https://dx.doi.org/10.1016/j.rse.2018.05.024

An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions

Xiaolin Zhu[a*], Eileen H. Helmer[b]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong

[b] International Institute of Tropical Forestry, USDA Forest Service, Río Piedras, Puerto Rico, 00926, USA

*Corresponding author:

Xiaolin Zhu

Address: The Hong Kong Polytechnic University, Room ZS621, South Wing, Block Z, 181 Chatham Road South, Kowloon, Hong Kong.

Phone: 852-2766-5976; Email: xiaolin.zhu@polyu.edu.hk

Abstract: Clouds and cloud shadows block land surface information in optical satellite images. Accurate detection of clouds and cloud shadows can help exclude these contaminated pixels in further applications. Existing cloud screening methods are challenged by cloudy regions where most of satellite images are contaminated by clouds. To solve this problem for landscapes where the typical frequency of cloud-free observations of a pixel is too small to use existing methods to mask clouds and shadows, this study presents a new Automatic Time-Series Analysis (ATSA) method to screen clouds and cloud shadows in multi-temporal optical images. ATSA has five main steps: (1) calculate cloud and shadow indices to highlight cloud and cloud shadow information; (2) obtain initial cloud mask by unsupervised classifiers; (3) refine initial cloud mask by analyzing time series of a cloud index; (4) predict the potential shadow mask using geometric relationships; and (5) refine the potential shadow mask by analyzing time series of a shadow index. Compared with existing methods, ATSA needs fewer predefined parameters, does not require a thermal infrared band, and is more suitable for areas with persistent clouds. The performance of ATSA was tested with Landsat-8 OLI images, Landsat-4 MSS images, and Sentinel-2 images in three sites. The results were compared with a popular method, Function of Mask (Fmask), which has been adopted by USGS to produce Landsat cloud masks. These tests show that ATSA and Fmask can get comparable cloud and shadow masks in some of the tested images. However, ATSA can consistently obtain high accuracy in all images, while Fmask has large omission or commission errors in some images. The quantitative accuracy was assessed using manual cloud masks of 15 images. The average cloud producer's accuracy of these 15 images is as high as 0.946 and the average shadow producer's accuracy reaches 0.884. Given that it can be applied to old satellite sensors and it is capable for cloudy regions, ATSA is a valuable supplement to the existing cloud screening methods.

Key words: Cloud detection; Cloud shadow; Mask; Optical satellite images; Time series

1. Introduction

Optical satellite images with bands ranging from visible to shortwave infrared are widely used for mapping land cover and land use, monitoring ecosystems, and estimating land surface parameters (Hansen and Loveland, 2012; Zhu and Liu, 2015, 2014). Unfortunately, optical satellite images are easily contaminated by clouds and cloud shadows. This contamination obscures land surface features and alters the reflectance of ground objects, reducing the availability of optical images for applications (Fisher, 2013; Zhu and Woodcock, 2014). Masking clouds and cloud shadows is often the first and a necessary step of image preprocessing in most optical remote sensing applications. Although manual digitization can obtain accurate cloud and shadow masks, it requires a lot of time and effort. Therefore, an automatic method for screening clouds and shadows is needed, especially when processing large numbers of images.

Automatic detection of clouds and cloud shadows is challenging (Zhu and Woodcock, 2014). First, different types of clouds have different spectral signatures and are easily confused with some cloud-free bright objects on the land surface, especially in images with limited spectral bands, such as Landsat Multispectral Scanner (MSS) images. The spectral signals of clouds are usually determined by cloud height, optical thickness, particle size, etc. (Platnick et al., 2003). As a result, cloud brightness ranges widely in visible and near infrared bands, and some clouds are easily confused with bright land surfaces, such as concrete surfaces, sand or snow. Second, blurry cloud edges and thin clouds partially obscure land surfaces, making their signal a mixture of cloud and land surface elements and making them difficult to separate from clear observations (Cahalan et al., 2001). Another challenge comes from cloud shadows. They are easily confused with dark land surfaces, such as moist soil, water bodies and topographic shadow (Fisher, 2013).

Despite the above challenges, several methods have been developed to automatically screen clouds and cloud shadows in optical images. These methods use one or more of the following rules based on cloud and cloud shadow properties: 1) clouds are generally brighter than ground objects, so they have high reflectance in visible, near and shortwave infrared bands; 2) clouds are generally colder than most ground objects, so they have lower brightness in thermal infrared bands; 3) shadows are generally darker than surrounding land surfaces, so they have lower reflectance in visible, near and shortwave infrared bands; 4) shadows are paired with clouds, so cloud location and solar angles can help locate cloud shadows; and 5) in a sequence of images, pixels affected by clouds and shadows have larger temporal variations than clear observations in the time series. In general, existing methods for masking clouds and cloud shadows can be divided into two categories: single-image methods (Choi and Bindschadler, 2004; Fisher, 2013; Helmer et al., 2012; Huang et al., 2010; Hughes and Hayes, 2014; Irish et al., 2006; Li et al., 2015, 2017; Luo et al., 2008; Martinuzzi et al., 2007; Roy et al., 2010; Scaramuzza et al., 2012; Wilson and Oreopoulos, 2013; Zhu and Woodcock, 2012) and multi-temporal or bi-temporal methods (Goodwin et al., 2013; Hagolle et al., 2010; Jin et al., 2013; Wang et al., 1999; Zhu and Woodcock, 2014).

Most existing single-image methods use either predefined thresholds or adaptive thresholds to screen clouds in individual images. For example, Luo et al. (2008) identify clouds in MODIS images if pixel reflectance satisfies these predefined thresholds: ($B1 > 0.18$ or $B3 > 0.20$) and $B6 > 0.16$ and Maximum ($B1$, $B3$) $> B6\times0.67$, where B1, B3, and B6 are reflectance of MODIS bands 1 (blue), 3 (red), and 6 (shortwave infrared), respectively. This MODIS cloud screening method was further adopted for Landsat-8 images (Wilson and Oreopoulos, 2013). Huang et al. (2010) use adaptive thresholds defined in the reflectance-temperature space to mask clouds in Landsat TM and ETM+ images. These adaptive thresholds are defined by the mean and standard deviation of pixel values of individual bands in the whole image. The Automatic cloud cover assessment

(ACCA) algorithm consists of twenty-six filters and rules applied to Landsat bands to detect clouds (Irish et al., 2006). ACCA was used to produce web-enable Landsat data (WELD), a consistent, long-term, and large-area data record (Roy et al., 2010). The multi-feature combined (MFC) method uses thresholds in spectral, geometric and texture features to detect clouds in GaoFen-1 imagery (Li et al., 2017). Zhu and Woodcock (2012) proposed a method called function of mask (Fmask) for detecting clouds in Landsat TM and ETM+ images. Fmask uses all Landsat image bands and several band indices, such as the normalized difference vegetation index (NDVI) and the normalized difference snow index (NDSI). It employs more than 20 predefined and adaptive thresholds to mask clouds. Besides the above methods using predefined or adaptive thresholds, machine-learning algorithms have been employed to model the complex relationships between image features and clouds using a training dataset. Then, the trained model is used to screen clouds in other images. These machine learning algorithms include decision trees (Scaramuzza et al., 2012), neural networks (Hughes and Hayes, 2014) and support vector machines (Li et al., 2015). Of several tested cloud and shadow masking algorithms that use only a single image, Fmask is globally the most accurate one that requires a thermal band (Foga et al., 2017). Of methods not requiring a thermal band, a version of ACCA (Irish et al., 2006) that uses a simulated thermal band is better overall, but it is not as accurate as Fmask with the thermal band (Foga et al., 2017). Recently, Fmask was further improved for mountainous areas through integrating Digital Elevation Models (DEMs) into the detecting process (Qiu et al., 2017).

In these single-image methods, shadow detection is often subsequent to cloud detection. In general, the possible shadow locations can be calculated from the geometric relationship between sun, sensor, and clouds. The calculation requires cloud heights, which can be estimated with brightness temperature derived from thermal infrared bands, because temperature declines with elevation (Qiu et al., 2017; Zhu and Woodcock, 2012). Some methods also use the fact that cloud shadows are dark to confirm whether the possible shadow location estimated from geometry is real cloud shadow, including Fmask (Zhu and Woodcock, 2012) and MFC (Li et al., 2017). In Fmask, predefined thresholds in the near infrared (NIR) band are used to produce a potential shadow mask, which is further compared to the possible shadow locations. If there is a high similarity between potential shadow masks and possible shadow locations, the shadow pixels are finally confirmed (Zhu and Woodcock, 2012).

For multi-temporal methods, temporal information in the images acquired at different times is used to detect clouds and shadows. Wang et al. (1999) used the brightness difference between a target image and a reference cloud-free image to detect clouds. Lyapustin et al. (2008) developed an algorithm, abbreviated as MAIAC CM, to detect clouds in time series of MODIS images. The general idea of MAIAC CM is to use the low covariance between reference cloud-free image blocks and cloudy image blocks as a criterion to identify clouds in the time series. Hagolle et al. (2010) computes differences in the blue band between a target image and a cloud-free reference image. It then flags cloud pixels if variations are larger than a threshold. Goodwin et al. (2013) uses filters to smooth the time-series and then identify clouds and shadows based on reflectance differences between each point in the time series and the smoothed time series. Zhu and Woodcock (2014) propose a new algorithm called multiTemporal mask (Tmask) to improve Fmask. Tmask fits a time series model of each pixel using remaining clear pixels based on an initial cloud mask from Fmask. Then, it compares model estimates with observations in the time series to detect cloud and shadow pixels which are omitted in the initial screening by Fmask. In general, these multi-temporal methods are better at detecting clouds and cloud shadows than single-image methods. The temporal information is a valuable complement to the spectral information for differentiating cloud, cloud shadow and clear observations over land surfaces (Goodwin et al., 2013; Zhu and Woodcock, 2014).

However, these multi-temporal methods still face challenges in areas with persistent cloud cover, such as tropical and subtropical regions (Ju and Roy, 2008). First, in these areas cloud-free observations may be the exception rather than the rule, making it difficult to know whether the fit of a time series represents clear or cloudy conditions, which limits the application of existing time-series methods (Foga et al., 2017). Example limitations include the requirement by the MAIAC CM method of a cloud free image as a reference image (Lyapustin et al., 2008), and the recommendation for Tmask of 15 cloud-free observations for estimating the time series model (Zhu and Woodcock, 2014). Second, most existing methods were designed for images of a specific sensor, so they lack flexibility. For example, Fmask and Tmask were designed for Landsat TM, ETM+, and OLI images, so they cannot be directly applied to the old Landsat MSS data with limited bands. Third, most of the current methods use predefined fixed thresholds to detect clouds and shadows in an entire scene. For instance, in Tmask, a pixel with observed green band reflectance of 0.04 higher than the time series model estimation will be identified as cloud. Considering the complex situation of clouds and shadows and the diversity of objects on land surfaces and in coastal areas, these fixed thresholds may not always obtain satisfactory results.

To overcome the above limitations of existing methods in cloudy regions, the objective of this study is to develop a new automatic method for accurately screening clouds and cloud shadows in multi-temporal optical images in places with persistent clouds. Our scope of inference is landscapes where are so cloudy that the typical frequency of cloud-free observations of a pixel is too small to use existing methods to mask clouds and shadows with image time series. The new method should have the following strengths: 1) it needs fewer predefined parameters; 2) it is suitable for areas with persistent clouds; and 3) it needs a minimal number of bands. Automatic Time-Series Analyses (ATSA) method was developed in this study and tested in three pilot sites using Landsat OLI and MSS images, and Sentinel-2 images. Its performance was compared with Fmask, a widely recognized method.

2. Test Sites and Data

2.1. A cloudy urban site

Urban landscapes bring more challenges to automatic screening of clouds and shadows than other landscapes. The bright built-up area often leads to large commission errors in cloud detection. To test the effectiveness of the proposed method in such challenging cases, we selected Hong Kong, a cloudy subtropical dense city with complex and mixed land-cover types (Fig. 1). This site has an area of 1, 620 km^2 (1200×1500 Landsat pixels), and the central coordinates are 22.367° N and 114.123° E. It is covered by the Landsat scene of Worldwide Reference System 2 (WRS-2) Path 122 Row 44. All 23 available Landsat-8 OLI level-1 images in 2015 were downloaded from USGS Earth Explorer. These images were then converted to Top of Atmosphere (TOA) reflectance with the scaling coefficients in the metadata file. The corresponding Fmask cloud masks of these images were also downloaded from USGS Earth Explorer. Based on Fmask cloud masks, only two images are clear, while the other images have total cloud and shadow coverage ranging from 5.5% to 97%. Sixteen of them have total cloud and shadow coverage larger than 60%, indicating Landsat imagery in this site is seriously contaminated by clouds (Table 1).

Fig. 1. True color composition of a Landsat-8 image of 2015, DOY003 in Hong Kong

Table 1. Summary of cloud and shadow coverage of Landsat-8 images from the year 2015 over the Hong Kong site using Fmask product. Only two images have no clouds.

DOY	Cloud coverage %	Shadow coverage %	Total cloud and shadow coverage%
3	0.0	0.0	0.0
19	0.0	0.0	0.0
35	97	0.0	97
51	65	3.8	69
67	90	0.65	90
83	87	3.1	90
99	89	0.0	89
115	82	0.06	82
131	32	5.4	38
147	94	0.25	94
163	64	3.7	68
179	62	10	72
195	48	6.8	55
211	83	2.7	85
227	95	0.00	95
243	95	0.36	95
259	93	0.52	93
275	78	4.2	82
291	6.6	2.2	8.7
307	93	0.71	93
323	42	12	55
339	3.7	1.8	5.5
355	79	14	93

2.2. A cloudy forest site

Dense time series data are needed for monitoring vegetation dynamics, and monitoring tropical and subtropical forests is very important to quantifying their important role in the global carbon cycle. However, persistent cloud cover poses challenges when monitoring tropical forest vegetation. To investigate the accuracy of the proposed method to screen clouds and shadows in cloudy tropical forest regions, the second site is northeastern Puerto Rico (Fig. 2). This site has an area of 1,836 km^2 (1200×1700 Landsat pixels), and the central coordinates are 18.321° N and 65.838° W. The major land cover type is forest, including the EI Yunque National Forest, where extensive tropical montane cloud forests occur that by definition are persistently cloudy. This site also includes bright, wet and dark features that are easily confused with clouds or cloud shadows. It includes much of the capital city of Puerto Rico, San Juan, coastal areas with features like sand, rock and coral reefs, topographic shadow associated with steep topography and many fields with bright, wet or bare soils. The Landsat WRS-2 scene Path 4 and Row 47 covers the area. A total of 18 Landsat 8 OLI images from May 26 2013 to May 29 2014 (i.e., one-year length) and their corresponding Fmask cloud layers were downloaded from USGS Earth Explorer. The total cloud and shadow coverage of the images as estimated by Fmask ranges from 5% to 81%, and the mean coverage is 45%, indicating this site is also seriously affected by clouds. In this site, another 11 Landsat-4 MSS images from the year 1983 were collected to test the performance of the proposed method for screening clouds and shadows in images with limited bands and low radiometric resolution. For these MSS images, corresponding Fmask cloud masks are not available from USGS Earth Explorer because Fmask uses thermal bands, which are not included in MSS images. Through visual inspection, these 11 MSS images have diverse cloud and shadow coverage, from almost cloud-free to fully covered by clouds.

Fig. 2. True color composition of a Landsat-8 image from 2013 (DOY210) of northeastern Puerto Rico.

Table 2. Summary of cloud and shadow coverage of Landsat 8 OLI images for the Puerto Rico site using Fmask product

DOY	Cloud coverage %	Shadow coverage %	Total cloud and shadow coverage%
146	41	4.1	45
178	43	8.0	51
210	4.3	1.6	5.9
226	39	5.6	45
242	40	8.2	48
258	28	6.5	35
274	67	14	81
290	27	8.1	35
306	8.5	1.1	9.6
322	30	6.7	37
354	58	13	71
5	24	11	35
21	43	13	56
53	41	12	54
69	31	9.5	41
117	37	6.5	44
133	38	3.0	41
149	59	14	72

2.3. A seasonal-change site

Strong seasonality is another challenge for most multi-temporal cloud screening methods. The large variation of spectral values due to seasonality may be confused with the variation due to occurrence of clouds and cloud shadows. To investigate the accuracy of the proposed method to screen clouds and shadows in regions with strong seasonality, the third site is Beijing metropolis and its surrounding rural areas (Fig. 3). This site is covered by an entire Sentinel-2 tile (about 12,000 km^2), and the central coordinates are 40.154° N and 116.495° E. This site has a lot of bright land surface and its vegetation is deciduous with strong seasonality. Images from different seasons in Fig. 3 show that vegetation grows to a peak greenness in summer and loses leaves in winter. In addition, the high mountains in this site bring difficulties for both cloud and cloud shadow detection. Twenty Sentinel-2 images in 2016 with varying cloud cover were downloaded from USGS Earth Explorer (Table 3). The Fmask cloud masks of these Sentinel-2 images were obtained using the Matlab code (Version 3.3; https://github.com/prs021/fmask) specific for Sentinel-2 images (Zhu et al., 2015). The total cloud and shadow coverage of the images as estimated by Fmask ranges from 0.1% to 100%, and 7 images have less than 20%, indicating this site has more clear images than the other two sites.

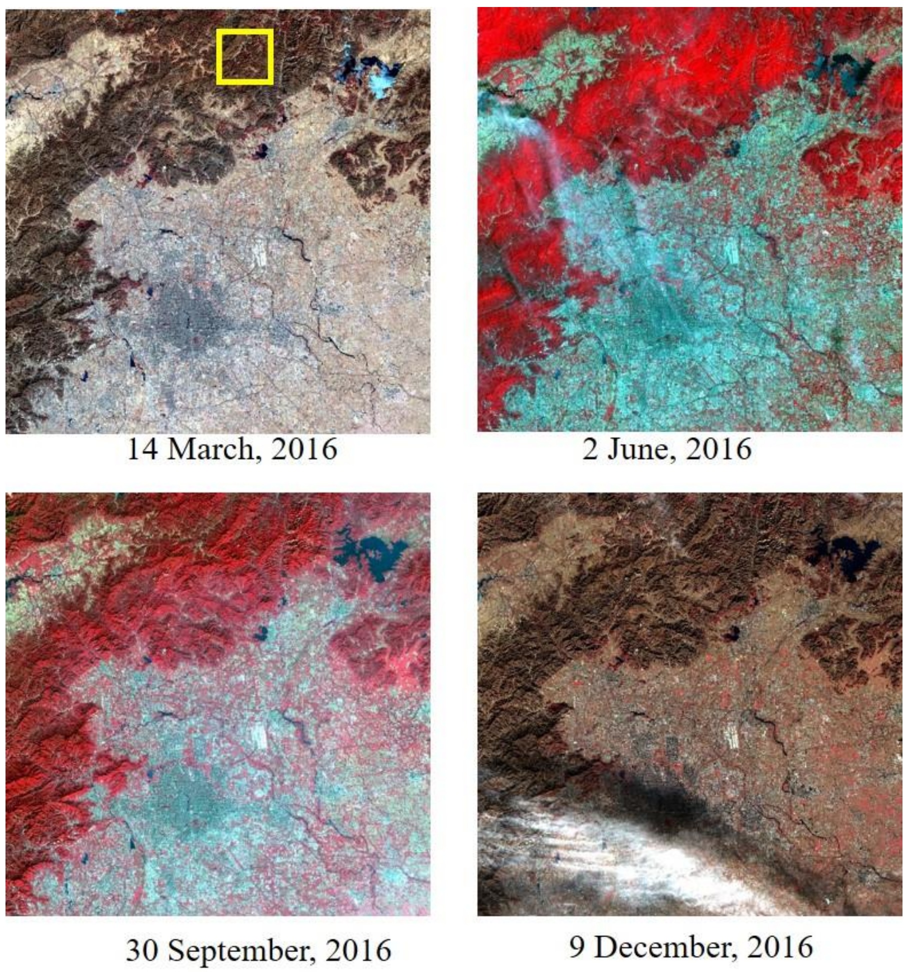


Fig. 3. False-color Sentinel-2 images in Beijing from different seasons in 2016 (the yellow box in upper left image is a forest region of interest (ROI) used to demonstrate the seasonality in Fig. 9)

Table 3. Summary of cloud and shadow coverage of 20 Sentinel-2 images in 2016 for the Beijing site using cloud masks by Fmask

Date	Cloud coverage %	Shadow coverage %	Total cloud and shadow coverage%
Jan.26	42.4	17.7	60.1
Mar.14	0.7	0.5	1.2
Mar.24	0.1	0.0	0.1
Apr.3	0.2	0.2	0.4
Jun.2	13.2	1.8	15.0
Jun.12	56.4	6.2	62.6
Jul.22	90.3	9.7	100.0
Aug.1	78.6	4.6	83.2
Aug.11	41.2	9.4	50.6
Aug.21	22.6	6.1	28.7
Aug.31	7.6	4.6	12.2
Sep.20	36.4	3.5	39.9
Sep.30	34.4	8.1	42.5
Oct.10	19.3	7.3	26.6
Oct.20	100.0	0.0	100.0
Oct.30	75.7	9.3	85.0
Nov.19	2.6	1.8	4.4
Nov.29	100.0	0.0	100.0
Dec.9	16.0	8.8	24.8
Dec.29	7.1	4.1	11.2

3. Methodology

There are five main steps in ATSA (Fig. 4). Either TOA reflectance or surface reflectance data can be used as inputs. The five main steps are: (1) compute a cloud index and a shadow index from the image bands to highlight cloud and shadow pixels; (2) detect clouds initially with unsupervised clustering of these indices for individual images in the time series; (3) refine the cloud pixels through analyzing the time series of the cloud index; (4) predict the potential shadow locations through the geometric relationships among the sun, clouds, and the Earth surface; (5) confirm the real shadow pixels through analyzing the time series of the shadow index. We detail these steps below.

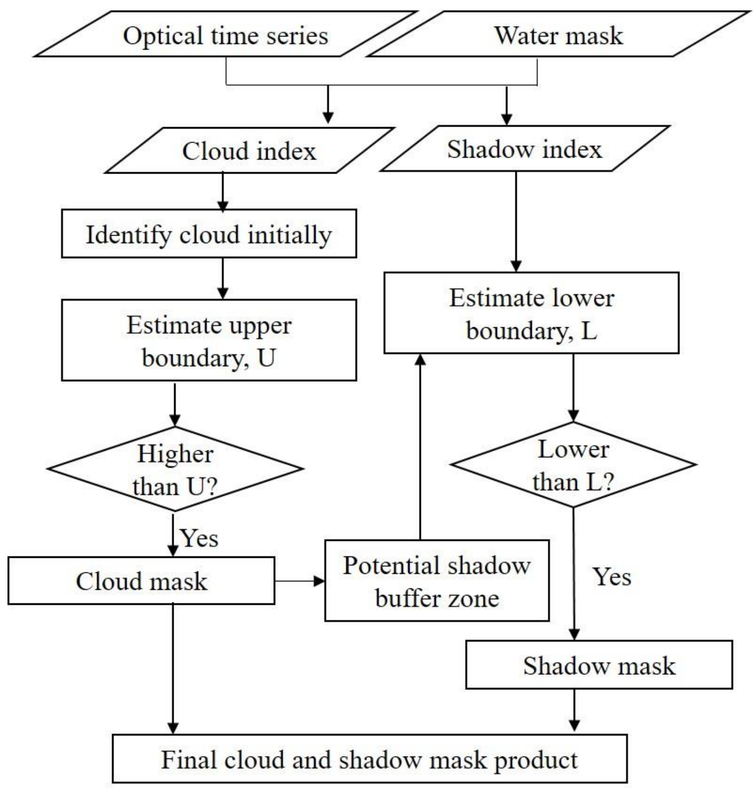


Fig. 4. Flowchart for the Automatic Time-Series Analysis (ATSA) to screen clouds and cloud shadows

3.1. Calculate cloud index and shadow index

Given the wide ranges of reflectance values exhibited by diverse cloud- and Earth surface types, individual spectral bands from one image cannot accurately differentiate clouds, cloud shadows and clear observations. With image time series, significant seasonality of some land cover types (*e.g.,* natural vegetation and agriculture) and land-cover change (*e.g.*, deforestation and urbanization) lead to large temporal variability of reflectance in image time series, which is easily confused with temporal variability caused by clouds and cloud shadows. Therefore, there is a need to combine or transform individual bands to get indices that highlight the clouds and cloud shadows while compressing variability in other land cover types, so that clear observations have values that are as stable as possible in the time series.

As land and water surfaces have very different spectral characteristics (Zhu and Woodcock, 2012), the cloud and shadow indices are designed separately for land and water surfaces. A water mask is needed in our method. Fortunately, a water mask can be easily obtained through classifying a cloud-free image in the time series or from an existing water mask. There are now several water masks available at different resolutions, such as a 30-m water mask from a Landsat-based global land cover product (Chen et al., 2015) and a 250-m global water mask from MODIS data (available in http://landcover.org/data/watermask/). In our test experiments, we classified a cloud-free image to obtain the water mask.

For land surfaces, we used the haze optimal transformation (HOT) as a cloud index. The HOT transformation is derived from an analysis of Red-Blue spectral space. These two bands have a perfect linear relationship for diverse land cover types under clear-sky conditions (Zhang et al., 2002), and Zhang et al. (2002) name this perfect line the clear-sky line (see the red line in Fig. 5 a). For pixels contaminated by haze and clouds, their spectral response in Red-Blue space is very different from the clear-sky line, so the HOT index was designed to quantify the perpendicular distance of a pixel from the clear-sky line:

$$HOT = \frac{|a \times B_{Blue} - B_{Red} + b|}{\sqrt{1 + a^2}} \quad (1)$$

where B_{Blue} and B_{Red} are pixel values of blue and red bands respectively, and a and b are the slope and intercept of the clear-sky line.

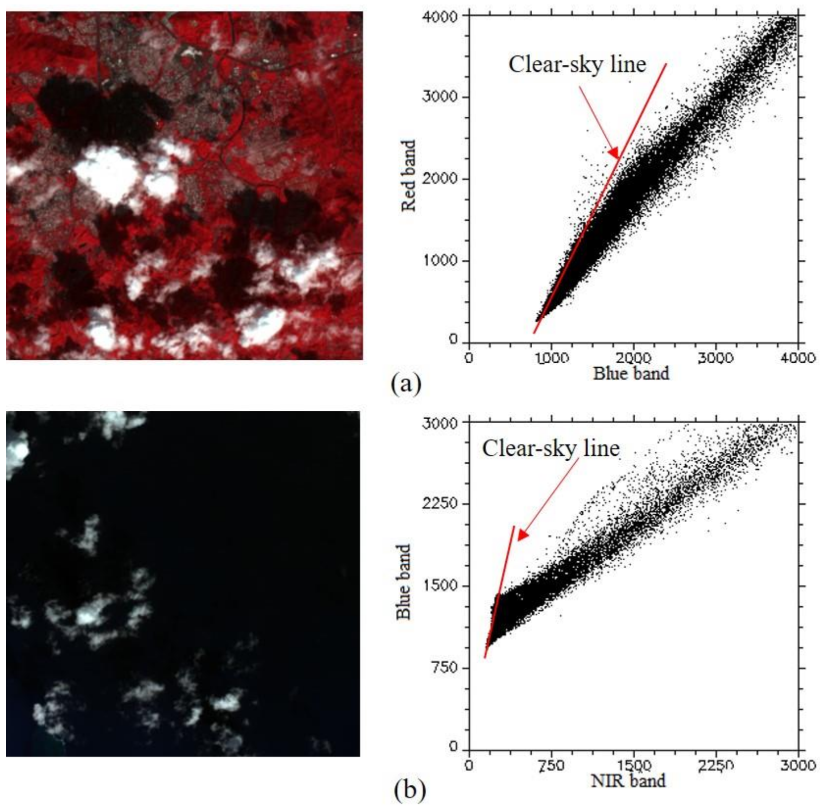


Fig. 5. A land-surface subset of a Landsat-8 image and its Red-Blue scatter plot (a) and a water-surface subset Landsat-8 image and its NIR-Blue scatter plot (b)

In the original HOT transformation (Zhang et al., 2002), the clear-sky line comes from regressing spectral values of pixels selected from areas of a scene that visually are deemed to be the clearest. To make our method automatic, we employed a bin-based approach to search for the clearest pixels in each scene. This approach has three steps: (1) divide the 0-0.15 range of blue reflectance values into 50 bins with equal intervals, because cloud-free pixel values of most land-cover types are within this range; (2) for each bin, select the 20 pixels with the largest reflectance values in the Red band and compute the average value of these selected pixels for red and blue bands, respectively, yielding a pair of red and blue reflectance values for each bin (B_{Redi}, B_{Bluei}) ; (3) for all 50 pairs of (B_{Redi}, B_{Bluei}), regress B_{Redi} against B_{Bluei} to get the clear-sky line using the least absolute deviation (LAD) regression method to avoid the effect of outliers (Bassett and Koenker, 1978). If some images in the time series are completely covered by clouds, no clear pixels can be found for estimating the clear-sky line. For these completely cloud-covered images, the average slope and intercept of clear-sky lines derived from other images in the time series are used to compute the HOT index. To demonstrate the effectiveness of a bin-based automatic approach, the retrieved clear-sky line in a sub-image was compared with the result using manually selected clear pixels (Fig. 6). The slope and intercept of the clear-sky line from the bin-based approach is very similar to the results from the manual approach.

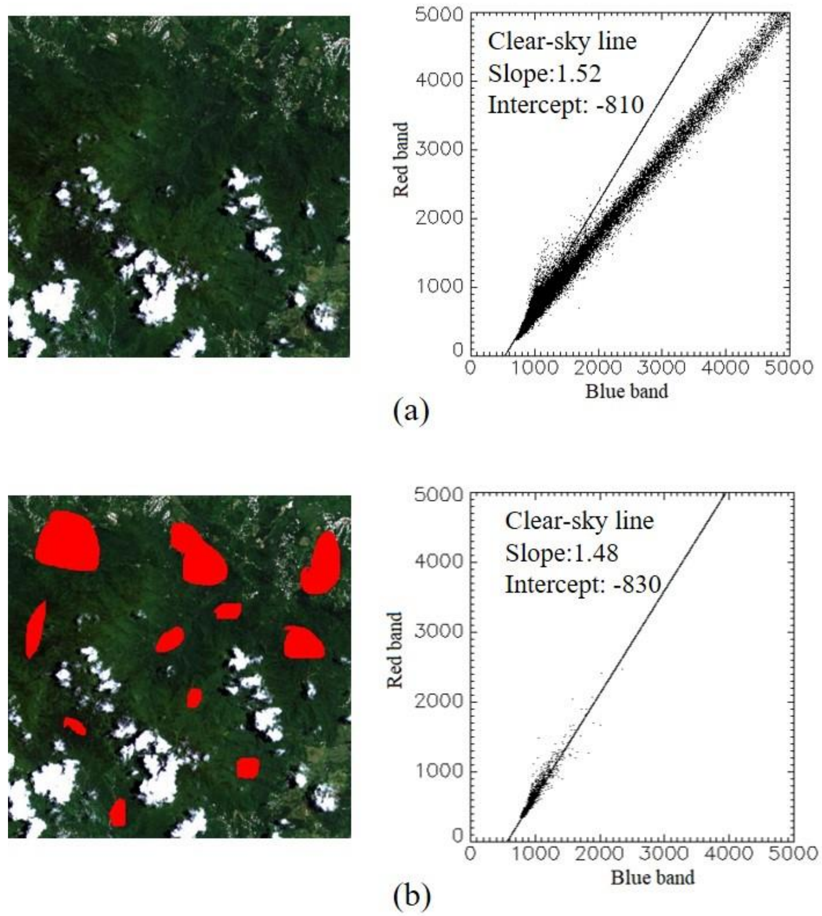


Fig. 6. Comparison between the clear-sky line of a sub-image estimated by the proposed automatic bin-based approach (a) and that using manually selected clear pixels marked by red ROIs (b).

For water surfaces, the cloud-free pixel values of the red and blue bands are not on the clear-sky line, leading to large HOT values that are confused with thin clouds. Consequently, a new HOT index, designed specifically for water surfaces, is needed. In the Blue-NIR space, the spectral response of cloud-free water pixels, including turbid or shallow water and coral reefs, is very different from cloudy pixels (Fig. 5b). A new HOT index for water surface, HOT_w is given as:

$$HOT_w = \frac{|a_w \times B_{NIR} - B_{Blue} + b_w|}{\sqrt{1 + {a_w}^2}} \quad (2)$$

where a_w and b_w are the slope and intercept of the clear-sky line for water bodies and are obtained through the same method as for the land surface. Then, the HOT indices for land and water surfaces are combined to yield a cloud index map (Fig. 7b). In this cloud index map, we can see that the HOT transformation yields an index with a larger difference between cloud and bright non-cloud objects than the individual visible bands. All clouds and haze are highlighted by larger values (i.e. white color) while all cloud-free pixels have a very low value (i.e. dark color).

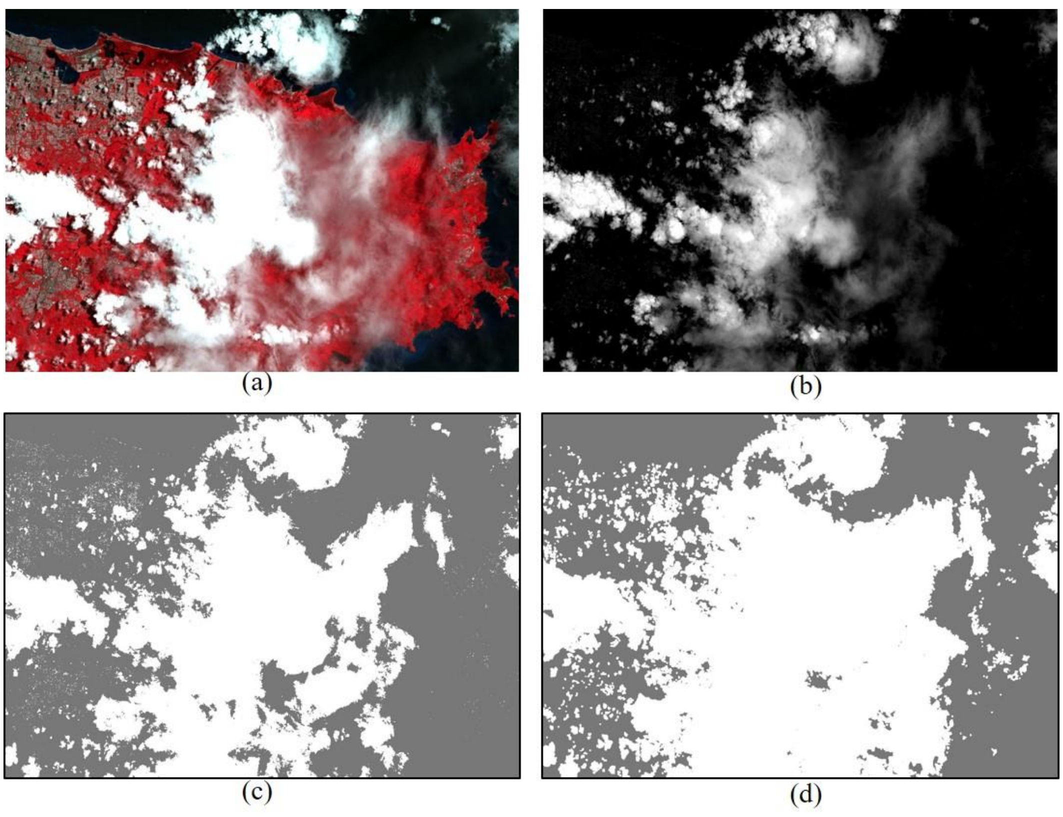


Fig. 7. False color Landsat-8 image of DOY149 in the Puerto Rico site (a), its corresponding HOT cloud index (b), initial cloud mask (c), and final cloud mask (d). In (c) and (d): gray is clear pixels and white is clouds. The time series analysis adds thin clouds to the initial cloud mask, and the minority analysis removes scattered bright pixels in urban and coastal areas in the upper

left of panel (c), which would otherwise be confused with clouds.

To further compare the ability of the original bands and HOT for discriminating clouds and clear land surface, the relative difference (*RD*) between cloud and cloud-free pixels in each image was computed:

$$RD = \frac{\bar{B}_{cloud} - \bar{B}_{clear}}{\bar{B}_{cloud}} \tag{3}$$

where $\bar{B}_{cloud}$ and $\bar{B}_{clear}$ are average values of cloudy pixels and clear pixels respectively. *RD* ranges from 0 to 1 and larger values indicate a higher separability between cloudy and clear pixels. Fig. 8 shows the *RD* values of the Blue band and the HOT index of Landsat-8 images which contain both clear and cloudy pixels in the Hong Kong site. Hong Kong includes both forests and considerable bright urban surfaces. It is a challenging site for cloud detection. We can see that in these images HOT index is better than the original Blue band at separating clouds from clear land surfaces. The comparisons of RD values between the Red band and the HOT index, and between the NIR band and the HOT index, have a similar pattern (results not shown).

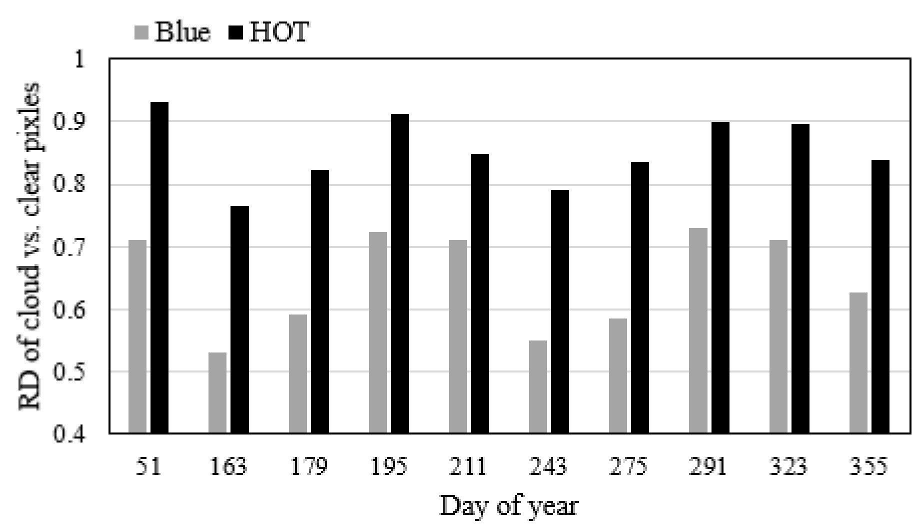


Fig. 8. Relative difference (*RD*) between the average value of cloudy and clear pixels of the Blue

band and the HOT index in Hong Kong Landsat-8 images. A larger *RD* indicates higher separability between cloud and clear pixels.

For cloud shadows, direct solar radiation is blocked by clouds, so the shadow pixels are illuminated by scattered light. Because the atmospheric scattering is weaker at longer wavelengths, the NIR and SWIR bands of shadow pixels are much darker than surrounding clear pixels (Zhu et al., 2015). Therefore, the shadow index (*SI*) is defined as:

$$SI = B_{NIR} + B_{SWIR} \tag{4}$$

However, water also absorbs most radiation at longer wavelengths, so water pixels not obstructed by clouds are as dark as shadow pixels in NIR and SWIR bands (Li et al., 2017). Consequently, for water surfaces, the shadow index is calculated with the blue and green bands:

$$SI_w = B_{Blue} + B_{Green} \tag{5}$$

For old satellite images with fewer bands, such as Landsat MSS images with only green, red, and 2 NIR bands, the green band replaces the blue band in Eqs. (1, 2, and 5), because it is highly correlated with blue band. Also, the second NIR band replaces the SWIR band in Eq. (4), because both the NIR and SWIR bands are good indicators of cloud shadows. Similarly, for other sensors without SWIR bands, such as IKONOS, we anticipate that only one NIR band would be used as the shadow index for land surfaces.

3.2. Detect cloud initially

All cloud index images of the time-series are classified by an unsupervised classifier, k-means, to get the initial cloud mask. First, a certain number (e.g. 10,000) of sample pixels are selected by systematic sampling of all cloud index images. Selecting samples from all images in the time series ensures that samples of clear surfaces, thin clouds, and thick clouds are included. Using the selected samples rather than all pixels speeds up the k-means optimization in the next step. Second, these samples are classified with the k-means method into three classes. The three classes are labeled based on the relative value of the class means, i.e., the lowest class mean is clear pixels, the middle one is thin clouds, and the highest one is the thick clouds. The k-means method uses an iterative procedure. At each iteration, each sample is assigned to one class based on the closeness to the class means obtained from the last iteration, and new class means are updated using new class labels of samples. The iterative process will be ended when the class labels no longer change (Lloyd, 1982). Third, individual pixels in each cloud index image of the time series are identified as thin clouds, thick clouds, or cloud-free observations based on which class has the smallest the cloud-index distance from the class means of the sample pixels. Finally, an initial cloud mask is produced for each image by combining thin clouds with thick clouds (Fig. 7c). The ranges of cloud index values for the three classes (thin clouds, thick clouds and clear), being derived from all pixels in the time series, form a set of thresholds that are adapted to a time series rather than a single image.

3.3. Detect remaining extremely thin clouds and remove bright pixels that are not cloud or haze

Although the initial cloud mask already identifies most cloudy pixels, it may omit some cloudy pixels, especially extremely thin clouds and cloud edges with lower values of the cloud index. Therefore, the initial cloud masks need to be further improved with temporal information. For each pixel, its time series may include both cloudy points and cloud-free observations. In general, cloudy points have larger variations in spectral values than clear observations. This temporal property can help to identify cloudy points (Zhu and Woodcock, 2014). However, due to changes in vegetation phenology or land cover, clear observations of some pixels also undergo temporal variations. However, compared with the original spectral bands, the cloud index derived from the HOT transformation depresses the temporal variations from different vegetation growth stages, soil inundation, or land-cover change. For instance, in a forest ROI from the Beijing site (marked by a yellow box in Fig. 3), the time series of the Red band shows a stronger seasonality than the HOT index (Fig. 9). The Red band is used to compute the HOT index. It has lower values in summer due to more absorption by vegetation. In contrast, the HOT index is more stable across different seasons and with smaller variability than the original Red bands.

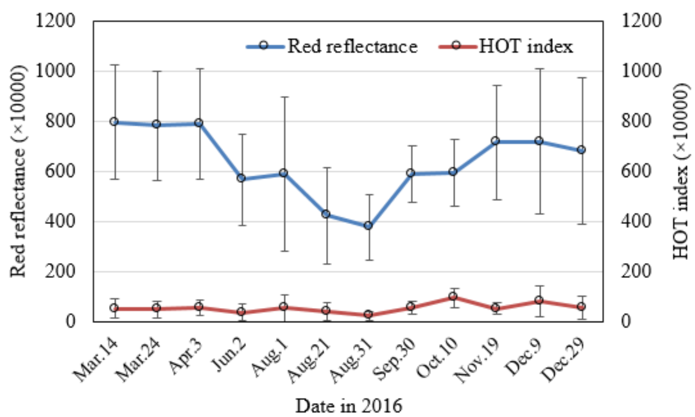


Fig. 9. Red reflectance and HOT index of cloud-free pixels in a forest ROI in Beijing (marked by a yellow box in Fig. 3) across different seasons. The circles are mean values and error bars are ± 1 standard deviation.

Therefore, an analysis of the cloud index time series is conducted to refine the initial cloud mask. Fig. 10 gives an example of a cloud index time series of a pixel from the Puerto Rico site (column1173, row 1092). There are two points identified as clouds in the initial mask (the red points). Most of the other points (the black ones) should be clear observations. They are used to find an upper boundary in the HOT index for clear pixels, $U(i)$. Points above this threshold are then also designated as cloudy. For ith pixel:

$$U(i) = \text{mean}\{HOT(i,t)|(i,t) \notin \mathbf{C}\} + A \times \text{sd}\{HOT(i,t)|(i,t) \notin \mathbf{C}\} \quad (6)$$

where sd{·} is the standard deviation of the HOT index through the time series, $HOT(i, t)$ is the HOT index value of the ith pixel at time t, and **C** is the set of cloudy points from the initial masks for ith pixel. A is a standard deviation multiplier that defines the upper boundary. A can be assigned a recommended value from 1 to 2. Smaller values would be able to identify thinner clouds, but meanwhile increase the risk of commission errors, *i.e.*, identifying "clear" observations as cloudy points. In existing methods, this parameter is a constant for all pixels in the image (Goodwin et al., 2013; Hagolle et al., 2010). However, cloud frequency is different in different parts of the image, so some pixels may include more cloud points in the HOT time series that are omitted in the initial detection than others. Therefore, we need to consider this difference among pixels when we set the value of parameter A. In general, clouds cause large variations in the HOT time series. We introduced a new variable, the normalized difference range index (NDRI), to tune the parameter A:

$$NDRI(i) = (T_{kmeans} - Range_i)/(T_{kmeans} + Range_i) \quad (7)$$

$$Range_i = \max\{HOT(i,t)|(i,t) \notin \mathbf{C}\} - \min\{HOT(i,t)|(i,t) \notin \mathbf{C}\} \quad (8)$$

where T_{kmeans} is the minimum HOT value of all cloud pixels identified by K-means in section 3.2. $NDRI$ is further used to adjust the parameter A in Eq. (6) as a pixel-wise parameter $A(i)$:

$$A(i) = A + NDRI(i) \quad (9)$$

The value for $A(i)$ is used in Eq. (6) to calculate the pixel-level upper boundary U for each pixel in the time series. $A(i)$ further tunes the pixel-level upper boundary U by adapting the standard deviation multiplier to the temporal variability of each pixel. Because $NDRI$ is added to A, we recommend an A value from 0.5 to 1.5 (instead of 1 to 2). Pixels with larger variation in the HOT time series will have a lower upper boundary, i.e. a stricter threshold. Any points above the upper boundary, e.g. the dashed line in Fig. 10, will be identified as clouds.

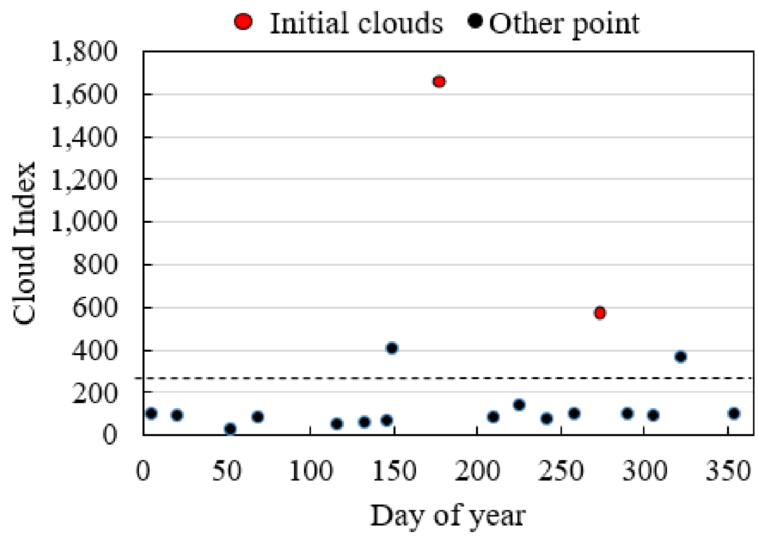


Fig. 10. An example of cloud index time series: the two black points above the dashed line are identified as clouds based on the time series analysis.

The cloudy points detected from the time series analysis are the final cloud mask (Fig. 7d). This step adds more thin clouds to the initial mask and also contributes to filtering bright non-cloud objects. For instance, very bright land surfaces (e.g., airport runways and beach sand) may show consistently high values in the cloud-index time series, leading to a high threshold in Eq. (6). As a result, pixels of these bright land surfaces are not likely to be identified as clouds because their cloud index values are unlikely to exceed the high threshold. In addition, assuming that clouds are generally wider than a few pixels at Landsat spatial resolution, isolated pixels identified as being cloudy are removed from the cloud mask using a repeated minority analysis. We removed cloud pixels if 4 or fewer pixels in the 3-by-3 neighborhood of a pixel are cloud pixels. This step removes any remaining isolated, bright pixels in urban and coastal areas that are not clouds. Finally, similar to Fmask, all cloud patches are buffered with a width of 1 pixel to further reduce omission errors around cloud edges.

3.4. Estimate potential shadow zones

Shadow pixels are easily confused with dark objects and topographic shadow even in shadow index maps. However, clouds always accompany cloud shadows except at scene edges. This characteristic can help reduce commission (e.g., wet soil, topographic shadow) errors of cloud shadow detection. Actually, the location of cloud shadows can be calculated by the precise geometric relationship among clouds, cloud shadows and the position of the sun (Zhu and Woodcock, 2012). For a cloudy pixel with coordinates (x, y), the location of its corresponding shadow pixel (x', y') can be calculated using following equations (Luo et al., 2008):

$$x' = x - H \times \tan\theta \sin\phi \tag{10}$$

$$y' = y - H \times \tan\theta \cos\phi \tag{11}$$

where H is the height of clouds above the land surface, and θ and Φ are solar zenith and azimuth angles. Values of θ and Φ can be extracted from the image metadata files, but H is unknown for each cloud patch. In most existing methods, the brightness temperature (BT), derived from thermal infrared bands, is used to estimate cloud height with lapse rates for air temperature, such as -9.8K/Km for dry air and -6.5K/Km for moist air (Goodwin et al., 2013; Huang et al., 2010; Zhu and Woodcock, 2012). However, there are two problems with estimating cloud height when locating cloud shadows: (1) the lapse rate varies in different atmospheric conditions, and the BT of thin clouds is also influenced by the land surface; (2) some sensors, especially old ones, do not have thermal infrared bands, such as the Landsat MSS sensor, the CBERS IRMSS sensor and the Sentinel MSI sensors. For Landsat 8 also, a method for masking clouds and shadows that does not require a thermal band is needed. The Thermal Infrared Sensor (TIRS) has some error and intermittent availability and has a shorter design life than the multispectral Operational Land

Imager (OLI). That thermal data may not always be available is one obstacle to improving Landsat 8 cloud and shadow masks with image time series (Foga et al., 2017; Scaramuzza et al., 2012).

To make the proposed method able to process historical images without thermal infrared bands, a range of possible cloud heights are used to estimate all possible shadow locations of a cloud. We can use a default value of 200 m for minimum cloud heights because it is suitable for most areas (Zhu and Woodcock, 2012). The maximum cloud heights can be determined empirically by visually checking the maximum horizontal distance (D_{max}) between clouds and their shadows, or using 12 km based on previous studies (Fisher, 2013; Luo et al., 2008) :

$$H_{max} = \frac{D_{max}}{\sqrt{(\tan\theta \sin\phi)^2 + (\tan\theta \cos\phi)^2}} \tag{12}$$

Fig. 11b shows an example of potential shadow zones of a subset of image DOY146 in the Puerto Rico site. We can see that the real shadows are located within the potential shadow zones.

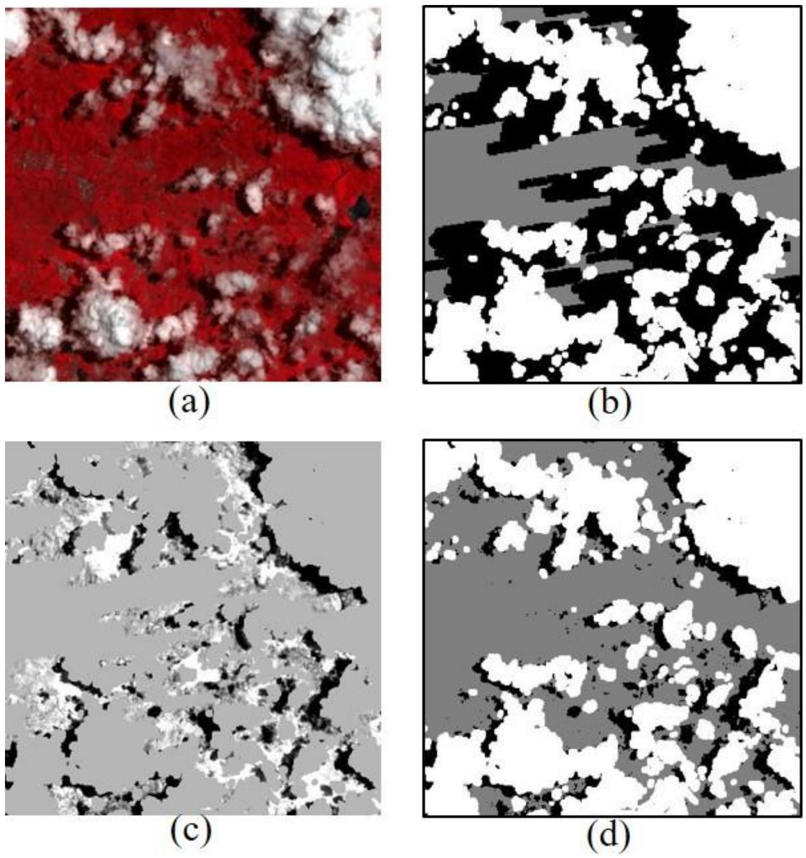


Fig. 11. A subset of the Landsat-8 image DOY146 in the Puerto Rico site (a), its potential shadow zones (b), shadow darkness as estimated by Inverse Distance Weighting (IDW) (c), and the initial shadow detected by K-means (d).

3.5. Detect shadow within potential shadow zones

The potential shadow zones mark the possible locations of cloud shadows. They overestimate the real shadow areas. Therefore, all the pixels within the shadow zones need to be further confirmed as to whether they are real shadow pixels. In the shadow index images, shadows are located at places with regional minima (i.e. “holes”) due to their being relatively dark in optical bands. Some existing methods use flood-fill transformation to predict the image without shadows and compare it with real images to identify shadow pixels (Li et al., 2017; Zhu and Woodcock, 2012). However, this approach may often mislabel dark objects, such as water, as cloud shadow (Li et al., 2017). Here, a similar idea is employed, but the new strategy reduces errors as compared with the flood-fill method. First, in the shadow index images, pixels in potential shadow zones are predicted from surrounding clear pixels with an inverse distance weighted (IDW) interpolator. Second, for those pixels in potential shadow zones, we estimate their “darkness” as their original shadow index minus the predicted values (Fig. 11c, a darker color means higher darkness). This darkness only shows how cloud shadows lower the pixel brightness compared with surrounding clear pixels. Third, similar to initial cloud detection, K-means clustering is applied to these darkened pixels (i.e. pixels with negative darkness values) to classify these pixels into two classes, clear observation and cloud shadow, to yield an initial shadow mask (Fig. 11d).

After the initial shadow detection, we apply a time series analysis, similar to the cloudy point refinement, to refine the initial shadow mask. This process aims to reduce both omission and commission errors in the initial shadow mask. Because cloud shadows have darkening effects, which lead to lower shadow index values in the time series of a pixel, a lower boundary L is used as a threshold to identify real shadow points. Considering differences in earth-sun-sensor geometry, atmospheric conditions and vegetation phenology, the shadow index of land surfaces needs to be normalized to minimize these differences prior to the time series analysis. Here, the histogram matching method is used given its simplicity (Helmer and Ruefenacht, 2005). Although histogram matching is a linear correction, and changes in vegetation phenology across an image can be nonlinear (Helmer and Ruefenacht, 2007), we found that histogram matching worked well for mitigating the temporal variability in shadow-index time series. First, the image with the fewest clouds in the time series is selected as a base image. Then, the shadow index of other images is normalized to this base image using the gain and bias:

$$gain = \frac{\sigma_B}{\sigma_t}$$
$$bias = \mu_B - \mu_t \times gain \tag{13}$$

where μ_B and μ_t are the mean value of clear pixels in the base image and the image at time t respectively, σ_B and σ_t are the standard deviations of clear pixels in the base image and the image at time t respectively. The normalized shadow index value of image at time t, $SI_N(i, t)$, can be computed as:

$$SI_N(i,t) = SI(i,t) \times gain + bias \tag{14}$$

This lower boundary L is defined using "good" points, which are those points not identified as shadow in the initial shadow mask (Fig. 12):

$$L(i) = \text{mean}\{SI_N(i,t)|(i,t) \in \text{"good"}\} - B \times \text{sd}\{SI_N(i,t)|(i,t) \in \text{"good"}\} \tag{15}$$

where B is a standard deviation multiplier that serves as a parameter to tune the threshold, $L(i)$ is the lower threshold for pixel i, *mean* is the mean shadow index (SI) of pixel i for the time series, and *sd* is the standard deviation of the SI for the time series of pixel i. Pixels with SI brighter than $L(i)$ are deemed too bright to be cloud shadow. The recommended value of B is from 1 to 3, and a larger value will select *darker* shadows, *i.e.,* it will darken the threshold for designating whether pixels are shadow. Therefore, the parameter B should be set to balance the omission and commission errors for shadow detection. For the initial shadow points, they are confirmed as real shadow if their shadow index values are lower than the mean value of "good" points. This step reduces the commission errors in initial shadow detection. For other points in the time series which are marked as potential shadow using sun-cloud geometry, they will be identified as final shadow points if their shadow index values are lower than L (Fig. 12). This step reduces the omission errors in the initial shadow detection. It should be noted that although the potential shadow zones and time series analysis can greatly prevent classifying topographic shadow as cloud shadow, topographic shadow within the potential shadow zones may be identified as cloud shadows if it is as dark as cloud shadow.

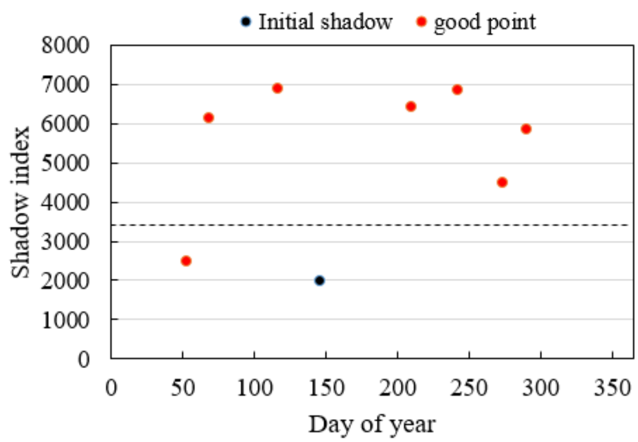
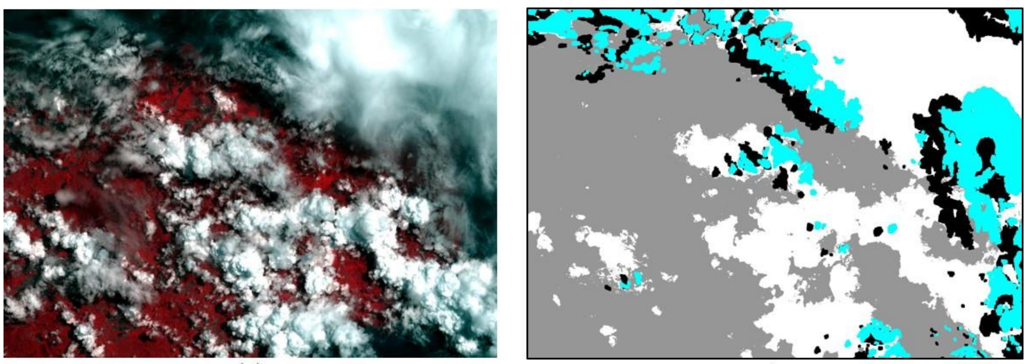


Fig. 12. An example of shadow index time series analysis: the points below the dashed line are identified as cloud shadow. The dashed line represents *L*, the lower threshold.

Similar to the cloud mask, isolated shadow pixels are also filtered out by a repeated minority analysis in a 3-by-3 neighborhood, and then all shadows are buffered, with a width of 1 pixel, to obtain the final shadow mask. The final shadow mask is combined with the final cloud mask to get the final product of cloud and shadow mask.

3.6. Evaluation and comparison

To demonstrate the accuracy and effectiveness of the proposed method, it was compared with Fmask (Zhu and Woodcock, 2012), one of the most advanced single-image methods and used by USGS to produce the standard cloud mask for Landsat images. The results of Fmask can be considered as a benchmark to assess the performance of ATSA. Both ATSA and Fmask were applied to Landsat-8 OLI and Sentinel-2 images, while only ATSA was applied to Landsat-4 MSS images in the second site, because MSS images lack not only thermal bands but also other bands that are needed by Fmask. We found that Fmask detected many clouds as snow in some images in the Hong Kong and Puerto Rico sites (Fig. 13). Because these two test sites are subtropical and never have snow, we merged snow into clouds before the comparison, but this adjustment was not made for cloud masks in Beijing site because it can snow in winter.



(a) (b)

Fig. 13. Landsat-8 image DOY178 in the Puerto Rico site (a) and its original Fmask cloud mask (b) showing where clouds are classified as snow (light blue color).

In the comparison, the agreement between these two methods was evaluated. First, the percentage of clouds and cloud shadows of all methods were plotted together to check their difference. Second, matrices were built comparing the proposed ATSA and Fmask methods, and the overall agreement derived from these matrices was used to assess the pixel-wise agreement between ATSA and Fmask. Third, representative images selected from the time series were digitized to produce reference cloud and shadow masks. The digitizing work was done by experienced experts who were not involved in the development of ATSA. Then, these digitized maps were used to quantitatively evaluate the accuracy of both methods. It should be noted that the manual mask of cloud and cloud shadow is not 100% accurate. It may include some commission or omission errors.

4. Results

4.1. Hong Kong site

Among 23 images to which we applied the two cloud and shadow masking methods, the two methods detect similar cloud cover for 19 of them (Fig. 14), while for the other 4 images (DOY 131, 179, 195, and 339) there are large differences. For the 19 images with similar cloud coverage, visual inspection confirms that both methods successfully detect clouds (see image of DOY 51 as an example in Fig. 16). For images of DOY 131 and 339, ATSA detected many more clouds than Fmask. On the other hand, for images of DOY 179 and 195, ATSA detected far fewer clouds than Fmask. Unlike cloud coverage, shadow coverage detected by the two methods slightly differs in most of the 23 images except the image DOY 355 (Fig. 14). Visual inspection of this image confirms that Fmask detected all water surface as cloud shadow. There are 5 images in the time series with large disagreement between ATSA and Fmask (Fig. 15). Agreement between ATSA and Fmask for the images of DOY 131, 179 and 339 is even lower than 50%.

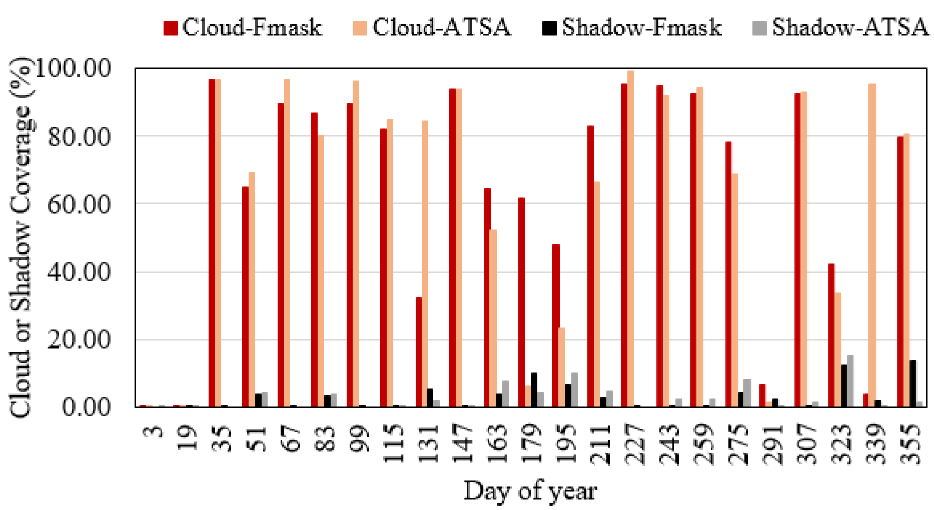


Fig. 14. Cloud and shadow coverage of 23 Landsat-8 OLI images in the Hong Kong site detected by ATSA and Fmask.

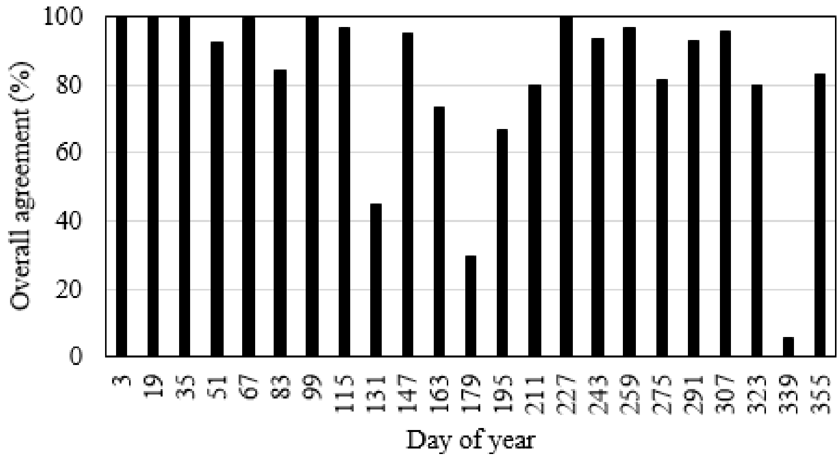


Fig. 15. Overall agreement of cloud and shadow masks of 23 Landsat-8 OLI images in the Hong Kong site between ATSA and Fmask.

In the cloud masks of the three images with the least agreement between ATSA and Fmask, it is clear from Fig. 16 that ATSA more accurately identified clouds. Fmask underestimated clouds in two images of DOY 131 and 339, and it overestimated clouds in the image of DOY 179. Specifically, Fmask failed to screen many of the thin clouds in the center of the image of DOY 131, and it failed to identify many of the thick clouds in the image of DOY339, even though these clouds appear very bright in all visible and NIR bands. In the image of DOY179, Fmask misidentified most of the clear water and some clear land surface (see the island in the lower right) as clouds, which led to serious overestimation of cloud cover. For the cloud shadows, it appears that ATSA successfully identified most shadows adjacent to clouds. Fmask identified some clear pixels as shadow that were near the misidentified cloud patches (see image of DOY 179 in Fig. 16).

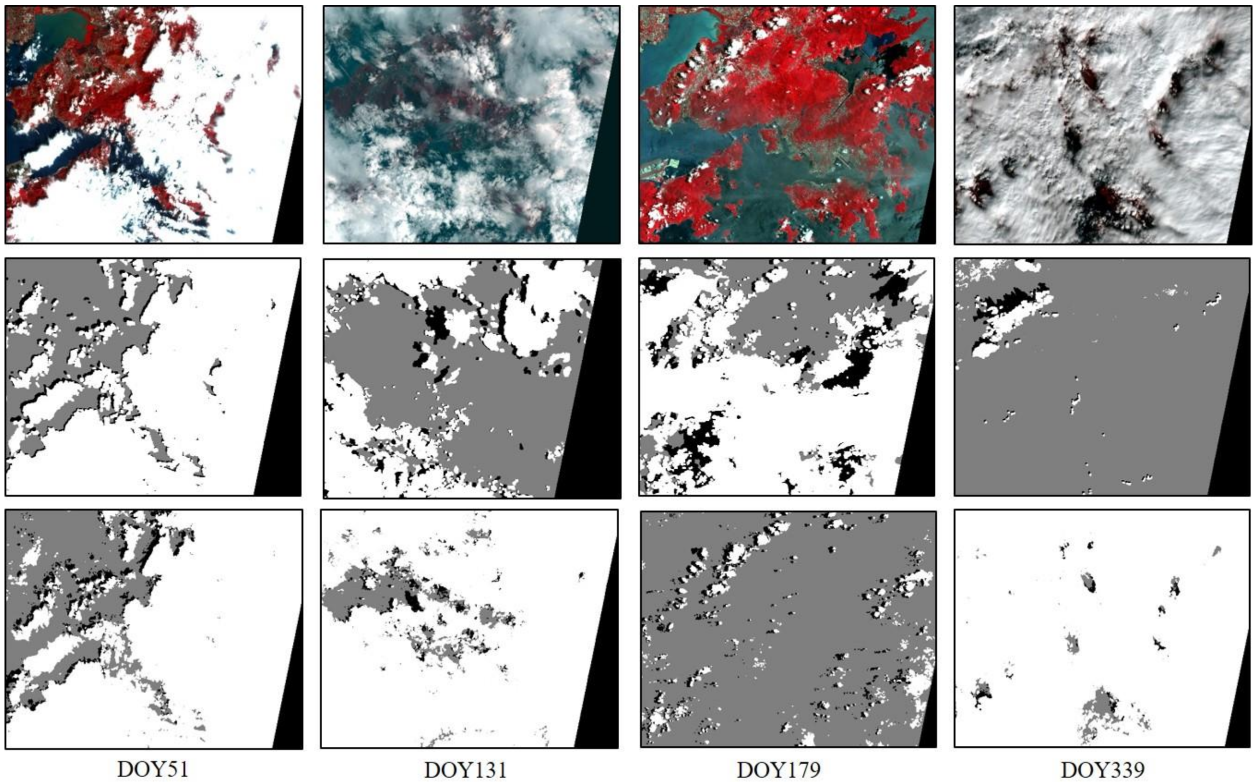


Fig. 16. False color composite of selected Landsat images (upper row) and their cloud masks by Fmask (middle row) and ATSA (lower row) for the Hong Kong site (gray: clear pixels; black: shadows; white: clouds)

Quantitative accuracy assessment for the four images in Fig. 16 using manual masks shows that ATSA and Fmask obtain comparable overall accuracy for the image of DOY51, but ATSA's overall accuracy is much higher than Fmask for the other three images (Table 4). For cloud detection, ATSA obtained user's accuracies ranging from 0.85 to 0.99 and producer's accuracies ranging from 0.89 to 0.99. The accuracy of ATSA cloud mask for the image of DOY179 is lower than that of the other images due to the errors in haze detection on the water surface (see haze in the lower right part of this image). In contrast, the cloud producer's accuracy of Fmask is low for images of DOY131 (0.41) and DOY339 (0.04) because of large omission errors. The cloud user's accuracy of Fmask is low for images of DOY179 (0.08) because it misidentified many clear pixels as clouds. For shadow detection, ATSA can obtain producer's accuracy higher than 0.82. The shadow user's accuracy is also high except the image of DOY179, in which ATSA overestimated the shadow area. In the context of applications with cloudy images, the producer's accuracy is more important than user's accuracy, because end users hope to exclude all contaminated pixels in their analysis, and meanwhile they can allow commission errors to some extent (Zhu and Woodcock, 2012). Both user's and producer's accuracies of shadow detection by Fmask are much lower than ATSA. Specifically, Fmask detected fewer shadows in the image of DOY51 and identified clear pixels near the wrong cloud patches as shadow in other three images in Fig. 16.

Table 4 Accuracy assessment of cloud masks of the 4 images in Fig. 16 in the Hong Kong site: overall accuracy (*oa*), user's accuracy (*ua*) and producer's accuracy (*pa*).

			Cloud		Shadow	
DOY		*oa*	*Ua*	*pa*	*ua*	*pa*
51	Fmask	0.93	0.97	0.98	0.60	0.49
	ATSA	0.99	0.99	0.99	0.95	0.87
131	Fmask	0.45	0.99	0.41	0.04	0.10
	ATSA	0.98	0.99	0.99	0.93	0.82
179	Fmask	0.29	0.08	0.98	0.08	0.26
	ATSA	0.97	0.85	0.89	0.67	0.90
339	Fmask	0.06	0.95	0.04	0.00	0.00
	ATSA	0.99	0.99	1.00	0.89	1.00

4.2. Puerto Rico site

4.2.1. Landsat-8 OLI images

Among the 18 images, ATSA and Fmask obtained similar cloud coverage in 15 images, while ATSA detected many more clouds in three images (DOY178(2013), 306(2013), and 053(2014)) than Fmask (Fig. 17). For shadow coverage, ATSA detected slightly more shadows than Fmask in most images. Through visual inspection of these images, we found that Fmask underestimated shadows surrounding small cloud patches, which leads to smaller shadow percentage than ATSA. On the other hand, in the images DOY 274(2013), 354(2013), 53(2014), and 149(2014), ATSA detected fewer shadows than Fmask. Visual inspection shows that these four images only have large cloud patches. Fmask overestimated shadow cover of these large cloud patches. The quantitative assessment of pixel-wise agreement between the two methods is good (higher than 80%) for masks of most images, but the masks of two images, DOY 178(2013) and 306(2013) have agreement between ATSA and Fmask that is lower than 60% (Fig. 18). For the images with good agreement between ATSA and Fmask, both methods successfully detect clouds (e.g. image DOY146(2013) in Fig. 19). In the two images with the least agreement between ATSA and Fmask, Fmask omitted a lot of thin clouds in west region in the image of DOY178(2013), and it missed a lot of cloudy pixels, even of thick clouds, in the image of DOY306(2013) (Fig. 19).

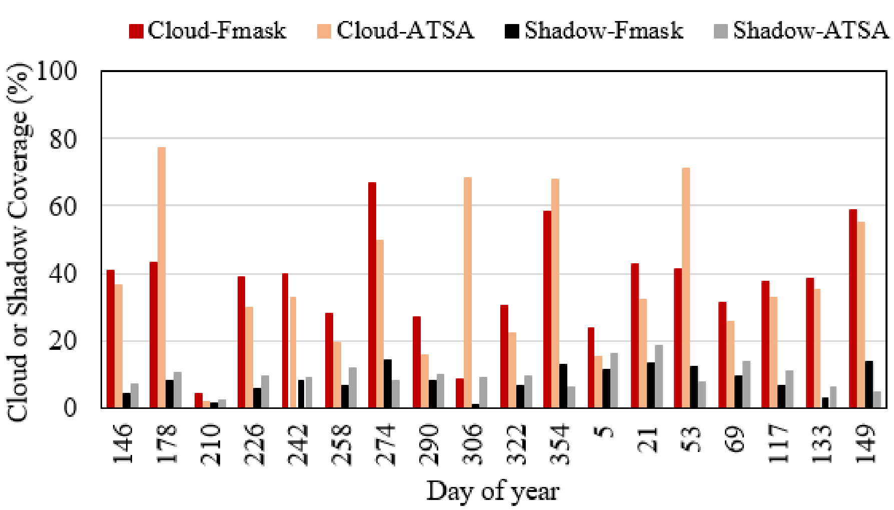


Fig. 17. Cloud and shadow coverage of 18 Landsat-8 OLI images in the Puerto Rico site detected by ATSA and Fmask

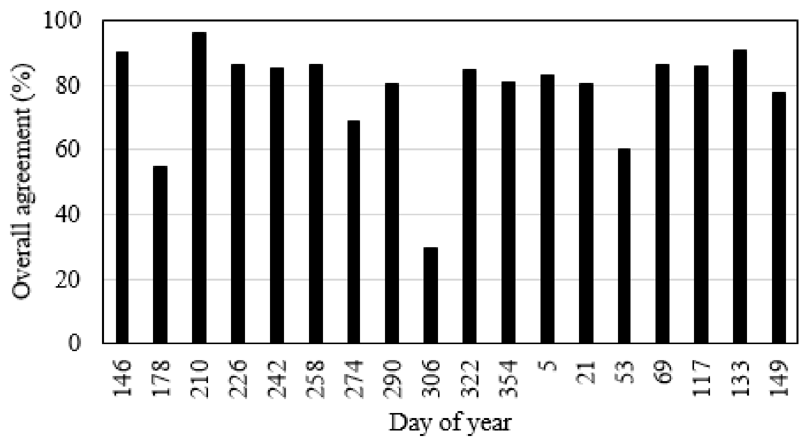


Fig. 18. Overall agreement of cloud and shadow mask between ATSA and Fmask for the Puerto Rico site.

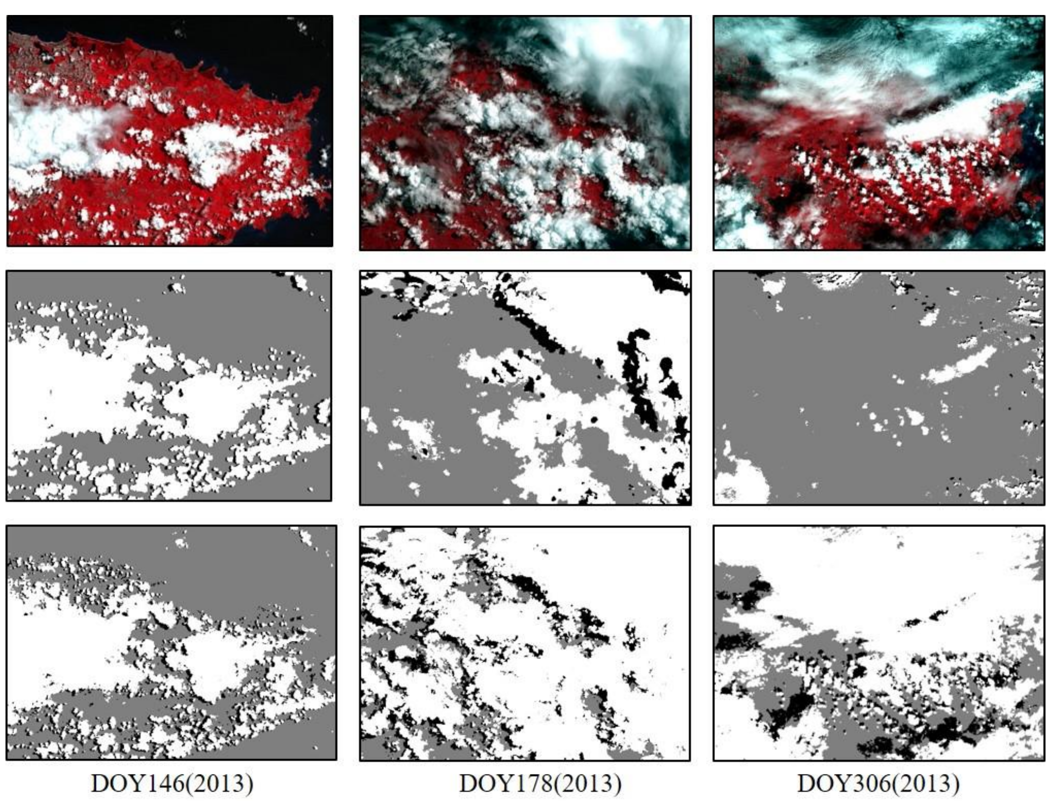


Fig. 19. False color composite of the three Landsat images in the Puerto Rico site (upper row)

and their cloud and shadow masks by Fmask (middle row) and ATSA (lower row) (gray: clear pixels; black: shadows; white: clouds)

The quantitative accuracy assessment of the cloud masks of these three images in Fig. 19, using manual masks, shows that the overall accuracy of ATSA cloud and shadow masks ranges from 0.97 to 0.98, which is much higher than Fmask (Table 5). Cloud producer's and user's accuracy of ATSA reaches 0.97 in all three images, while cloud producer's accuracies of Fmask are only 0.12 to 0.52 for the image of DOY178 and 306. Shadow producer's and user's accuracies of ATSA are lower than the cloud mask accuracy, but it is still much higher than Fmask. ATSA omitted some thin shadows on land surfaces in the lower part of image of DOY306 (Fig. 19) leading to a relatively lower producer's accuracy of 0.86. Similar to the Hong Kong site, Fmask detected fewer shadows than the real situation, leading to low producer's accuracy in shadow detection.

Table 5. Accuracy assessment of cloud masks of three images in Fig. 19 in the Puerto Rico site: overall accuracy (*oa*), user's accuracy (*ua*) and producer's accuracy (*pa*).

			Cloud		Shadow	
DOY(Year)		*oa*	*ua*	*pa*	*ua*	*pa*
146(2013)	Fmask	0.90	0.85	0.98	0.58	0.34
	ATSA	0.98	0.97	1.00	0.92	0.94
178(2013)	Fmask	0.53	0.98	0.52	0.07	0.08
	ATSA	0.98	0.98	0.99	0.97	0.96
306(2013)	Fmask	0.28	0.98	0.12	0.03	0.00
	ATSA	0.97	1.00	0.98	0.97	0.86

4.2.2. Landsat-4 MSS images

Only ATSA was applied to the 11 Landsat-4 MSS images in the Puerto Rico site, because Fmask needs more bands than MSS images have. Four images with representative cloud coverages

(6.8% to 99% clouds; 0% to 12.6% cloud shadows) were selected for further assessment (Fig. 20). The cloud coverages of these four selected images are 6.8% (DOY40), 31.1% (DOY280), 41.4% (DOY24), and 66.1% (DOY200). Visual inspection shows that ATSA successfully identified most clouds and shadows in these MSS images (Fig. 21), including the thin clouds in the image of DOY200. We can see that the cloud user's and producer's accuracy of all four MSS images are higher than 0.95 (Table 6), indicating that ATSA successfully screened clouds in these images with very small omission and commission errors. For shadow accuracy, in terms of producer's accuracy (more important for applications in our opinion), it is high enough in image DOY 40, and 280, reaching 0.97. The image DOY200 has shadow producer's accuracy of 0.83 which is caused by the identification of shadows as clouds in the lower right part of this image (Fig. 21).

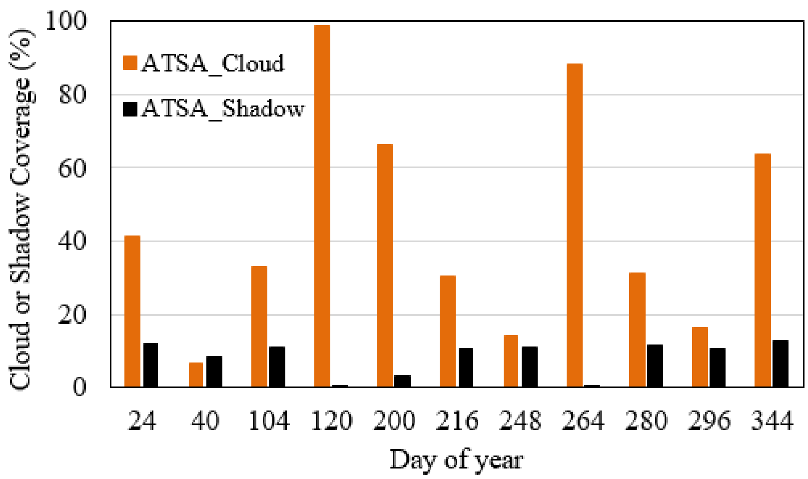


Fig. 20. Cloud and shadow coverage of 11 Landsat MSS images in the Puerto Rico site detected by ATSA.

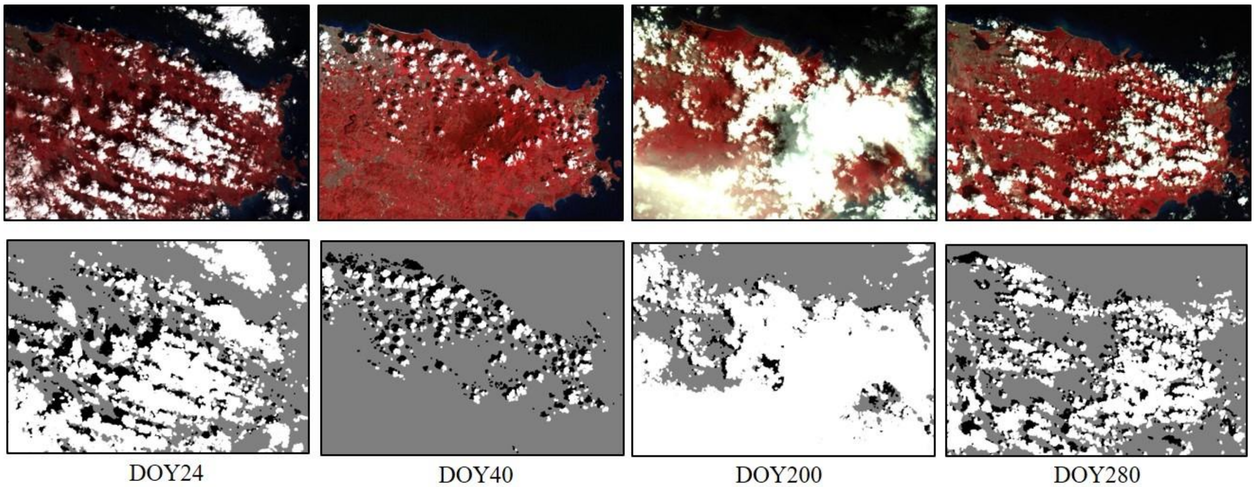


Fig. 21. False color composite of the four representative Landsat MSS images in the Puerto Rico site and their cloud and shadow masks by ATSA (gray: clear pixels; black: shadows; white: clouds).

Table 6. Accuracy assessment of cloud masks of four MSS images in Fig. 21 in the Puerto Rico site: overall accuracy (*oa*), user's accuracy (*ua*) and producer's accuracy (*pa*)

		Cloud		Shadow	
DOY	*oa*	*Ua*	*pa*	*ua*	*pa*
24	0.96	0.98	0.99	0.88	0.87
40	0.99	1.00	0.99	0.89	1.00
200	0.97	0.99	0.97	0.90	0.83
280	0.98	1.00	0.98	0.92	0.97

4.3. Beijing site

In general, the cloud coverage detected by ATSA is smaller than that of Fmask in the 20 Sentinel-2 images. The cloud coverage difference between ATSA and Fmask is larger than or equal to 20% in three images, Aug. 21, Sep.30, and Oct.10 (Fig. 22). Through visual inspection of these cloud masks, we found that both ATSA and Fmask misclassify some pixels in very bright urban surfaces as clouds, but this commission error of Fmask is more serious than ATSA. Fmask detected nearly all urban pixels as cloud or snow in the three images of Aug.21, Sep.30, and Oct.10. Similar to cloud coverage, shadow coverage detected by ATSA is generally lower than Fmask, except for images dated Oct. 30 and Nov. 29. The larger shadow coverage detected by Fmask results from the commission errors of cloud detection. In other words, Fmask detected many clear pixels as shadow surrounding areas wrongly detected as clouds. Regarding the pixel-wise agreement between ATSA and Fmask (Fig. 23), there are 7 images with overall agreement lower than 80% and 3 images lower than 70%. These three images are Jan.26, Aug.11, and Sep.30.

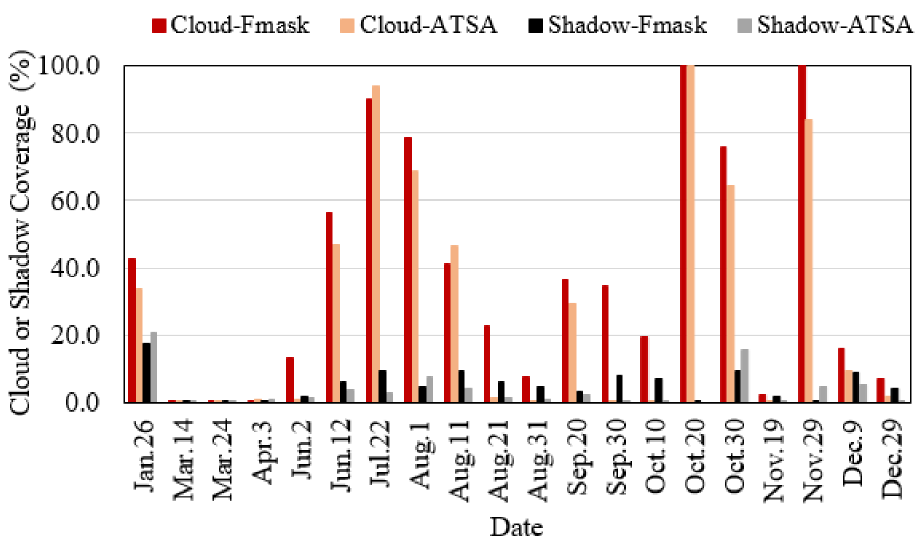


Fig. 22. Cloud and cloud shadow coverage of 20 Sentinel-2 images in the Beijing site detected by ATSA and Fmask.

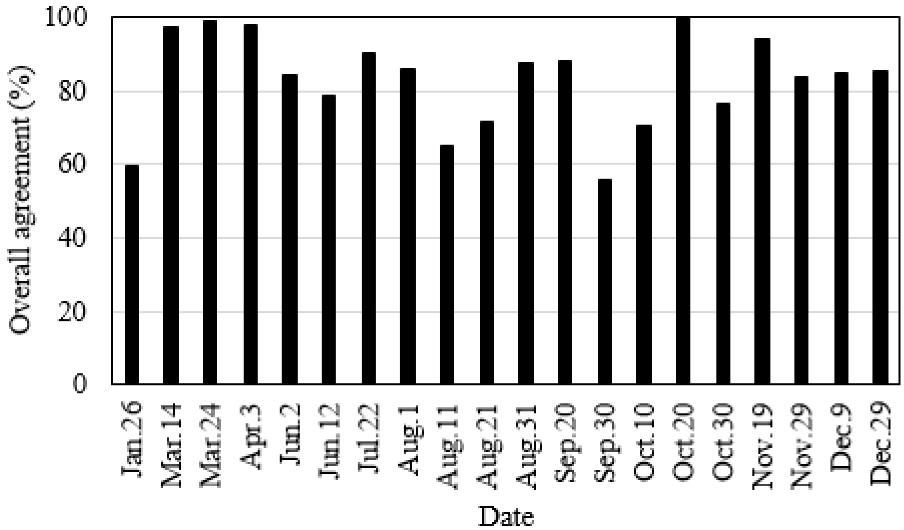


Fig. 23. Overall agreement of cloud and shadow mask between ATSA and Fmask for the 20 Sentinel-2 images from the Bejing site.

For the images with high agreement between ATSA and Fmask, both methods successfully detect clouds and shadows (see Sep.20 image as an example in Fig. 24). In the three images with the least agreement between ATSA and Fmask, ATSA is generally more successful than Fmask for identifying clouds and shadows. Specifically, in the Jan. 26 image, both ATSA and Fmask detected most of clouds. ATSA does not have snow detection step, so the snow in the northwest was misclassified as cloud in ATSA, while Fmask successfully detected these snows but it also identified many clear pixels as snow. Fmask also detected many clear and thin-cloudy pixels as snow in summer image where no snow events should happen. In addition, Fmask detected many clear urban pixels as cloud in the Sep. 30 image (Fig. 24). For cloud shadow detection, ATSA is more successful although it detected topographic shadows as cloud shadows in the Jan. 26 image. Images of mountainous areas have more topographic shadows in spring and winter due to the lower sun elevation. It may lead to larger commission errors in cloud shadow detection if these topographic shadows are within the potential shadow zones. In contrast, Fmask failed to detect shadows which are distant from the cloud patches. The possible reason is that Fmask for Sentinel-2 assumes the cloud height between 200 m and 1,200 m for all images (Zhu et al., 2015).

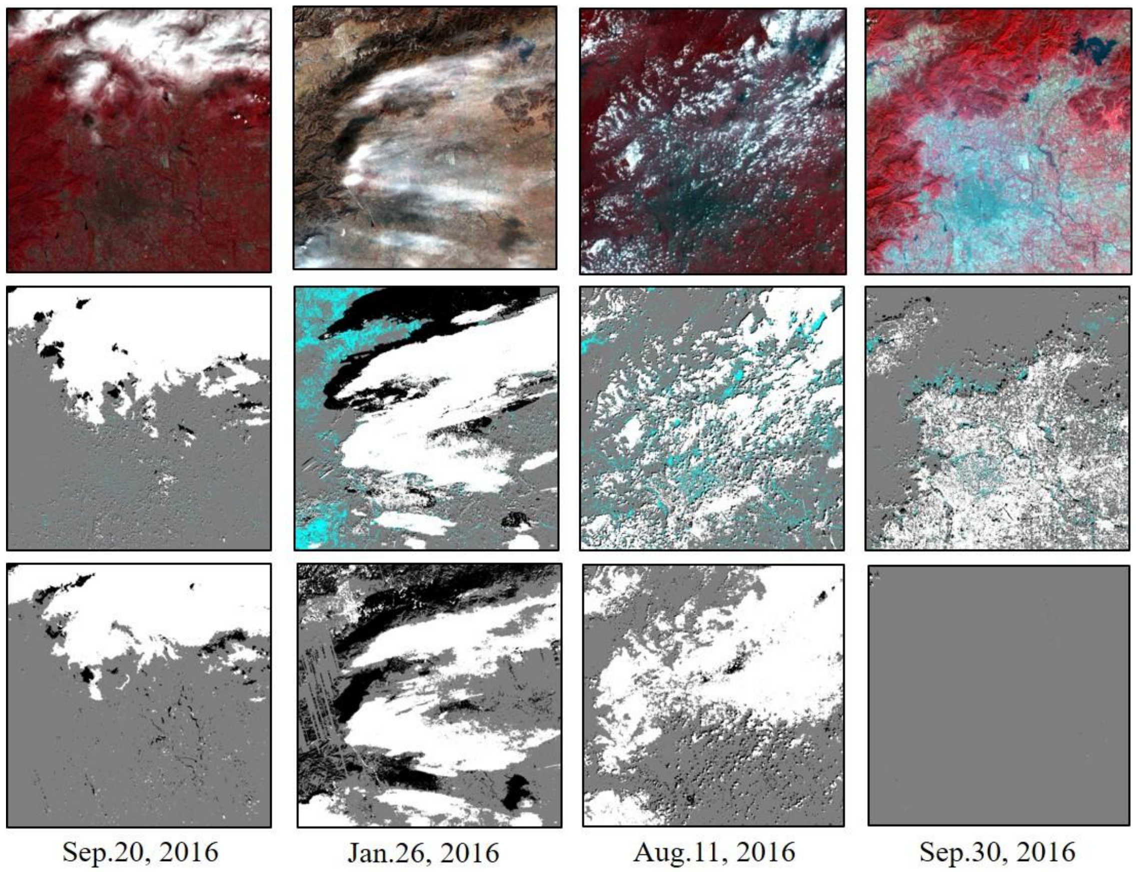


Fig. 24. False color composite of the four Sentinel-2 images in the Beijing site (upper row) and their cloud mask by Fmask (middle row) and ATSA (lower row) (gray: clear pixels; black: shadows; white: clouds; snow: light blue)

The quantitative accuracy assessment also demonstrates that ATSA can obtain more accurate cloud and shadow masks than Fmask (Table 7). For the Sep. 20 image, both ATSA and Fmask can obtain acceptable accuracy in cloud detection. For other three images, cloud producer's accuracy of ATSA ranges from 0.81 to 0.96 and cloud user's accuracy ranges from 0.92 to 0.99. In contrast, cloud user's accuracy of Fmask for the Sep.30 image is very low because of large commission errors, and the low cloud producer’s accuracy of Fmask for the Aug.11 image is caused by large omission errors. For the Jan. 26 image, the producer’s accuracy of cloud detection for both Fmask and ATSA is only 0.81 because both methods omitted extremely thin clouds. For shadow detection, ATSA obtains good producer’s accuracy ranging from 0.81 to 0.96, which is much higher than Fmask ranging from 0.20 to 0.50, indicating that Fmask omitted considerable cloud shadow in these images. For the Jan. 26 image, the user’s accuracy of shadow detection by ATSA is only 0.5, because it detects many black rocks and topographic shadows as cloud shadows. In this mountainous area, some snow and ice pixels were misclassified as clouds, which makes the black rocks and topographic shadows within the potential shadow zone. This issue can be solved if the commission error in cloud detection is reduced, especially for distinguishing snow and ice from clouds.

Table 7. Accuracy assessment of cloud masks of images in Fig. 24: overall accuracy (*oa*), user’s accuracy (*ua*) and producer’s accuracy (*pa*).

			Cloud		Shadow	
Date		*oa*	*ua*	*pa*	*ua*	*pa*
Sep.20	Fmask	0.89	0.80	1.00	0.20	0.41
	ATSA	0.98	0.99	1.00	0.62	0.81
Jan.26	Fmask	0.67	0.74	0.81	0.33	0.50
	ATSA	0.79	0.92	0.81	0.50	0.87
Aug.11	Fmask	0.68	0.77	0.65	0.21	0.46
	ATSA	0.97	0.99	0.96	0.94	0.86
Sep.30	Fmask	0.58	0.00	0.92	0.00	0.20
	ATSA	1.00	0.99	0.85	0.96	0.96

5. Discussion and conclusions

Masking clouds and cloud shadows is necessary for many applications of optical satellite images, because it is difficult to acquire totally cloud-free images in most places, particularly when time series are needed to monitor change. Many methods have been developed to screen clouds and cloud shadows automatically in optical images. However, they may not perform well in very cloudy regions. Aiming to produce more accurate cloud and shadow masks of optical imagery in cloudy regions, an automatic time series analysis based method, ATSA, was developed in this study. ATSA was tested in three sites with different dominant land covers. Landsat-8 OLI images, Landsat-4 MSS images, and Sentinel-2 images were used to evaluate the performance of ATSA for screening clouds and cloud shadows in images with different band configurations and quality. Results show that ATSA can obtain accurate cloud and shadow masks in all sites and all data sets except the images with snow and ice cover. The comparison with an advanced algorithm, Fmask, also confirms that ATSA can yield robust and accurate cloud and shadow masks in cloudy regions. The good performance of ATSA can be attributed to the following strengths.

First, ATSA only needs a minimum number of input bands. Only 5 bands, blue, green, red, NIR, and SWIR bands, are required, and this requirement can be reduced to 3 bands if the images do not have blue and SWIR bands. The low requirement of input bands brings two advantages. The first advantage is that in the regions tested, the results can be more robust than existing methods when processing images with various conditions. Although the spectral similarities among different land surfaces and clouds and cloud shadows are complex, being different among locations and times, a common characteristic is that they affect the pixel values from visual to near infrared bands, i.e., clouds brighten these bands and shadows darken them. In general, adding more bands into the screening process, such as thermal bands or a cirrus band, can improve the accuracy of cloud and shadow masks, especially for the single-image cloud detection methods (Foga et al., 2017; Zhu et al., 2015). However, it may also lead to more uncertainties and errors in some extreme cases. For example, Fmask uses visible, near infrared, and thermal bands to identify clouds; it also uses the cirrus band in Landsat-8 images to detect clouds (Zhu et al., 2015), while ATSA uses neither the thermal nor cirrus bands. In the Hong Kong site, for the image DOY339, Fmask misses most clouds. A further investigation of all bands of this image reveals that the thermal band is cooler in a small sub-area of this mostly cloudy image (Fig. 25). As a result, Fmask only detects clouds in this cold area and omits other warmer clouds. According to the USGS product guide, Fmask has a known issue that either too large or too small temperature differentials will lead to errors in cloud detection. The second advantage of using fewer bands is that the algorithm is more flexible and applicable than existing methods when processing images from different optical sensors. For cloudy places, we expect ATSA to: (1) extend the history for automated Landsat time series analyses with cloud and cloud shadow masks that are highly accurate, but automatically derived, back to the MSS era of the 1970s (instead of only the TM era of the 1980s); and (2) in the era of Sentinal-2, allow for denser time series in intra-annual analyses such as those examining vegetation phenology. The past and ongoing optical sensors have different configurations of spectral bands. However, most of these optical sensors have visible and near infrared bands. ATSA can be applied to all images with these basic bands, which is very important and necessary when we process historical satellite images with limited bands.

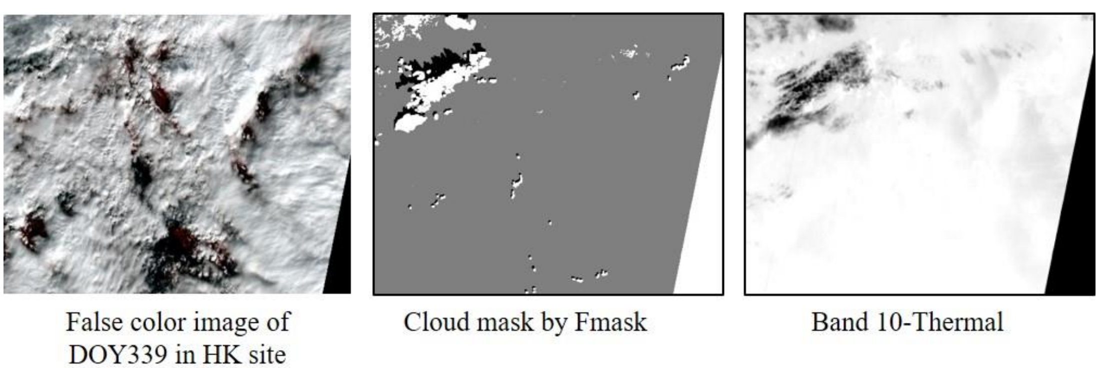


Fig. 25. False color composite of the Landsat-8 images DOY339 in the Hong Kong site (left), its cloud mask by Fmask (center), and the thermal band of this image (right)

Second, ATSA has fewer predefined parameters than most existing methods. In ATSA, there are only two important predefined parameters, i.e., A in Eq. (8) and B in Eq. (12) tune the threshold for identifying clouds and shadows in the time series respectively. As standard deviation multipliers of variation through the time series, these two parameters regulate the degree of strictness for masking clouds and shadows. In other words, they balance the omission errors and commission errors of cloud and shadow detection. In our tests, A and B are 0.5 and 1.5 for the Hong Kong site, 1.0 and 1.5 for the Puerto Rico site, and 1.2 and 2.0 for the Beijing site. Fig. 26 shows the cloud user's and producer's accuracy for the Landsat-4 MSS image DOY200 in Fig. 21 when using different values of parameter A within the recommended range 0.5-1.5. Larger values of parameter A improve the user's accuracy but meanwhile decrease the producer's accuracy. The parameter B shows a similar effect on the accuracy of cloud shadow detection. Fig. 26 also suggests that the detection accuracy is not very sensitive to the parameter. There is a wide range of parameter A able to obtain both producer's accuracy and user's accuracy higher than 0.95. Users can tune these two parameters according to their specific applications. For example, studies using time series to model land surface parameters, such as forest biomass and crop yield, are very sensitive to clouds, even the extremely thin clouds. These studies may hope to mask out all possible clouds and accept some commission errors, so smaller values of parameter A and B should be used. In addition, ATSA also use the statistics of each image in the time series to determine some parameters to increase the adaptability of ATSA. For example, the HOT transformation has been used in many cloud screening methods, such as MFC (Li et al., 2017) and Fmask (Zhu and Woodcock, 2012). However, these methods apply one HOT formula to all images. For instance, both MFC and Fmask use $HOT=B_{blue}-0.5B_{red}$ for all images. However, the coefficients in the HOT transformation vary from scene-to-scene, so it is necessary to estimate the HOT parameter for individual images (Chen et al., 2016; Zhang et al., 2002). ATSA regresses the coefficients in HOT transformation model in each image by an automated strategy which can get optimal cloud index images.

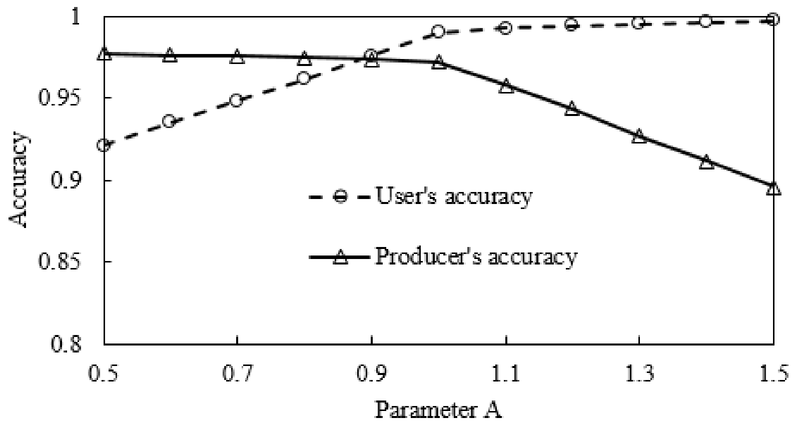


Fig. 26. User’s and producer’s accuracy of cloud detection for the Landsat-4 MSS image DOY200 in the Puerto Rico site using different values of parameter A

Third, ATSA uses the minimal clear observations in image time series over cloudy regions to ensure accurate cloud and shadow masking without fitting a time series model of these observations. For both cloud and shadow detection, there are two hierarchies in ATSA. In the first hierarchy, ATSA selects samples from all images in the time series for optimizing the class centers in the K-means classifier. As we know, it is quite common that image scenes are totally covered by clouds. If the K-means classifier (K=2 or 3) is applied to each individual image, it cannot detect all clouds in a totally cloud-covered image. In the second hierarchy, ATSA only uses “clear” observations in the time series to estimate the adaptive threshold, and further detect clouds and shadows omitted in the first hierarchy. Another multi-temporal method, Tmask, also uses clear observations in the time series to refine the initial cloud mask from Fmask. It can detect more thin clouds than Fmask (Zhu and Woodcock, 2014). However, Tmask is not appropriate in our test sites in cloudy regions. Fig. 27 shows the number of clear observations of individual pixels in the time-series data of Hong Kong and Puerto Rico site. We can see that both sites have considerable pixels with fewer than 6 clear observations. The clear observations are not enough for Tmask (15 clear observations are recommended) to accurately estimate the parameters in the time series model, leading to misclassifying cloudy pixels as clear pixels (Foga et al., 2017). Unlike Tmask, ATSA does not fit a time series model using many clear observations. It can be an alternative to Tmask for screening clouds in time-series data of cloudy regions or short time series (e.g., one year) which is unlikely to have enough clear observations.

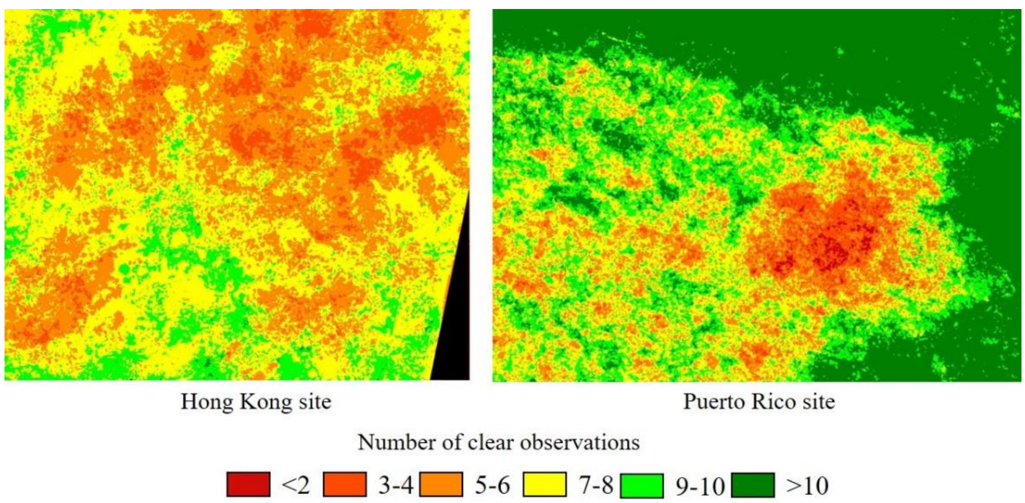


Fig. 27. Number of clear observations of individual pixel in the Landsat-8 time series at both Hong Kong and Puerto Rico sites.

There are also some limitations in ATSA. First, ATSA currently does not have a snow detection module. In tropical and subtropical regions, which are among the cloudiest regions (Ju and Roy, 2008), images in these regions do not have snow in all seasons except at the highest elevations. If the images include snow, ATSA is likely to detect snow as clouds (see Jan.26 image in Fig. 24). This outcome may be acceptable in many applications, such as vegetation studies, in which, like clouds, snow would often be excluded. Actually, most current algorithms often confuse snow and clouds even if they have a snow detection module, like the Fmask results shown in Fig. 13 and Fig. 24. If more powerful snow detection methods are developed in the future, they can be integrated with ATSA. Second, although the HOT transformation can suppress the pixel values of various land covers (also see an experiment in a desert landscape shown in the Supplementary Data), the very bright pixels may be identified as clouds. A recent study proposed an iterative HOT (IHOT) algorithm which can better suppress surface reflectance (Chen et al., 2016), but it needs more computing time. IHOT can be used as an alternative to HOT if the computing time is not a restriction factor. Third, land cover changes may happen in the time series. It may bring temporal variability in the HOT time series which could further affect the cloud detection by ATSA. An experiment reported in the Supplementary Data shows that ATSA may be not affected by many types of land cover changes, but other methods (e.g. Tmask) which can model land cover change may obtain better results than ATSA when substantial land cover changes exist. Fourth, ATSA may omit some cloud shadows on water surfaces or cloud shadows on the land surface that are extremely thin. Omission of cloud shadows on water surfaces may not affect mapping the water bodies, but it may affect water quality modeling. Thin cloud shadow on land surfaces may also affect quantitative information retrieval. Omission errors from missed cloudy pixels are the most common errors in cloud shadow masking methods (Foga et al., 2017); however, more accurate cloud detection with ATSA in the types of landscapes tested should reduce this error. These errors can be corrected by a further manual checking. Fifth, ATSA requires a time series, albeit with fewer dates than existing methods. Last, due to the limitation of resources and support, ATSA was tested in several typical sites and on data sets from three satellite sensors. More comparison and validation are needed, and they are our future studies. Due to its simple principles, ATSA has an acceptable efficiency for processing time-series data. ATSA only used 11 minutes and 13 minutes for the Landsat-8 time series in Hong Kong and Puerto Rico sites respectively (program coded in interactive data language and run on a windows laptop with a 2.50GHz CPU and 8 GB RAM). We welcome other researchers to test ATSA in more areas and different data sets. The code of ATSA is available upon request.

In conclusion, a new cloud and cloud shadow screening method, ATSA, was developed in this study. Its target is time series optical images in cloudy regions. ATSA is a valuable supplement to the family of cloud and cloud shadow masking algorithms. It will support studies of land surface dynamics using dense optical time series, such as studies of forest phenology in tropical regions using Landsat or Sentinel-2 images.

Acknowledgements

This study was supported by the Research Grants Council of Hong Kong (project no.25222717), the National Natural Science Foundation of China (project no.41701378), the grant 1-ZE6Q from The Hong Kong Polytechnic University, the USDA Forest Service International Institute of Tropical Forestry (Cooperative Agreement 13-CA-11120101-029), the Southern Research Station Forest Inventory and Analysis Program, and the National Science Foundation of the USA (1010314). We thank Dr. Zhe Zhu at Texas Tech University for sharing the Fmask code and for his assistance during the implementation of Fmask. We thank Yujing Chen, Jiaqi Tian, and Zheyan Shen for creating cloud reference data. We also thank Ronn Kling for his earlier input on cloud and cloud shadow masking.

Appendix A. Supplementary data

Supplementary data to this article can be found online at ##.

References

Bassett, G.J., Koenker, R., 1978. Asymptotic Theory of Least Absolute Error Regression. J. Am. Stat. Assoc. 73, 618–622. doi:10.2307/2286611

Cahalan, R.., Oreopoulos, L., Wen, G., Marshak, A., Tsay, S.-C., DeFelice, T., 2001. Cloud characterization and clear-sky correction from Landsat-7. Remote Sens. Environ. 78, 83–98. doi:10.1016/S0034-4257(01)00251-6

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., Mills, J., 2015. Global land cover mapping at 30m resolution: A POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27. doi:10.1016/j.isprsjprs.2014.09.002

Chen, S., Chen, X., Chen, J., Jia, P., Cao, X., Liu, C., 2016. An Iterative Haze Optimized Transformation for Automatic Cloud/Haze Detection of Landsat Imagery. IEEE Trans. Geosci. Remote Sens. 54, 2682–2694. doi:10.1109/TGRS.2015.2504369

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. doi:10.1016/j.rse.2004.03.007

Fisher, A., 2013. Cloud and cloud-shadow detection in SPOT5 HRG imagery with automated morphological feature extraction. Remote Sens. 6, 776–800. doi:10.3390/rs6010776

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. doi:10.1016/j.rse.2017.03.026

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. doi:10.1016/j.rse.2013.02.019

Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. Remote Sens. Environ. 114, 1747–1755. doi:10.1016/j.rse.2010.03.002

Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. 122, 66–74. doi:10.1016/j.rse.2011.08.024

Helmer, E.H., Ruefenacht, B., 2007. A comparison of radiometric normalization methods when filling cloud gaps in Landsat imagery. Can. J. Remote Sens. 33, 325–340. doi:10.5589/m07-028

Helmer, E.H., Ruefenacht, B., 2005. Cloud-free satellite image mosaics with regression trees and histogram matching. Photogramm. Eng. Remote Sens. 71, 1079–1089. doi:10.14358/PERS.71.9.1079

Helmer, E.H., Ruzycki, T.S., Benner, J., Voggesser, S.M., Scobie, B.P., Park, C., Fanning, D.W., Ramnarine, S., 2012. Detailed maps of tropical forest types are within reach: Forest tree communities for Trinidad and Tobago mapped with multiseason Landsat and multiseason fine-resolution imagery. For. Ecol. Manage. 279, 147–166. doi:10.1016/j.foreco.2012.05.016

Huang, C., Thomas, N., Goward, S.N., Masek, J.G., Zhu, Z., Townshend, J.R.G., Vogelmann, J.E., 2010. Automated masking of cloud and cloud shadow for forest change analysis using Landsat images. Int. J. Remote Sens. 31, 5449–5464. doi: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. doi:10.3390/rs6064907

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. doi:10.14358/PERS.72.10.1179

Jin, S., Homer, C., Yang, L., Xian, G., Fry, J., Danielson, P., Townsend, P.A., 2013. Automated cloud and shadow detection and filling using two-date Landsat imagery in the USA. Int. J. Remote Sens. 34, 1540–1560. doi:10.1080/01431161.2012.720045

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. doi:10.1016/j.rse.2007.08.011

Li, P., Dong, L., Xiao, H., Xu, M., 2015. A cloud image detection method based on SVM vector machine. Neurocomputing 169, 34–42. doi:10.1016/j.neucom.2014.09.102

Li, Z., Shen, H., Li, H., Xia, G., Gamba, P., Zhang, L., 2017. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. Remote Sens. Environ. 191, 342–358. doi:10.1016/j.rse.2017.01.026

Lloyd, S.P., 1982. Least Squares Quantization in PCM. IEEE Trans. Inf. Theory 28, 129–137. doi:10.1109/TIT.1982.1056489

Luo, Y., Trishchenko, A.P., Khlopenkov, K. V., 2008. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. Remote Sens. Environ. 112, 4167–4185. doi:10.1016/j.rse.2008.06.010

Lyapustin, A.I., Wang, Y., Frey, R., 2008. An automatic cloud mask algorithm based on time series of MODIS measurements. J. Geophys. Res. Atmos. 113. doi:10.1029/2007JD009641

Martinuzzi, S., Gould, W., González, O., 2007. Creating cloud-free Landsat ETM+ data sets in tropical landscapes: cloud and cloud-shadow removal. Gen. Tech. Rep. IITF-GTR-32 1–18. doi:97208-3890

Platnick, S., King, M.D., Ackerman, S. a., Menzel, W.P., Baum, B. a., Riédi, J.C., Frey, R. a., 2003. The MODIS cloud products: Algorithms and examples from terra. IEEE Trans. Geosci. Remote Sens. 41, 459–472. doi:10.1109/TGRS.2002.808301

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. doi:10.1016/j.rse.2017.07.002

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010. Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. Remote Sens. Environ. 114, 35–49. doi:10.1016/j.rse.2009.08.011

Scaramuzza, P.L., Bouchard, M.A., Dwyer, J.L., 2012. Development of the landsat data continuity mission cloud-cover assessment algorithms. IEEE Trans. Geosci. Remote Sens. 50, 1140–1154. doi:10.1109/TGRS.2011.2164087

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. doi:10.1109/TGRS.2012.2203823

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. doi:10.1016/S0034-4257(02)00034-2

Zhu, X., Liu, D., 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. ISPRS J. Photogramm. Remote Sens. 102, 222–231. doi:10.1016/j.isprsjprs.2014.08.014

Zhu, X., Liu, D., 2014. Accurate mapping of forest types using dense seasonal landsat time-series. ISPRS J. Photogramm. Remote Sens. 96, 1–11.

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. doi:10.1016/j.rse.2014.12.014

Zhu, Z., Woodcock, C.E., 2014. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. Remote Sens. Environ. 152, 217–234. doi:10.1016/j.rse.2014.06.012

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94. doi:10.1016/j.rse.2011.10.028