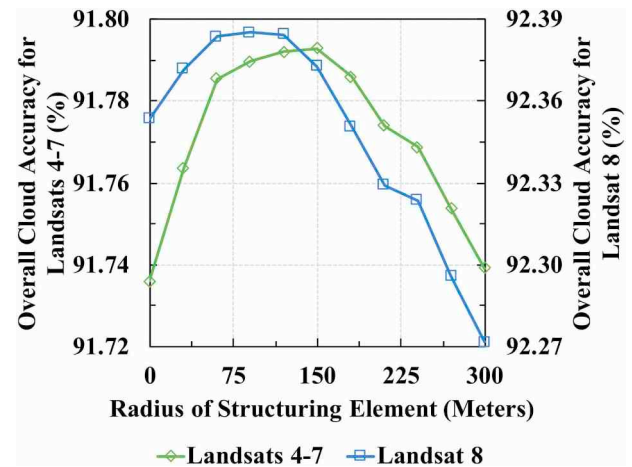


Fig. 16. Selection of the radius of structuring element. The radius of 150 m achieves the highest overall cloud accuracy for Landsats 4–7 and 90 m for Landsat 8 based on 90 reference images.

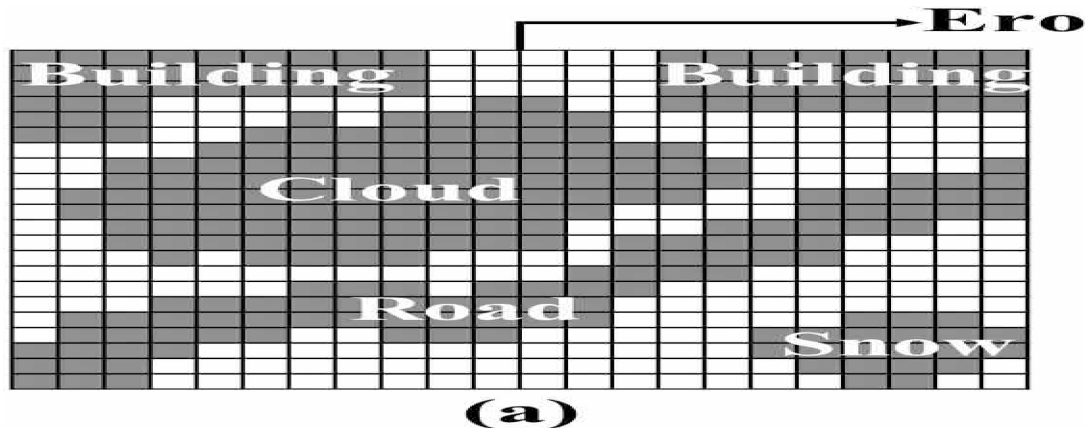


Fig. 15. Example of erosion and dilation operations with a radius of 3 pixels. (a) Potential false positive cloud layer. (b) Remaining pixels after erosion. (c) Cloud recovered by dilation.

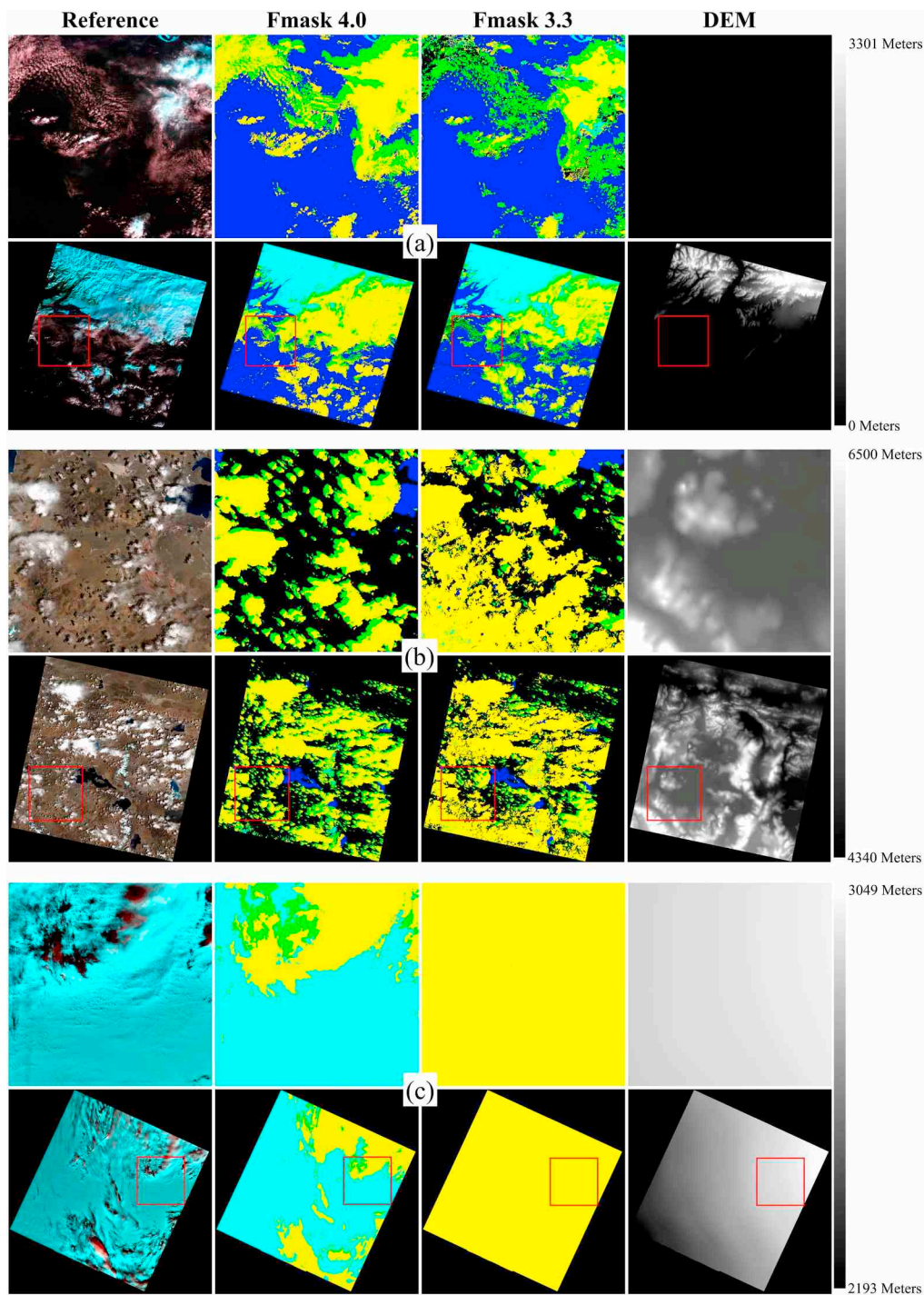


Fig. 17. Comparisons of Fmask 4.0 and Fmask 3.3 for five Landsat images. (a) Landsat 7 image located in Alaska Gulf, U.S. acquired on Mar. 28th, 2002 at Path 65/Row 18. (b) Landsat 8 image located in Tibetan Plateau, China acquired on May 16th, 2016 at Path 143/Row 38. (c) Landsat 8 image located in Greenland, Denmark acquired on May 27th, 2014 at Path 6/Row 10. (d) Landsat 8 image located in Mount Qomolangma, China acquired on Oct. 2nd, 2016 at Path 140/Row 40. (e) Landsat 8 image located in Los Angeles, U.S. acquired on Dec. 23rd, 2013 at Path 41/Row 37. The upper images in a, b, c, and d are enlargements of the lower entire images with a size of $60 \times 60 \text{ km}^2$. The upper images in e are enlargements of the lower entire images with a size of $15 \times 15 \text{ km}^2$. The reference images in a, c, and d are shown with SWIR, NIR, and red bands composited. The reference images in b and e are shown with NIR, red, and green bands composited. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

detecting clouds and cloud shadows than Fmask 3.3 did for Landsats 4–8 images. Fig. 17a is a Landsat 7 image located in Alaska Gulf where there are well-behaved clouds over a large area of water region. Fmask 4.0 successfully identified the clouds (especial thin clouds) over water that Fmask 3.3 missed. Fig. 17b shows the results of a Landsat 8 image located in Tibetan Plateau with low vegetation coverage and large topographic differences. Fmask 3.3 cloud mask showed large commission

errors in cloud detection and failed to identify the associated cloud shadows. Nevertheless, Fmask 4.0 detected the clouds and their shadows successfully. Fig. 17c is a Landsat 8 image acquired from Greenland where thick and thin clouds over very bright snow/ice sheets. Fmask 3.3 incorrectly identified the entire image as clouds, but Fmask 4.0 successfully separated clouds from the bright snow/ice. Fig. 17d is a Landsat 8 image acquired from Mount Qomolangma with large

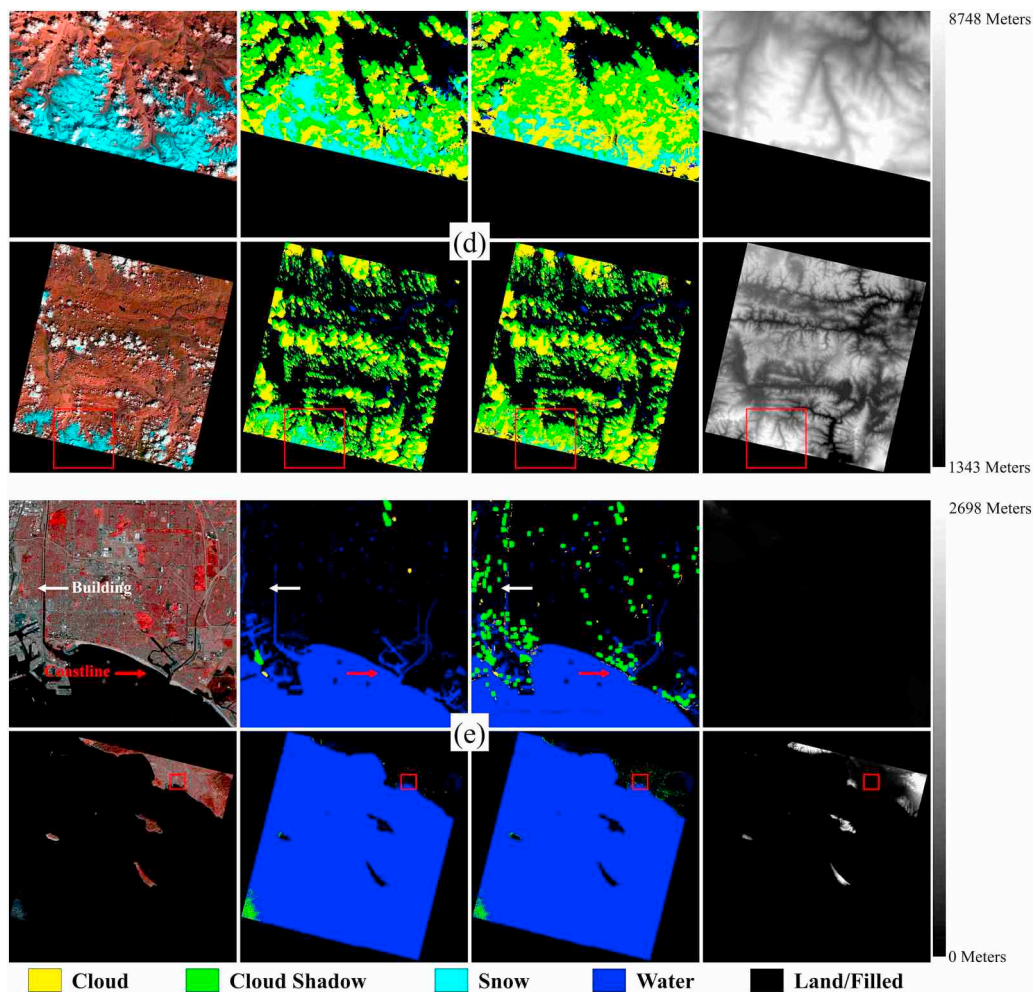


Fig. 17. (continued)

Table 3
Confusion matrices for assessing the accuracies of Fmask 4.0 and Fmask 3.3 for Landsats 4–7 images.

		Visual			
		Cloud	Cloud Shadow	Clear	User (%)
Fmask 4.0	Cloud	633	4	2	99.06
	Cloud Shadow	31	45	13	50.56
	Clear	55	13	757	91.76
	Producer (%)	88.04	72.58	98.06	92.40
Fmask 3.3	Cloud	607	0	1	99.84
	Cloud Shadow	37	42	11	46.67
	Clear	75	20	760	88.89
	Producer (%)	84.42	67.74	98.45	90.73

Table 4
Confusion matrices for assessing the accuracies of Fmask 4.0 and Fmask 3.3 for Landsats 8 images.

		Visual			
		Cloud	Cloud Shadow	Clear	User (%)
Fmask 4.0	Cloud	657	1	16	97.48
	Cloud Shadow	13	50	17	62.50
	Clear	16	21	761	95.36
	Producer (%)	95.77	69.44	95.84	94.59
Fmask 3.3	Cloud	652	3	18	96.88
	Cloud Shadow	15	40	20	53.33
	Clear	19	29	756	94.03
	Producer (%)	95.04	55.56	95.21	93.30

topographic gradients. For this image, Fmask 3.3 showed large false positive errors in cloud detection in areas covered by mountain snow/ice, while Fmask 4.0 significantly reduced this kind of commission errors. Fig. 17e is a Landsat 8 image acquired from Los Angeles (a coastal city) with areas of buildings. Compared to the results generated by Fmask 3.3, Fmask 4.0 did a good job reducing the false positive errors from urban/built-up (See the white arrow) and reduced the commission errors from the coastlines (See the red arrow).

By comparing with our validation samples quantitatively, the overall accuracy of Fmask 4.0 is higher than Fmask 3.3 (92.40% versus 90.73%) for Landsats 4–7. For cloud detection, the producer's accuracy of Fmask 4.0 is higher than Fmask 3.3 (88.04% versus 84.42%) (Table 3). The user's accuracy of Fmask 4.0 decreased slightly (0.78% lower than Fmask 3.3)

(Table 3), which suggests Fmask 4.0 may have more commission errors. For cloud shadow detection, Fmask 4.0 achieved higher producer's and user's accuracies than Fmask 3.3 (Table 3). Note that the user's accuracies for cloud shadow are relatively low mainly because of the 3 pixels' dilation. The improvements in Fmask 4.0 cloud shadow detection is mainly contributed by the enhancements of MFmask (Qiu et al., 2017).

For Landsat 8 images, Fmask 4.0 still shows higher accuracies than Fmask 3.3, in which overall accuracy is 1.29% higher (Table 4). For cloud detection, both producer's and user's accuracies of Fmask 4.0 (95.77% and 97.48%) are higher than Fmask 3.3 (95.04% and 96.88%). On the other hand, Fmask 4.0 achieved higher producer's (69.44%) and user's accuracies (62.50%) in cloud shadow detection than Fmask 3.3 (55.56% and 53.33%).

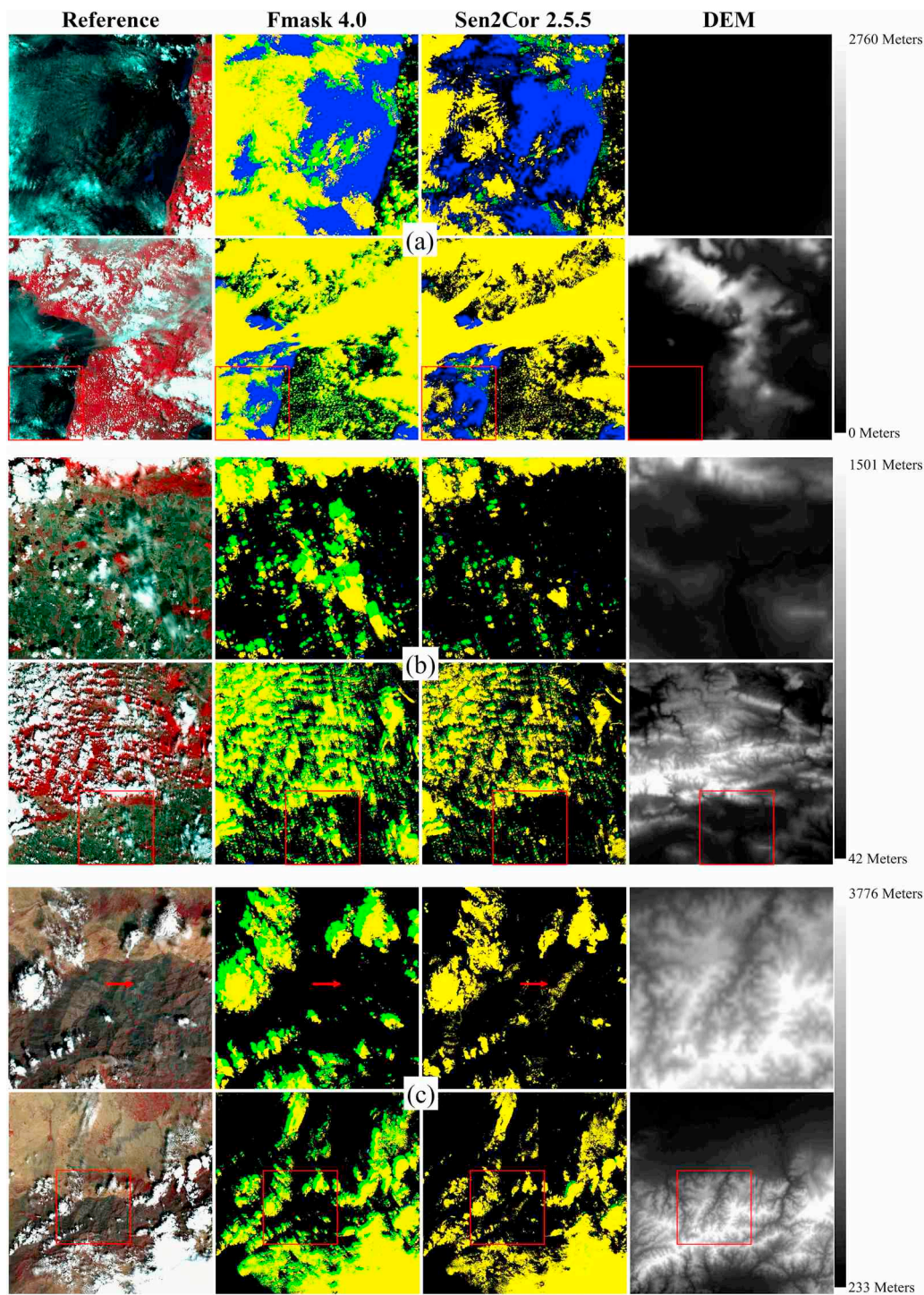


Fig. 18. Comparisons of Fmask 4.0 and Sen2Cor 2.5.5 for seven Sentinel-2 images. (a) Sentinel-2A image located in Boni Gulf, Indonesia acquired on Mar. 28th, 2017 at Tile 51MUR. (b) Sentinel-2A image located in Sliven, Bulgaria acquired on Sept. 23rd, 2017 at Tile 35TMH. (c) Sentinel-2A image located in Atlas Mountains, Morocco acquired on Aug. 25th, 2016 at Tile 29RNQ. (d) Sentinel-2B image located in Araouane, Mali acquired on Jan. 3rd, 2018 at Tile 30QUQ. (e) Sentinel-2A image located in Antarctica acquired on Mar. 9th, 2018 at Tile 16CEB. (f) Sentinel-2B image located in Longmen Mountains, China acquired on Apr. 20th, 2018 at Tile 48RUV. (g) Sentinel-2A image located in New York, U.S. acquired on Aug. 4th, 2017 at Tile 18TWL. The upper images in a, b, c, d, and e are enlargements of the lower entire images with a size of $40 \times 40 \text{ km}^2$. The upper images in f and g are enlargements of the lower entire images with a size of $20 \times 20 \text{ km}^2$. The reference images in a, b, c, d, and g are shown with NIR, red, and green bands composited. The reference images in e and f are shown with SWIR, NIR, and red bands composited. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Results for Sentinel-2

We compared the results from Fmask 4.0 with Sen2Cor 2.5.5. The Sen2Cor is a software for generating Sentinel-2 Level 2A product, which can provide categorical information of cloud, cloud shadow, water, and

snow at the pixel level (Main-Knorn et al., 2017). For the cloud category, Sen2Cor 2.5.5 provides subcategories, including medium probability cloud, high probability cloud, and thin cirrus cloud (Müller-Wilm et al., 2018). In this study, we combined them to generate one cloud mask and compare Sen2Cor 2.5.5 with Fmask 4.0 results. Visual

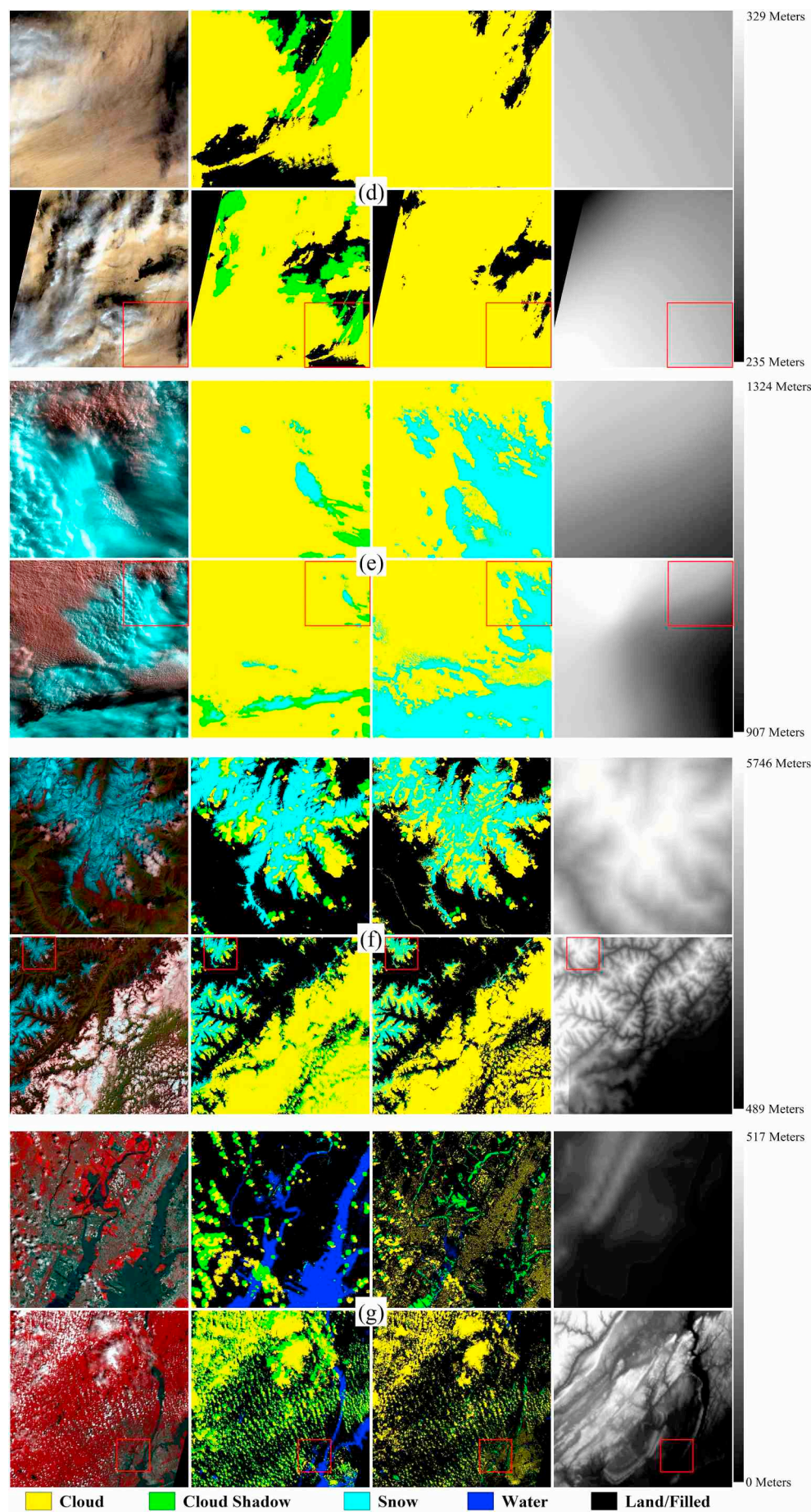


Fig. 18. (continued)
15

comparisons demonstrate that Fmask 4.0 works better in detecting clouds and cloud shadows than Sen2Cor 2.5.5 for Sentinel-2 images (Fig. 18). Fig. 18a is a Sentinel-2A image located in Boni Gulf where there are well-behaved clouds over a large area of water and highly vegetated areas. Both Fmask 4.0 and Sen2Cor 2.5.5 worked well in detecting clouds over highly vegetated surfaces, but Sen2Cor 2.5.5 missed many thin clouds over the water. Fig. 18b is a Sentinel-2A image located in Sliven where the land surfaces include vegetation and barren. Fmask 4.0 captured the clouds and cloud shadows missed by Sen2Cor 2.5.5. Fig. 18c is a Sentinel-2A image located in Atlas Mountains, which is characterized by large elevation differences and topographic reliefs. Both algorithms detected the clouds over low elevation surfaces, but Sen2Cor 2.5.5 incorrectly identified high mountains as clouds (See the red arrow). In contrast, Fmask 4.0 did not show this problem and detected cloud shadows as well. Fig. 18d is a Sentinel-2B image located in Araouane where lots of thin and thick clouds are over desert. Sen2Cor 2.5.5 failed to separate clouds from the bright desert, but Fmask 4.0 succeeded. Fig. 18e is a snow-covered Sentinel-2A image acquired from Antarctica, in which many cloud pixels are missed by Sen2cor 2.5.5 but identified by Fmask 4.0. Fig. 18f is a Sentinel-2B image located Longmen Mountains with large coverage of mountain snow/ice. Compared to Sen2Cor 2.5.5, Fmask 4.0 reduced those commission errors significantly. Fig. 18g is a Sentinel-2A image located in New York. Fmask 4.0 separated clouds from urban/built-up and identified their shadows as well, but Sen2Cor 2.5.5 has large commission errors of clouds and omission errors of cloud shadows.

By comparing with our reference samples, Fmask 4.0 shows higher accuracies for detecting clouds and cloud shadows than Sen2Cor 2.5.5 (Table 5). The overall accuracy of Fmask 4.0 is 94.30%, which is much higher than Sen2Cor 2.5.5 (87.10%). For cloud detection, the producer's and user's accuracies of Fmask 4.0 (93.65% and 97.27%) are higher than Sen2Cor 2.5.5 (83.43% and 93.35%). For cloud shadow detection, the producer's accuracy of Fmask 4.0 reached 69.64%, which is also higher than Sen2Cor 2.5.5 (39.29%). Note that Sen2Cor 2.5.5 has higher user's accuracy (73.33%).

Table 5

Confusion matrices for assessing the accuracies of Fmask 4.0 and Sen2Cor 2.5.5 for Sentinel-2.

		Visual			
		Cloud	Cloud Shadow	Clear	User (%)
Fmask 4.0	Cloud	678	2	17	97.27
	Cloud Shadow	16	39	7	62.90
	Clear	30	15	723	94.14
	Producer (%)	93.65	69.64	96.79	94.30
Sen2Cor 2.5.5	Cloud	604	4	39	93.35
	Cloud Shadow	4	22	4	73.33
	Clear	116	30	704	82.82
	Producer (%)	83.43	39.29	94.24	87.10

5. Discussions and conclusion

Auxiliary data (e.g., GSWO, DEM, slope, and aspect) are beneficial for improving cloud and cloud shadow detection. For example, without the GSWO layer, cloud probability for pixels covered by thick clouds will be calculated based on approaches designed for cloud detection over land areas, and this will lead to low cloud probability for clouds over water areas, especially when there are bright and cold land surfaces (e.g., snow/ice) (Fig. 19b). If clouds over water are correctly classified based on auxiliary data provided by GSWO, their cloud probabilities will be more reasonable (Fig. 19c). Also, DEM and its derivatives (slope and aspect) are critical for improving cloud and cloud shadow detection, especial for places with large topographic gradient (Qiu et al., 2017). Generally, we recommend using auxiliary data with the same spatial resolution as the satellite data, but we found DEM with 1-kilometer resolution, and land/water mask with 150-meter resolution can generate results of similar accuracies (Fig. 7). To make Fmask 4.0 easy to use, we have packaged global GTOPO30 (1 km) and GSWO (150 m) online (<https://github.com/gersl/fmask>).

Cloud detection in Sentinel-2 data has been quite challenging due to the lack of a thermal band. Though the use of cirrus band can alleviate this issue (Zhu et al., 2015), it is still problematic for low altitude clouds, which are almost invisible in cirrus band (Fig. 20b). With the inclusion of new HOT-based cloud probability, we can capture these cloud types again (Fig. 20c). Comparison between results with and without the new HOT-based cloud probability (Fig. 20c and d) indicates this new cloud probability is exceptionally effective for sensors without a thermal band. Note that the current Fmask 4.0 designed for Sentinel-2 images is developed and calibrated based on training images simulated from Landsat 8 images, and considering the spectral differences between Landsat 8 and Sentinel-2 (Drusch et al., 2012), training images from real Sentinel-2 images are needed for better calibrate the algorithm.

Fmask 4.0 defines new default cloud probability thresholds for Landsats 4–7 (0.1), Landsat 8 (0.175), and Sentinel-2 (0.2) based on the 90 reference images (Fig. 10). Note that Landsat 8 and Sentinel-2 have similar thresholds, but Landsats 4–7 have a relatively lower threshold. This is mainly because both Landsat 8 and Sentinel-2 has the cirrus band to help with detection of thin cirrus clouds, but Landsats 4–7 do not have this band, and in order to detect those thin cirrus clouds, a lower threshold is used. Additionally, those thresholds are global optima but not regional optima. Users are welcome to adjust the threshold to produce the best cloud masks for their study areas. Based on the preference of commission or omission errors, the user can decrease or increase the cloud probability threshold as demonstrated in Fig. 21.

Integration of spectral and contextual information is the key approach for separating clouds from other bright surfaces. By combining NDSI and the standard deviation of green band, the new SCSI index can identify 5% snow/ice pixels, which are enough to present the clear-sky surface and provide more accurate cloud probability estimation

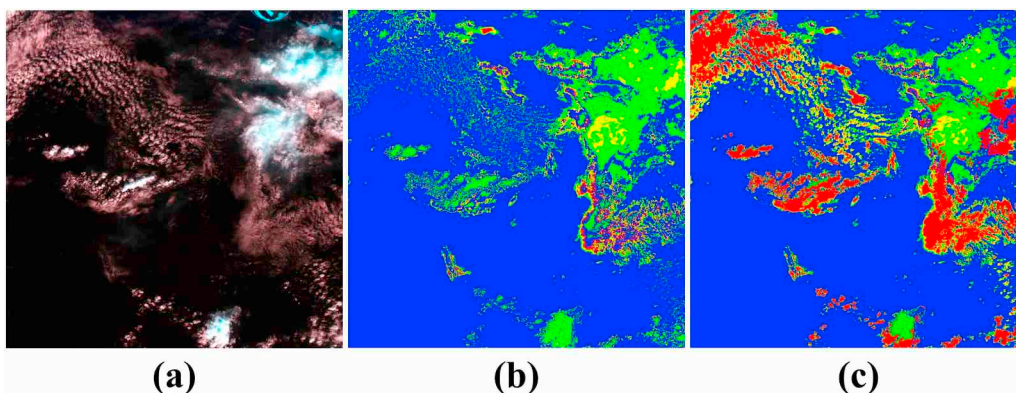


Fig. 19. Cloud probability over water (Same image in Fig. 17a). (a) False color composite image (SWIR, NIR, and red bands). (b) Density slice of the cloud probability over water generated without GSWO data. (c) Density slice of the cloud probability over water generated with GSWO data. Both of the two density slices were created by the same color ranges ($-\infty < \text{blue} < 0.5$; $0.5 < \text{green} < 1$; $1 < \text{yellow} < 1.5$; $1.5 < \text{red} < +\infty$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

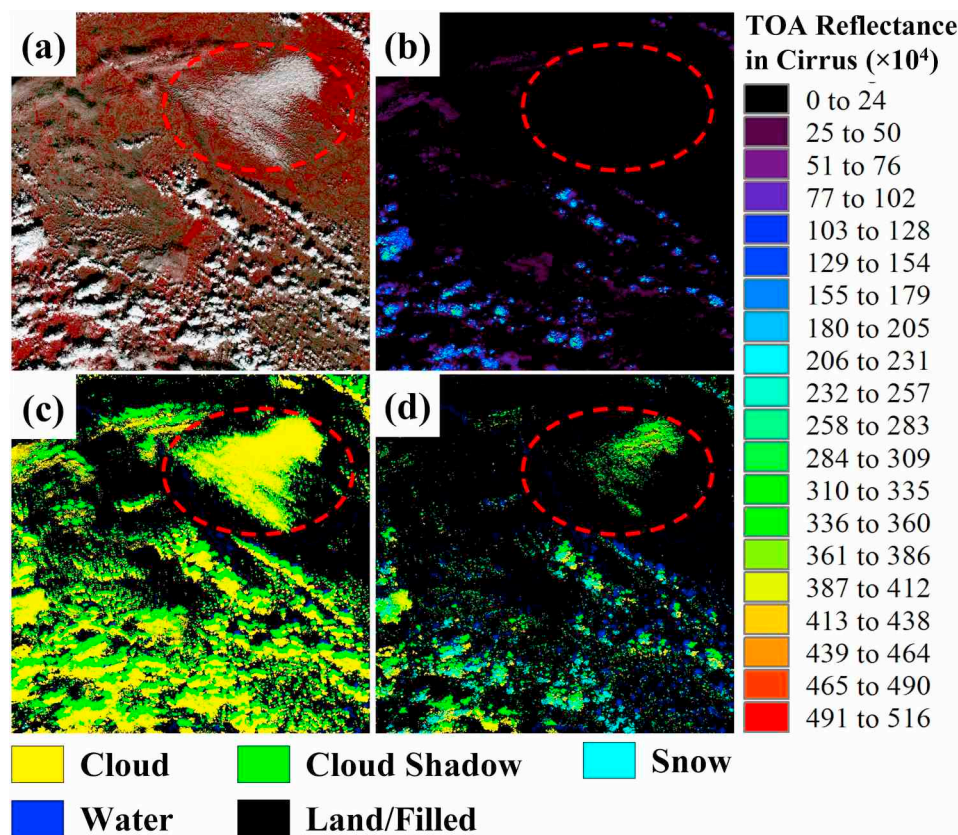


Fig. 20. Results for a Sentinel-2A image acquired on Oct. 18th, 2016 at Tile 31TCJ. (a) False color composite image (NIR, red, and green bands). (b) TOA reflectance at cirrus band. (c) Results derived from Fmask 4.0 with HOT-based cloud probability. (d) Results derived from Fmask 4.0 without HOT-based cloud probability. Note that the red circle indicates low clouds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 12). Meanwhile, a line-enhanced NDBI and morphological operators (erosion and dilation) can remove the remaining bright surfaces from cloud pixels, such as urban/built-up and mountain snow/ice. Unique for Sentinel-2 data, a 3D spectral-contextual-based index called CDI can be used to separate clouds from these bright surfaces (Frantz et al., 2018). This is particularly important, as Sentinel-2 cloud detection is seriously influenced by the lack of a thermal band. Finally, we found that the spectral-contextual processes (mainly the erosion operator) may also remove some small clouds and result in some omission errors. We think this kind of omission errors are less serious than persistent commission errors, which are quite common for these bright surfaces, especial for time series analyses (Fig. 22). This is because most of them can tolerate a few noisy pixels, but none of them will be able to work with persistent missing values (Dong et al., 2015; Huang et al., 2010a; Veh et al., 2018).

The accurate detection of cloud shadows is challenging because lots of dark surfaces (e.g., wetland, burned area, and terrain shadow) are easily confused with cloud shadows. Fmask detects cloud shadows

mainly based on the cloud-shadow projection geometry, which is beneficial to separate clouds from those spectrally dark surfaces. However, we still found some omission errors in cloud shadow detection (Tables 3-5). The main reason is that cloud shadow shape is calculated based on cloud object, which is inherently impacted by cloud detection results. Considering the cloud shadow size is usually a lot smaller than the cloud size, the omission of cloud shadows is generally less harmful than omission of clouds.

The accuracies were assessed against random samples from 16 locations with different land cover types (Fig. 4). To provide a rigorous global validation of cloud and cloud shadow detection results, more random samples around the world (random both in the spatial and temporal domain) should be created in the future. Considering that the original Fmask algorithm has achieved relatively high accuracy (Foga et al., 2017), the 1–2% higher overall accuracy of Fmask 4.0 is valuable for Landsats 4–8 data. Qiu et al. (2019) also applied the Fmask 4.0 algorithm for creating clear-sky Landsat Time Series (LTS) data and found it can make the LTS more consistent. As for Sentinel-2, we

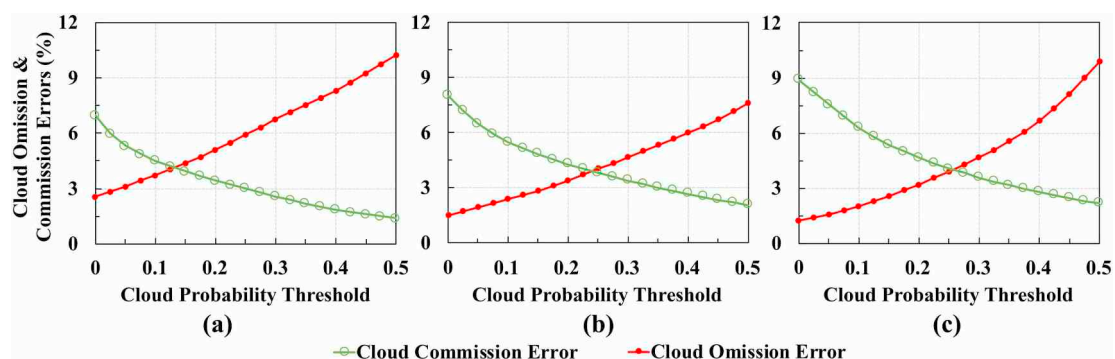


Fig. 21. The relationship between cloud probability thresholds and accuracies (omission errors and commission errors) for (a) Landsats 4–7, (b) Landsat 8, and (c) Sentinel-2 images based on 90 reference images.

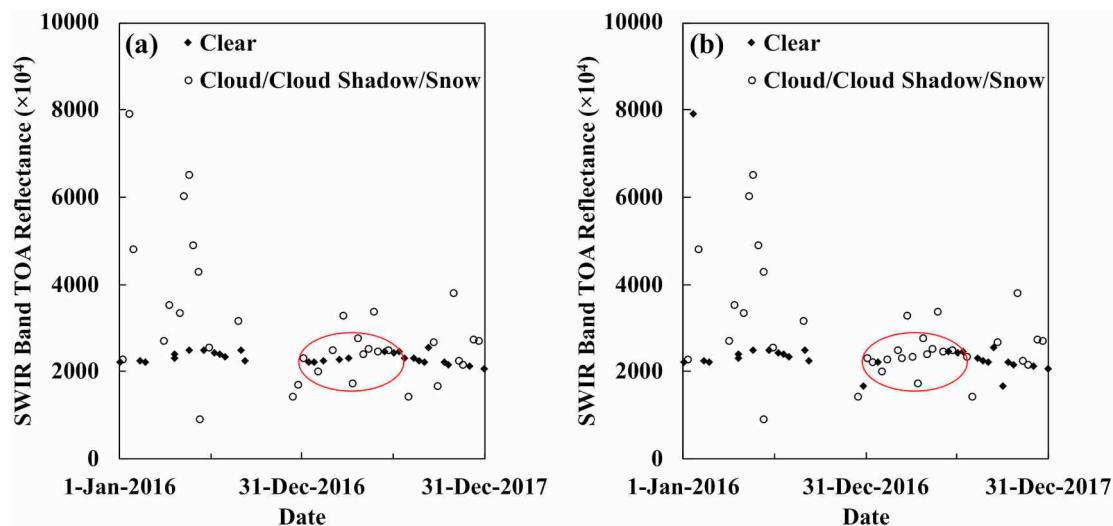


Fig. 22. Comparison between Fmask 4.0 (a) and Fmask 3.3 (b) for times series of SWIR Band TOA reflectance of a Sentinel-2 pixel located in the red arrow in Fig. 3. Red circles indicate the period where Fmask 4.0 identified the clear pixels while Fmask 3.3 misidentified them as clouds persistently. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

observed that Fmask 4.0 achieved 7% higher overall accuracy than Sen2Cor 2.5.5. Recently, the study comparing cloud detection algorithms demonstrated that this Fmask 4.0 algorithm again achieved 6% higher accuracy than Sen2Cor 2.5.5 and even comparable to multi-temporal approach such as MAJA (MACCS-ATCOR Joint Algorithm) (Baetens et al., 2019). However, Fmask 4.0 still has limitations compared to Sen2Cor 2.5.5. For example, Fmask 4.0 processes the Sentinel-2 image at 20-meter resolution, and subsequently this would result in missing very small clouds and their cloud shadows. At the same time, we also found that Sen2Cor 2.5.5 has fewer commission errors than Fmask 4.0 in cloud shadow detection (Table 5). In the future, combination of different cloud and cloud shadow results (e.g., Fmask and Sen2Cor) may be a better solution.

Although Fmask 4.0 is more complicated than Fmask 3.3, it has the similar efficiency as Fmask 3.3. Generally, the time required for Fmask 4.0 software to process one Landsat or Sentinel-2 image generally is between 1.5 and 10 min (depend on cloud cover percentage and land surface characteristics) on a 2.7 GHz computer with an Intel Core i7 processor and 16 GB of RAM. This Fmask package is publically available at <https://github.com/gersl/fmask>.

In conclusion, we proposed a new Fmask algorithm (4.0 version) that substantially improves cloud and cloud shadow detection results for Landsats 4–8 and Sentinel-2 images by integrating auxiliary data, new cloud probabilities, and novel spectral-contextual features.

Acknowledgments

We gratefully acknowledge the support of National Key Research and Development (R&D) Program of China (#2018YFD0200301), National Natural Science Foundation of China (#41671361), and Fundamental Research Fund for the Central Universities (#ZYGX2012Z005) to Binbin He. We also gratefully acknowledge the support of USGS-NASA Landsat Science Team (LST) Program for Toward Near Real-time Monitoring and Characterization of Land Surface Change for the Conterminous US (#G17PS00256) to Zhe Zhu. Shi Qiu specially thanks China Scholarship Council (#201706070113). We gratefully thank Yanru Zhou for preparing auxiliary data, and also acknowledge the CDI-based Fmask code from Frantz David and the feedbacks from Curtis Woodcock, Karen C. Seto, Yunkun Lin and other enthusiast researchers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2019.05.024>.

References

- Akbari, H., Pomerantz, M., Taha, H., 2001. Cool surfaces and shade trees to reduce energy use and improve air quality in urban areas. *Sol. Energy* 70, 295–310.
- Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sens.* 11, 433. <https://doi.org/10.3390/rs11040433>.
- Bouffies, S., Breon, F.M., Tanre, D., Dubuisson, P., 1997. Atmospheric water vapor estimate by a differential absorption technique with the polarisation and directionality of the Earth reflectances (POLDER) instrument. *J. Geophys. Res.* 102, 3831–3841.
- Burbank, D.W., Leland, J., Fielding, E., Anderson, R.S., Reid, M.R., Duncan, C., 1996. Bedrock incision, rock uplift and threshold hillslopes in the NW Himalaya. *Nature* 379, 505–510.
- 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, 237–242. <https://doi.org/10.1016/j.rse.2004.03.007>.
- Cibula, W.G., Zetka, E.F., Rickman, D.L., 1992. Response of thematic mapper bands to leaf water stress. *Int. J. Remote Sens.* 13, 1869–1890.
- Claverie, M., Masek, J., 2017. Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide. Web site: Authors. Natl. Aeronaut. Sp. Adm, Washington, DC, USA 2. <https://doi.org/10.13140/RG.2.2.33017.26725>.
- Coakley, J.A., Bretherton, F.P., 1982. Cloud cover from high resolution scanner data: detecting and allowing for partial fields of view. *J. Geophys. Res.* 87, 4917–4932.
- Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L., Jin, C., Zhou, Y., Wang, J., Biradar, C., Liu, J., Moore, B., 2015. Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. *Remote Sens. Environ.* 160, 99–113. <https://doi.org/10.1016/j.rse.2015.01.004>.
- Dozier, J., 1984. Snow reflectance from LANDSAT-4 thematic mapper. *IEEE Trans. Geosci. Remote Sens.* GE-22, 323–328. <https://doi.org/10.1109/TGRS.1984.350628>.
- Dozier, J., 1989. Spectral signature of alpine snow cover from the Landsat thematic mapper. *Remote Sens. Environ.* 28, 9–22. [https://doi.org/10.1016/0034-4257\(89\)90101-6](https://doi.org/10.1016/0034-4257(89)90101-6).
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>.
- Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., Schmidt, G.L., Dwyer, J.L., Joseph Hughes, M., Laue, B., 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* 194, 379–390. <https://doi.org/10.1016/j.rse.2017.03.026>.
- 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, 1242–1246. <https://doi.org/10.1109/LGRS.2015.2390673>.
- Frantz, D., Haß, E., Uhl, A., Stoffels, J., Hill, J., 2018. Improvement of the Fmask

- algorithm for Sentinel-2 images: separating clouds from bright surfaces based on parallax effects. *Remote Sens. Environ.* 215, 471–481.
- Gao, B.-C., Kaufman, Y.J., 1995. Selection of the 1.375- μ m MODIS channel for remote sensing of cirrus clouds and stratospheric aerosols from space. *J. Atmos. Sci.* 52, 4231–4237. [https://doi.org/10.1175/1520-0469\(1995\)052<4231:SOTMCF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<4231:SOTMCF>2.0.CO;2).
- Gao, B.-C., Kaufman, Y.J., 2003. Water vapor retrievals using moderate resolution imaging spectroradiometer (MODIS) near-infrared channels. *J. Geophys. Res. Atmos.* 108. [https://doi.org/10.1029/2002JD003023108\(D13\)](https://doi.org/10.1029/2002JD003023108(D13)).
- Gao, B.-C., Goetz, A.F.H., Wiscombe, W.J., 1993. Cirrus cloud detection from airborne imaging spectrometer data using the 1.38 μ m water vapor band. *Geophys. Res. Lett.* 20, 301–304. <https://doi.org/10.1029/93GL00106>.
- Goodwin, N.R., Collett, L.J., Denham, R.J., Flood, N., Tindall, D., 2013. Cloud and cloud shadow screening across Queensland, Australia: an automated method for Landsat TM/ETM+ time series. *Remote Sens. Environ.* 134, 50–65. <https://doi.org/10.1016/j.rse.2013.02.019>.
- Guindon, B., Zhang, Y., Dillabaugh, C., 2004. Landsat urban mapping based on a combined spectral-spatial methodology. *Remote Sens. Environ.* 92, 218–232. <https://doi.org/10.1016/j.rse.2004.06.015>.
- 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, 1747–1755. <https://doi.org/10.1016/j.rse.2010.03.002>.
- Hall, D.K., Riggs, G.A., Salomonson, V.V., 1995. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sens. Environ.* 54, 127–140. [https://doi.org/10.1016/0034-4257\(95\)00137-P](https://doi.org/10.1016/0034-4257(95)00137-P).
- Haralick, R.M., Sternberg, S.R., Zhuang, X., 1987. Image analysis using mathematical morphology. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-9, 532–550. <https://doi.org/10.1109/TPAMI.1987.4767941>.
- Hollstein, A., Segl, K., Guanter, L., Brell, M., Enesco, M., 2016. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images. *Remote Sens.* 8, 1–18. <https://doi.org/10.3390/rs8080666>.
- Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., Vogelmann, J.E., 2010a. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sens. Environ.* 114, 183–198. <https://doi.org/10.1016/j.rse.2009.08.017>.
- Huang, C., Thomas, N., Goward, S.N., Masek, J.G., Zhu, Z., Townshend, J.R.G., Vogelmann, J.E., 2010b. Automated masking of cloud and cloud shadow for forest change analysis using Landsat images. *Int. J. Remote Sens.* 31, 5449–5464. <https://doi.org/10.1080/01431160903369642>.
- 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, 4907–4926. <https://doi.org/10.3390/rs6064907>.
- Irish, R.R., 2000. Landsat 7 automatic cloud cover assessment, in: algorithms for multi-spectral, hyperspectral, and ultraspectral imagery vi. International Society for Optics and Photonics 348–356. <https://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, 1179–1188. <https://doi.org/10.14358/PERS.72.10.1179>.
- Jönsson, P., Cai, Z., Melaas, E., Friedl, M.A., Eklundh, L., 2017. A method for robust estimation of seasonality for Landsat and Sentinel-2 time series data. *Remote Sens.* 10, 635. <https://doi.org/10.3390/rs10040635>.
- Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sens. Environ.* 112, 1196–1211. <https://doi.org/10.1016/j.rse.2007.08.011>.
- Landsberg, H.E., 1981. *The Urban Climate*. Academic press.
- Li, J., Roy, D.P., 2017. A global analysis of Sentinel-2a, Sentinel-2b and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens.* 9, 902. <https://doi.org/10.3390/rs9090902>.
- Li, X., Gong, P., Liang, L., 2015. A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data. *Remote Sens. Environ.* 166, 78–90. <https://doi.org/10.1016/j.rse.2015.06.007>.
- Louis, J., Debaecker, V., Pflug, B., Mainknorn, M., Bieniarz, J., 2016. Sentinel-2 Sen2Cor: L2A processor for users. In: *Proceedings Living Planet Symposium 2016*. Spacebooks Online, pp. 1–8.
- Loveland, T.R., Reed, B.C., Ohlen, D.O., Brown, J.F., Zhu, Z., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* 21, 1303–1330. <https://doi.org/10.1080/014311600210191>.
- Mainknorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2. In: *Image and Signal Processing for Remote Sensing XXIII*. International Society for Optics and Photonics, pp. 3. <https://doi.org/10.1117/12.2278218>.
- Melaas, E.K., Friedl, M.A., Zhu, Z., 2013. Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data. *Remote Sens. Environ.* 132, 176–185. <https://doi.org/10.1016/j.rse.2013.01.011>.
- Melesse, Assefa M., Jordan, Jonathan D., 2002. A comparison of fuzzy vs. augmented-ISO DATA classification algorithms for cloud-shadow discrimination from Landsat images. *Photogramm. Eng. Remote Sens.* 68 (9), 90–912.
- Müller-Wilm, U., Devignot, O., Pessiot, L., 2018. Sen2Cor Configuration and User Manual - Ref. S2-PDGS-MPC-L2A-SUM-V2.5.5. <https://castellon.san.gva.es/documents/4434516/5663339/2015+cuadernillo+rel+prof.v4.pdf>.
- Nikolakopoulos, K.G., Kamaratakis, E.K., Chrysoulakis, N., 2006. SRTM vs ASTER elevation products. Comparison for two regions in Crete, Greece. *Int. J. Remote Sens.* 27, 4819–4838. <https://doi.org/10.1080/01431160600835853>.
- Oishi, Y., Ishida, H., Nakamura, R., 2018. A new Landsat 8 cloud discrimination algorithm using thresholding tests. *Int. J. Remote Sens.* 39 (23), 9113–9133. <https://doi.org/10.1080/01431161.2018.1506183>.
- Oreopoulos, L., Wilson, M.J., Várnai, T., 2011. Implementation on Landsat data of a simple cloud-mask algorithm developed for MODIS land bands. *IEEE Geosci. Remote Sens. Lett.* 8, 597–601. <https://doi.org/10.1109/LGRS.2010.2095409>.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man. Cybern.* 9, 62–66. <https://doi.org/10.1109/TSMC.1979.4310076>.
- Pasquarella, V.J., Holden, C.E., Woodcock, C.E., 2018. Improved mapping of forest type using spectral-temporal Landsat features. *Remote Sens. Environ.* 210, 193–207. <https://doi.org/10.1016/j.rse.2018.02.064>.
- Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540, 418–422. <https://doi.org/10.1038/nature20584>.
- Potapov, P., Turubanova, S., Hansen, M.C., 2011. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sens. Environ.* 115, 548–561. <https://doi.org/10.1016/j.rse.2010.10.001>.
- Qiu, S., He, B., Zhu, Z., Liao, Z., Quan, X., 2017. Improving Fmask cloud and cloud shadow detection in mountainous area for Landsats 4–8 images. *Remote Sens. Environ.* 199, 107–119. <https://doi.org/10.1016/j.rse.2017.07.002>.
- Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L., Zhu, Z., Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L., Zhu, Z., 2019. Making Landsat time series consistent: evaluating and improving Landsat analysis ready data. *Remote Sens.* 11 (1), 51. <https://doi.org/10.3390/RS11010051>.
- Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalsky, V., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010. Web-enabled Landsat Data (WELD)—Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sens. Environ.* 114 (1), 35–49. <https://doi.org/10.1016/j.rse.2009.08.011>.
- Roy, D.P., Wulder, M.A., Loveland, T.R., C.E., W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindischadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalsky, V., Lee, Z.P., Lyburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172. <https://doi.org/10.1016/j.rse.2014.02.001>.
- Scaramuzza, P.L., Bouchard, M.A., Dwyer, J.L., 2012. Development of the Landsat data continuity mission cloud-cover assessment algorithms. *IEEE Trans. Geosci. Remote Sens.* 50, 1140–1154. <https://doi.org/10.1109/TGRS.2011.2164087>.
- Schwanghart, W., Scherler, D., 2014. Short communication: TopoToolbox 2 - MATLAB-based software for topographic analysis and modeling in earth surface sciences. *Earth Surf. Dyn.* 2, 1–7. <https://doi.org/10.5194/esurf-2-1-2014>.
- Shao, Z., Deng, J., Wang, L., Fan, Y., Sumari, N.S., Cheng, Q., 2017. Fuzzy AutoEncode based cloud detection for remote sensing imagery. *Remote Sens.* 9, 1–19. <https://doi.org/10.3390/rs9040311>.
- Six, D., Fily, M., Alvain, S., Henry, P., Benoist, J.P., 2004. Surface characterisation of the Dome Concordia area (Antarctica) as a potential satellite calibration site, using Spot 4/Vegetation instrument. *Remote Sens. Environ.* 89, 83–94. <https://doi.org/10.1016/j.rse.2003.10.006>.
- Storey, J., Roy, D.P., Masek, J., Gascon, F., Dwyer, J., Choate, M., 2016. A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sens. Environ.* 186, 121–122. <https://doi.org/10.1016/j.rse.2016.08.025>.
- Sun, L., Mi, X., Wei, J., Wang, J., Tian, X., Yu, H., Gan, P., 2017. A cloud detection algorithm-generating method for remote sensing data at visible to short-wave infrared wavelengths. *ISPRS J. Photogramm.* 124, 70–88. <https://doi.org/10.1016/j.isprsjprs.2016.12.005>.
- Tachikawa, T., Hato, M., Kaku, M., Iwasaki, A., 2011. Characteristics of ASTER GDEM version 2. In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, pp. 3657–3660. <https://doi.org/10.1109/IGARSS.2011.6050017>.
- Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A.K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., 2015. Free and open-access satellite data are key to biodiversity conservation. *Biol. Conserv.* 182, 173–176. <https://doi.org/10.1016/j.biocon.2014.11.048>.
- Veh, G., Korup, O., Roessner, S., Walz, A., 2018. Detecting Himalayan glacial lake outburst floods from Landsat time series. *Remote Sens. Environ.* 207, 84–97. <https://doi.org/10.1016/j.rse.2017.12.025>.
- Vermote, E., Saleous, N., 2007. LEDAPS surface reflectance product description. University of Maryland Department of Geography, College Park.
- Wang, B., Ono, A., Muramatsu, K., Fujiwara, N., 1999. Automated detection and removal of clouds and their shadows from Landsat TM images. *IEICE Trans. Inf. Syst.* 82, 453–460.
- Wilson, M.J., Oreopoulos, L., 2013. Enhancing a simple MODIS cloud mask algorithm for the Landsat data continuity mission. *IEEE Trans. Geosci. Remote Sens.* 51, 723–731. <https://doi.org/10.1109/TGRS.2012.2203823>.
- Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindischadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., 2008. Free access to Landsat imagery. *Science* (80-) 320, 1011a. <https://doi.org/10.1126/science.320.5879.1011a>.
- Wulder, M.A., Hilker, T., White, J.C., Coops, N.C., Masek, J.G., Pflugmacher, D., Crevier, Y., 2015. Virtual constellations for global terrestrial monitoring. *Remote Sens. Environ.* 170, 62–76. <https://doi.org/10.1016/j.rse.2015.09.001>.
- Xu, H., 2007. Extraction of urban built-up land features from Landsat imagery using a thematic-oriented index combination technique. *Photogramm. Eng. Remote Sensing* 73, 1381–1391. <https://doi.org/10.14358/PERS.73.12.1381>.

- Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sens. Environ.* 144, 42–64. <https://doi.org/10.1016/j.rse.2014.01.006>.
- Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* 24, 583–594. <https://doi.org/10.1080/01431160304987>.
- Zhang, Y., Guindon, B., Cihlar, J., 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. *Remote Sens. Environ.* 82, 173–187. [https://doi.org/10.1016/S0034-4257\(02\)00034-2](https://doi.org/10.1016/S0034-4257(02)00034-2).
- Zhou, G., Zhou, X., Yue, T., Liu, Y., 2016. An optional threshold with SVM cloud detection algorithm and DSP implementation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* 41, 771–777. <https://doi.org/10.5194/isprsarchives-XLI-B8-771-2016>.
- Zhu, X., Helmer, E.H., 2018. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. *Remote Sens. Environ.* 214, 135–153. <https://doi.org/10.1016/j.rse.2018.05.024>.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>.
- Zhu, Z., Woodcock, C.E., 2014. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: an algorithm designed specifically for monitoring land cover change. *Remote Sens. Environ.* 152, 217–234. <https://doi.org/10.1016/j.rse.2014.06.012>.
- Zhu, Z., 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. <https://doi.org/10.1016/j.rse.2014.12.014>.
- Zhu, Z., Fu, Y., Woodcock, C.E., Olofsson, P., Vogelmann, J.E., Holden, C., Wang, M., Dai, S., Yu, Y., 2016. Including land cover change in analysis of greenness trends using all available Landsat 5, 7, and 8 images: a case study from Guangzhou, China (2000–2014). *Remote Sens. Environ.* 185, 243–257. <https://doi.org/10.1016/j.rse.2016.03.036>.
- Zhu, Z., Qiu, S., He, B., Deng, C., 2018. Cloud and cloud shadow detection for Landsat images: the fundamental basis for analyzing Landsat time series. *Remote Sens. Time Ser. Image Process.* 3–24. <https://doi.org/10.1201/9781315166636-10>.
- Zhu, Z., Wulder, M.A., Roy, D.P., Woodcock, C.E., Hansen, M.C., Radeloff, V.C., Healey, S.P., Schaaf, C., Hostert, P., Stroh, P., Pekel, J.F., 2019. Benefits of the free and open Landsat data policy. *Remote Sens. Environ.* 224, 382–385.
- Zi, Y., Xie, F., 2018. A cloud detection method for Landsat 8 images based on PCANet. *Remote Sens.* 10, 1–21. <https://doi.org/10.3390/rs10060877>.