Remote Sensing of Environment 152 (2014) 217–234

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change

Zhe Zhu *, Curtis E. Woodcock

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

ARTICLE INFO

Article history:
Received 22 January 2014
Received in revised form 15 June 2014
Accepted 15 June 2014
Available online xxxx

Keywords:
Tmask
Fmask
Cloud
Cloud shadow
Snow
Landsat
Multitemporal
Change

ABSTRACT

We developed a new algorithm called Tmask (multiTemporal mask) for automated masking of cloud, cloud shadow, and snow for multitemporal Landsat images. This algorithm consists of two steps. The first step is based on a single-date algorithm called Fmask (Function of mask) that initially screens most of the clouds, cloud shadows, and snow. The second step benefits from the extra temporal information from the remaining "clear" pixels and further improves the cloud, cloud shadow, and snow mask. Three Top Of Atmosphere (TOA) reflectance bands (Bands 2, 4, and 5 — Landsat-7 band numbering) are used in a Robust Iteratively Reweighted Least Squares (RIRLS) method to estimate a time series model for each pixel. By comparing model estimates with Landsat observations for the three spectral bands, the Tmask algorithm is capable of detecting any remaining clouds, cloud shadows, and snow for an entire stack of Landsat images. Generally, this algorithm will not falsely identify land cover changes as clouds, cloud shadows, or snow, as it is capable of modeling land cover change. The multitemporal images also provide extra information for better discrimination of cloud and snow, which is difficult for single-date algorithm. A snow threshold is derived for Band 5 TOA reflectance for each pixel at each specific time based on a modified Norwegian Linear Reflectance-to-Snow-Cover (NLR) algorithm. By comparing the results of Tmask with a single-date algorithm (Fmask) for multitemporal Landsat images located at Path 12 Row 31, significant improvements are observed for identification of clouds, cloud shadows, and snow. The most significant improvement is observed for cloud shadow detection. Many of the errors in cloud, cloud shadow, and snow detection observed in Fmask are corrected by the Tmask algorithm. The goal is development of a cloud, cloud shadow, and snow detection algorithm that results in a multitemporal stack of images that is free of "noise" factors and thus suitable for detection of land cover change.

1. Introduction

Landsat data has been widely used in remote sensing because of its medium spatial resolution (Woodcock & Strahler, 1987), accurate radiometric calibration (Chander, Markham, & Helder, 2009), high geometric precision (Lee, Storey, Choate, & Hayes, 2004; Masek, Honzak, Goward, Liu, & Pak, 2001), and long historical record (Markham, Storey, Williams, & Irons, 2004). The policy providing free access to Landsat data has made Landsat data even more popular (Woodcock et al., 2008) and has completely revolutionized the utilization of Landsat data (Wulder, Masek, Cohen, Loveland, & Woodcock, 2012). Take change detection as an example: previously, we detected land cover change by comparing two dates of clear Landsat images (Collins & Woodcock, 1996; Healey, Cohen, Yang, & Krankina, 2005; Masek et al., 2008), but now algorithms use tens (Huang, Goward, et al., 2010; Huang, Thomas, et al., 2010; Kennedy, Cohen, & Schroeder, 2007; Vogelmann, Tolk, & Zhu, 2009; Zhu, Woodcock, & Olofsson, 2012) or even hundreds (Zhu & Woodcock, 2014) of Landsat images at the same location. In this new data rich era, many preprocessing methods that require user input are no longer practical. One of the most immediate problems is cloud, cloud shadow, and snow detection in Landsat images.

Clouds, their shadows, and snow significantly influence optical sensors like Landsat (Dozier, 1989; Irish, Barker, Goward, & Arvidson, 2006; Zhu & Woodcock, 2012). The brightening effect of clouds and snow and the darkening effect of cloud shadows significantly influence the reflectance of different spectral bands. Screening of clouds, cloud shadows, and snow is especially crucial for remote sensing activities like change detection because undetected cloud, cloud shadow, or snow will likely result in identification of change where none occurred ("false positive errors"). Considering the relatively small areas of land cover change, this type of error significantly decreases change detection accuracy. Therefore, identification of clouds, cloud shadows, and snow is usually the first step in most remote sensing activities, and for certain applications like change detection, very accurate detection is required.

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

http://dx.doi.org/10.1016/j.rse.2014.06.012
0034-4257/

The detection of clouds, cloud shadows, and snow is not always easy, especially if we want to detect them accurately. Clouds are notoriously difficult to detect in Landsat images, due to the limited Landsat spectral bands and the complexity of clouds themselves (Zhu & Woodcock, 2012). Many types of clouds exist, and each kind may have a different spectral signature based on cloud properties like cloud optical thickness, particle effective radius, thermodynamic phase, and cloud height (Platnick et al., 2003). Moreover the spectral signature of optically thin clouds can be very similar to the signature of the Earth surfaces underneath, making them more difficult to separate from clear observations. Cloud shadow detection can be difficult as well due to the spectral similarity of cloud shadows to dark surfaces. Thin cloud shadows are even more difficult to detect, as their spectral signature can be almost the same as clear pixels due to the penetration of solar radiation. Snow detection is usually considered relatively easier as the Normalized Difference Snow Index (NDSI) is very helpful for snow detection (Salomonson & Appel, 2004). However, the NDSI values of snow pixels can also change significantly depending on the grain size, the thickness of snowpack, and the amount of impurities (Warren & Wiscombe, 1980; Wiscombe & Warren, 1980). Moreover, most of snow-covered surfaces are actually a mixture of snow and other land cover types. In forested areas, snow is mixed with trees, and the NDSI values of these pixels are much lower than pure snow pixels (Klein, Hall, & Riggs, 1998; Xin et al., 2012). Additionally, snow and clouds can be very difficult to separate in some circumstances. Certain clouds, such as ice clouds, can have very similar spectral signatures to snow. Sometimes, it is almost impossible to separate clouds from snow based only on the spectral information.

To detect clouds, cloud shadows, and snow, one common approach is to identify them manually based on hand-drawn polygons. This works fine for processing a few Landsat images, but if we want to use a large number of Landsat images, more automated algorithms are needed. Recently, many new automated algorithms have been developed based on a single Landsat image (Huang, Goward, et al., 2010; Huang, Thomas, et al., 2010; Irish et al., 2006; Masek et al., 2006; Oreopoulos, Wilson, & Várnai, 2011; Roy et al., 2010; Scaramuzza, Bouchard, & Dwyer, 2012; Zhu & Woodcock, 2012). The development of these automated algorithms has made it possible for various kinds of remote sensing activities that use many Landsat images. However, for certain kinds of applications such as change detection, the single-date masking algorithms are still not accurate enough. Some of the single-date algorithms are capable of providing masks with high accuracy, but, given the relatively small areas of land cover change in most environments, any errors in the masking process will pose major problems for change detection. To remove clouds as much as possible, one solution for single-date algorithms is to use a lower threshold in detecting clouds (Zhu & Woodcock, 2012). However, this will also overestimate clouds and their shadows, and many clear pixels will be classified as cloud or cloud shadow, making change detection algorithms difficult for these pixels because of insufficient data.

To better detect clouds, cloud shadows, and snow, new algorithms based on multitemporal images have been developed for a number of satellite sensors, including Landsat (Goodwin, Collett, Denham, Flood, & Tindall, 2013; Hagolle, Huc, Pascual, & Dedieu, 2010; Jin et al., 2013; Wang, Ono, Muramatsu, & Fujiwara, 1999), Systeme Probatoire d'Observation de la Terre (SPOT) (Tseng, Tseng, & Chien, 2008), Spinning Enhanced Visible and Infrared Imager (SEVIRI) (Derrien & Le Gléau, 2010), and Moderate Resolution Imaging Spectroradiometer (MODIS) (Liu & Liu, 2013; Lyapustin, Wang, & Frey, 2008). The basic idea of these algorithms is that clouds, cloud shadows, and snow will cause sudden changes to the reflectance, and by comparing a reference image without clouds to the observed image, clouds, cloud shadows, and snow will be easily detected. These algorithms are reported to have higher accuracies in detecting clouds and their shadows. Goodwin et al. (2013) found that their multitemporal algorithm will produce better results in detecting cloud shadow compared to the Function of mask (Fmask) algorithm (Zhu & Woodcock, 2012). Despite the reported better results in these multitemporal algorithms, there are also disadvantages. The biggest disadvantage is that they may cause problems for applications like change detection because land cover change will also result in sudden changes to satellite observations. Most of these multitemporal cloud, cloud shadow, and snow detection algorithms rely on the assumption that between the time of the reference image and the observed image there is not any land cover change and differences in reflectance only result from clouds, cloud shadows, and snow. This may be true for some sensors with high temporal frequency such as MODIS or SEVIRI if the reference image is very close in time with the observed image. For sensors like Landsat, this assumption is often invalid, especially for places where land cover change is common. There have been several approaches proposed for limiting the effect of land cover change on multitemporal cloud and cloud shadow identification. For example, some of the algorithms use the Band 7/Band 1 ratio (Zhu et al., 2012) or the Band 3/Band 1 relationship (Hagolle et al., 2010) to distinguish some kinds of frequent changes (e.g. agriculture) from clouds. Lyapustin et al. (2008) propose to use an internally derived surface change mask to prevent the possibility of identifying surface change as clouds. On the other hand, Goodwin et al. (2013) use a geometry-based approach to distinguish land cover change from cloud shadows. However, it is difficult to exclude all kinds of land cover change with these empirically derived spectral tests or include all possible changes in a surface change mask, particularly given the wide variety of kinds of land cover change. This kind of commission error — where land cover change is removed from images as part of the cloud/cloud shadow screening process — is particularly serious when the ultimate goal of the analysis is to monitor land cover change. Moreover, as both clouds and snow usually make the visible bands brighter, it is difficult to separate snow from clouds based on simple image differencing. Most multitemporal algorithms assume that there is no snow in the image and the pixels that are brighter than the reference values are only due to clouds (Goodwin et al., 2013; Hagolle et al., 2010; Jin et al., 2013; Wang et al., 1999).

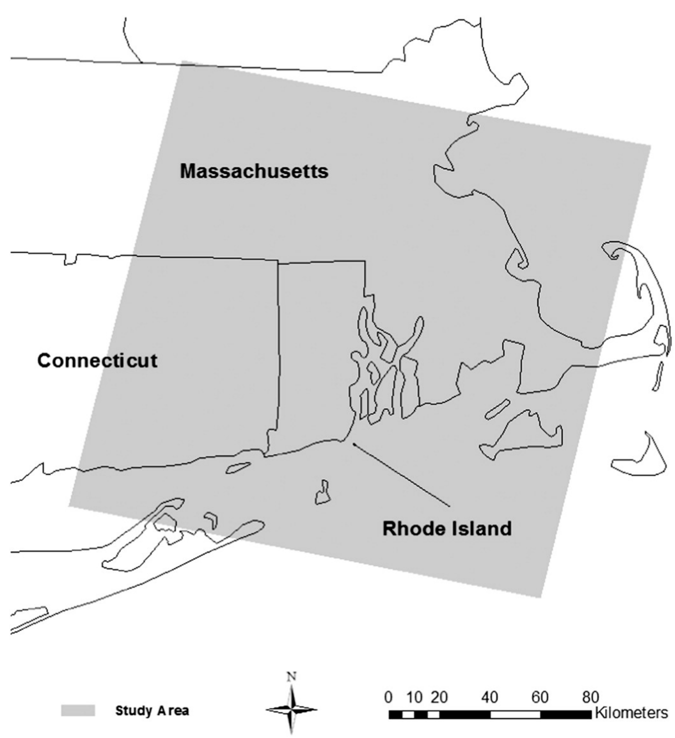

Fig. 1. Study area (Fig. 2 in Zhu & Woodcock, 2014).

The trade-off in multitemporal cloud, cloud shadow, and snow screening between higher accuracy and the possibility of masking out land cover change is significant. In an attempt to solve this problem, we developed a new algorithm designed specifically for change detection that employs many images called Tmask (multiTemporal mask) for automated detection of clouds, cloud shadows, and snow in Landsat images. This algorithm has the following advantages: 1) it achieves more accurate detection of clouds, cloud shadows, and snow; 2) it does not exclude land cover change; and 3) it has better discrimination of clouds and snow.

2. Data and study area

2.1. Study area

We use the Northeastern United States as our study area, which includes all of Rhode Island, parts of Eastern Connecticut, and much of Eastern Massachusetts (Fig. 1). It has been selected for the following reasons: 1) this coastal area is frequently influenced by both clouds and snow; 2) there are many kinds of land cover change occurring in this study area, including forest clearing, urbanization, and abandonment of agricultural fields (Zhu & Woodcock, 2014); and 3) there are large areas of forest in this study area and in winter the detection of snow pixels is difficult for pixels that include both forest and snow.

2.2. Landsat data

A total of 88 Level 1 Terrain corrected (L1T) Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) images from 2005 to 2006 at Worldwide Reference System (WRS) Path 12 and Row 31 are used. We use all available Landsat TM and ETM+ images with cloud cover less than 80% (determined by the Fmask algorithm).

3. Methods

3.1. Image pre-processing

The original Digital Number (DN) values of Landsat Bands 2, 4, and 5 (Landsat-7 band numbering) are converted to Top Of Atmosphere (TOA) reflectance with the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) atmosphere correction tool (Masek et al., 2006). Band 2, 4, and 5 TOA reflectance and Fmask results (see Section 3.2.1 for details) are then stacked to have the same upper left corner and same image size. We choose TOA reflectance instead of surface reflectance as Tmask input because TOA reflectance includes all atmosphere influences and atmosphere correction is meaningless for cloud pixels. Instead of using the blue band (Band 1) for cloud detection as many multitemporal algorithms do (Goodwin et al., 2013;

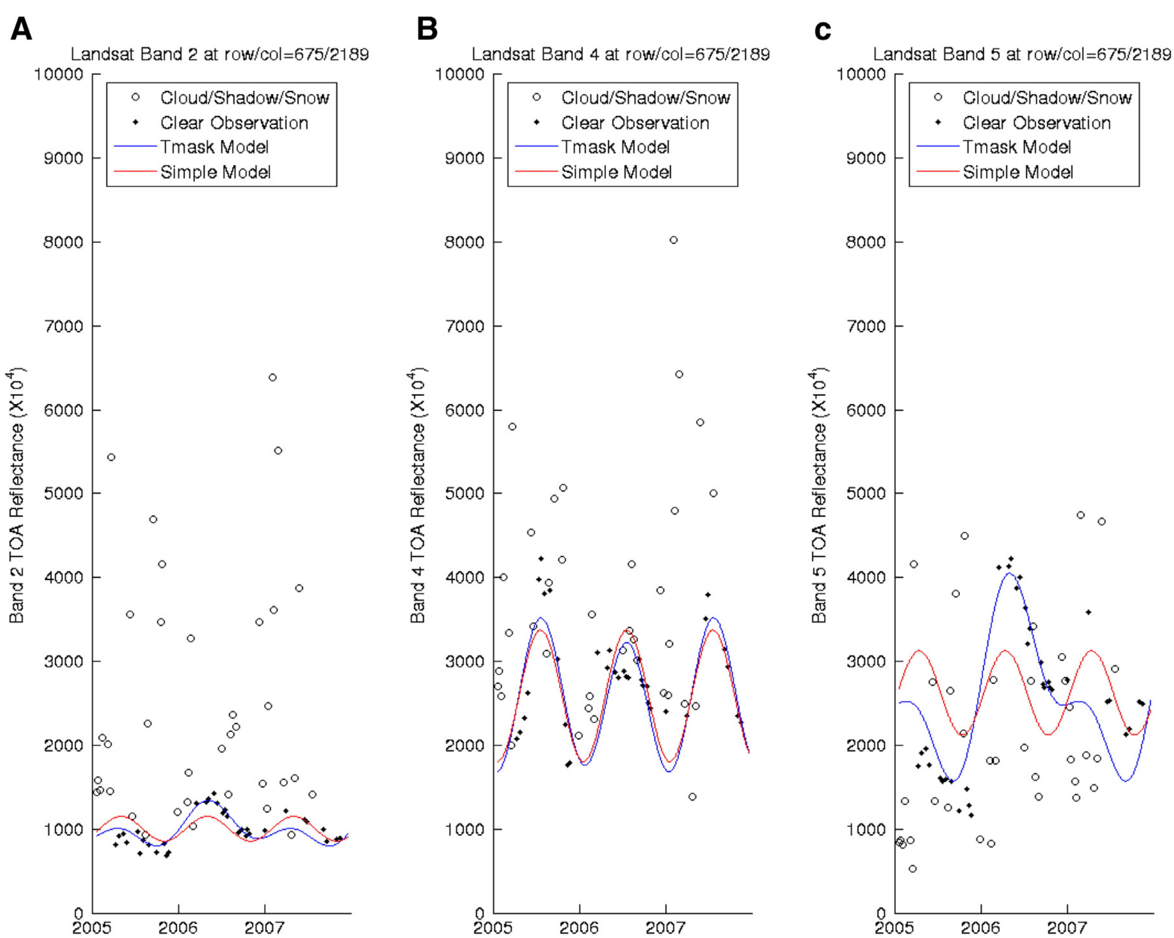

Fig. 2. Estimates for Bands 2 (A), 4 (B), and 5 (C) for TOA reflectance for a deforestation pixel based on two different time series models (Tmask model and a simpler model). Black dots are clear observations, and black circles are either clouds, cloud shadows, or snow identified by Fmask and are not used in the model estimation. The red line represents the time series model that does not include land cover change. The blue line represents the time series model that includes land cover change. Note that for this deforestation pixel the red line fails to capture the land cover change signal, especially for Band 5 TOA reflectance, and the blue line successfully captures most of the signal contributed by both seasonality and land cover change. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Hagolle et al., 2010; Jin et al., 2013; Wang et al., 1999), Tmask uses the green band (Band 2) for detecting clouds and snow. Though the blue band has the advantage of generally low reflectance for most of the Earth surface and relatively high reflectance for clouds and snow, this Landsat Band is also known to saturate frequently (Dozier, 1984, 1989). This will cause serious problems for the multitemporal algorithms that use the blue band for cloud detection because the presence of clouds and snow will not make saturated, bright surfaces any brighter. The other drawback for the blue band is that it is also sensitive to other atmospheric factors such as aerosols and smoke (Kaufman et al., 1997; Vermote, El Saleous, & Justice, 2002). On the other hand, the green band (Band 2) has proven to be less frequently saturated (Dozier, 1984, 1989) and at the same time less sensitive to atmospheric influences. The Near Infrared (NIR) band (Band 4) is mainly chosen for cloud shadow detection and snow and cloud shadow separation because cloud shadows make the NIR band darker, but snow makes the NIR band brighter. The Short Wave Infrared (SWIR) band (Band 5) is used to separate snow and clouds, and because snow and cloud shadows are generally dark in this band, but clouds are usually brighter, Band 5 also helps with cloud shadow detection. Note that we did not use the thermal band for this multitemporal cloud, cloud shadow, and snow detection, though it has been reported very helpful in detecting clouds in many single-date algorithms (Huang, Goward, et al., 2010; Huang, Thomas, et al., 2010; Irish et al., 2006; Masek et al., 2006; Zhu & Woodcock, 2012). The reason for not including the thermal band in this multitemporal algorithm is that the thermal band is quite sensitive to different physical phenomena that are not related to cloud, cloud shadow, or snow and if this band is included, many commission errors will occur.

3.2. First step — initial masking

The basic idea of the Tmask algorithm is to compare modeled or "predicted" TOA reflectance with Landsat observations to detect clouds, cloud shadows, and snow. The "predicted" TOA reflectance comes from a time series model. However, the time series analysis needs a dataset that is 100% free of clouds, cloud shadows, and snow, as a single missed outlier can bias the estimation of the time series model. On the other

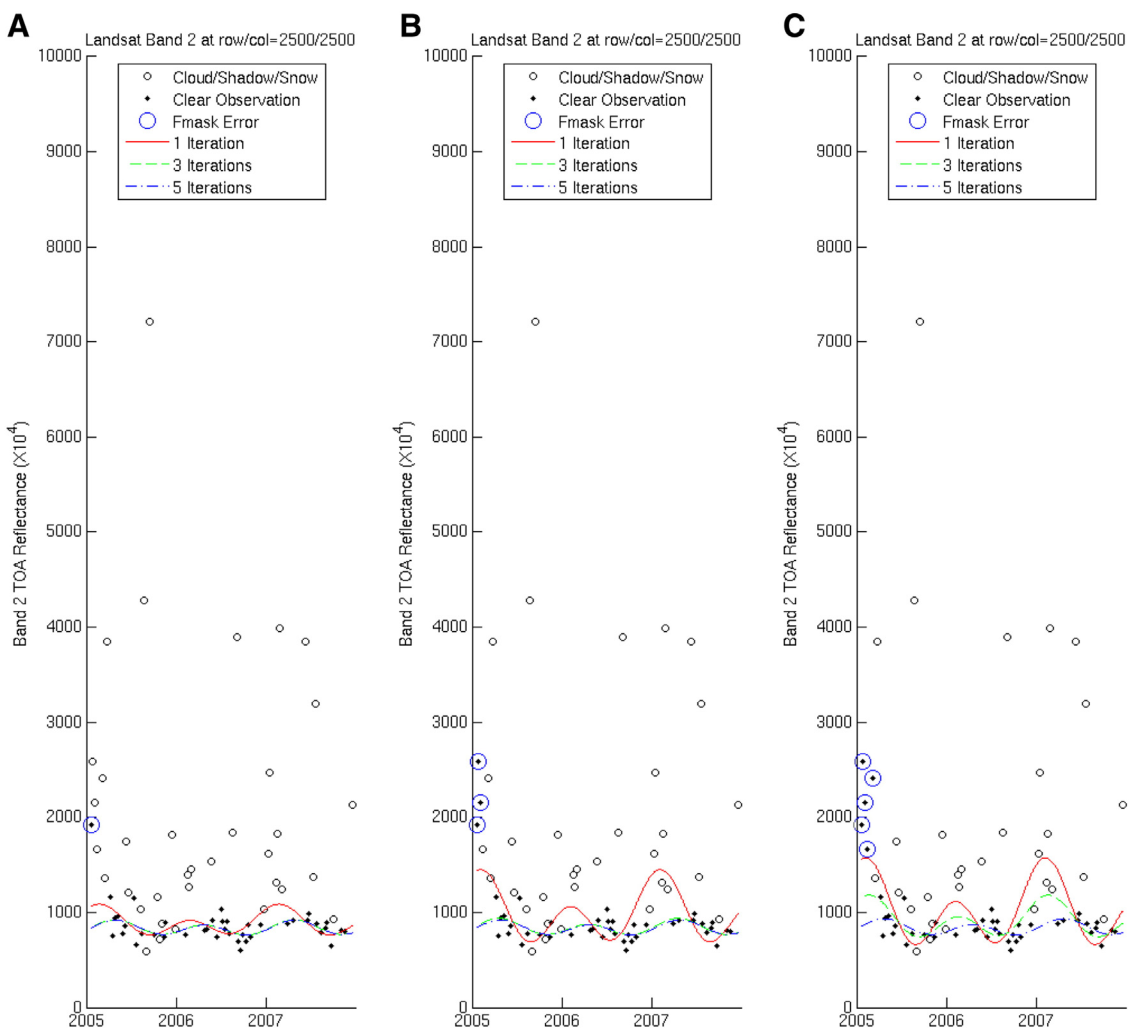

Fig. 3. Tests of the influence of different numbers (one, three, and five) of consecutive Fmask mistakes on the RIRLS method based on different numbers (one, three, and five) of maximum iterations for Band 2 TOA reflectance. Clear pixels are black dots. Clouds, cloud shadows, snow are black circles. Fmask mistakes are blue circles. The red, green, and blue lines correspond to the model estimates with a maximum of one, three, and five iterations. Fig. 3A shows the model estimation results when Fmask only makes one mistake. Fig. 3B shows the model estimation results when Fmask makes three mistakes in a row. Fig. 3C shows the model estimation results when Fmask makes five mistakes in a row. Note that when the RIRLS method uses a maximum of five iterations, the model estimation is not influenced even if Fmask misses clouds five consecutive times. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

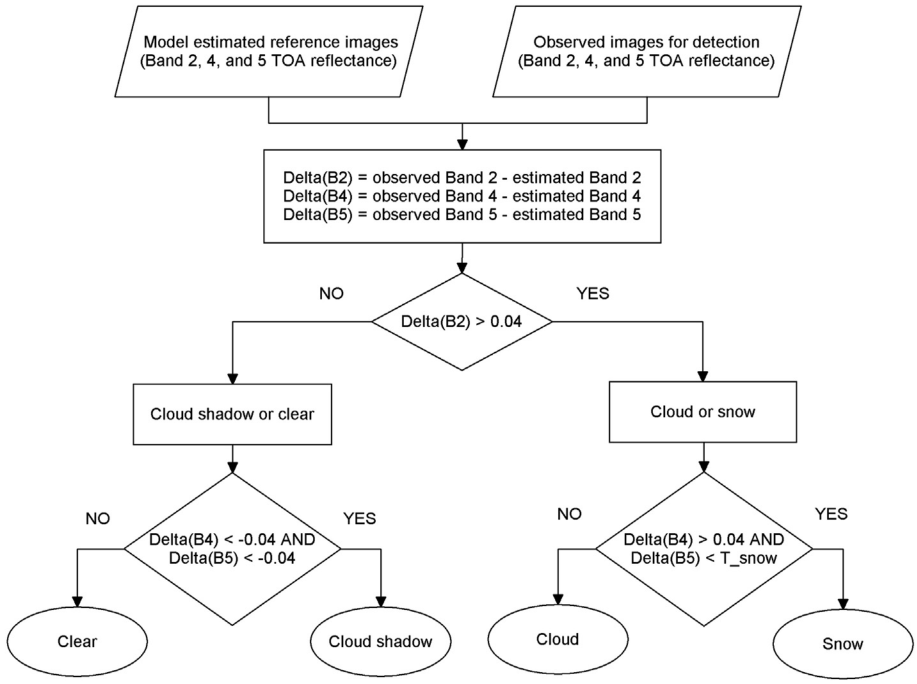

Fig. 4. Flow chart of Tmask spectral differencing algorithm.

hand, this 100% clean dataset is only possible if we can detect all clouds, cloud shadows, and snow without any error, which is not possible using any single-date algorithm. To avoid this "chicken and egg" problem, we used a progressive method that first uses a single-date algorithm (Fmask) to exclude most of the clouds, cloud shadows, and snow.

3.2.1. The Fmask algorithm

In this paper we used the newly developed single-date Fmask algorithm (Zhu & Woodcock, 2012) for this initial masking because of its high accuracy in cloud detection (96.4% in overall accuracy) and its capacity to provide cloud shadow and snow masks. The Fmask algorithm is applied to individual images (i.e. single dates) and provides a cloud, cloud shadow, and snow mask for each image. In this algorithm, clouds and snow are identified based on many spectral tests, and cloud shadow is initially extracted based on a flood-fill transformation and then confirmed based on an object-based cloud and cloud shadow match approach. We applied the Fmask algorithm to all 88 Landsat images. We also dilated all clouds, cloud shadows, and snow by 3 pixels in all 8 connected directions to remove the surrounding pixels that may be partially influenced.

3.2.2. Backup algorithm for Fmask

For most of the single-date algorithms (including Fmask), one of the major problems is that they may exclude pixels consistently detected as clouds if they are cold, bright, and white. This will make estimation of the time series model fail because of insufficient clear observations (less than 15 "clear" observations). For each pixel, the Tmask model needs at least 15 "clear" observations for robust model estimation. If the number of total "clear" observations for a pixel is less than 15, model estimation may fail easily and when this happens, a backup algorithm is applied. This backup algorithm is based on the assumption that clouds cannot stay at the same place persistently, and Band 2 TOA reflectance for most cloud pixels will always be higher than the median values of the entire time series at the same location. Therefore, if Band 2 TOA reflectance of a pixel is less than or equal to the median value plus 0.04, it is identified as a "clear" pixel (Eq. 1). The reason for adding 0.04 for thresholding is based on the fact that clouds will make Band 2 brighter and based on the testing of all the 88 images, a threshold of 0.04 detects most of the clouds (even including some very thin clouds) and will not misidentify other changes (e.g., seasonal changes, soil wetness changes) as clouds. Note that all the pixels used for calculating

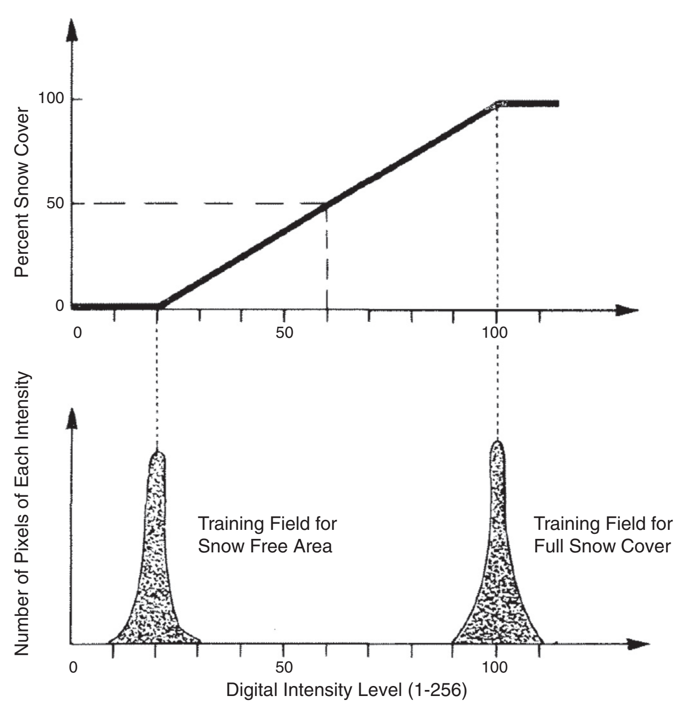

Fig. 5. Illustration of the Norwegian Linear Reflectance-to-snow-cover (NLR) algorithm in which a pixel's digital intensity level is linearly transformed to a percentage of snow cover (Fig. 1 in Andersen, 1982).

the median value of Band 2 TOA reflectance should be non-snow pixels, as snow can persist in certain areas and if we do not exclude snow beforehand, the median value of Band 2 TOA reflectance may be from a snow pixel, which will make multitemporal cloud masking fail (the "predicted" pixel is not a clear pixel). Therefore, all the time series observations used for the backup algorithm are from non-snow pixels identified by the Fmask algorithm. For places of perennial snow, even this backup algorithm will not be able to work due to the lack of non-snow observations. For these pixels, we will use the single-date algorithm results (Fmask results in this paper) directly for labeling (cloud, cloud shadow, snow, or clear). Note that we will not exclude cloud shadow pixels in this backup algorithm because the areas covered by cloud shadows are much smaller compared to those covered by clouds. Because the second step (Tmask algorithm) is capable of handling outliers, this backup algorithm will work as long as it excludes most of the outliers that are caused by clouds and snow.

$$\rho\left(2, x_j\right) \leq median\left(\rho\left(2, x_{\{1,2,3\ldots K\}}\right)\right) + 0.04 \quad (1)$$

where:

x is the Julian date
j is the jth non-snow pixel
K is the total number of non-snow pixels
$\rho(2, x_j)$ is the Observed Landsat Band 2 TOA reflectance at Julian date x for the jth non-snow pixel.

3.3. Second step — Tmask algorithm

The Tmask algorithm is based on the results of initial masking and further improves the initial masking results by using temporal information. It is based on the idea that we have already identified most of the cloud, cloud shadow, and snow in the first step, and by using an empirically estimated time series models for the rest of the "clear" observations it is possible to model or "predict" TOA reflectance with Landsat observations to better detect clouds, cloud shadows, and snow. When observations differ dramatically from the "predicted" values, it may be due to clouds, cloud shadows or snow. The simplified version of the Tmask algorithm has been successfully applied for applications like monitoring forest disturbance (Zhu et al., 2012) and continuous change detection and classification of land cover (Zhu & Woodcock, 2014). However, in these applications, there was no

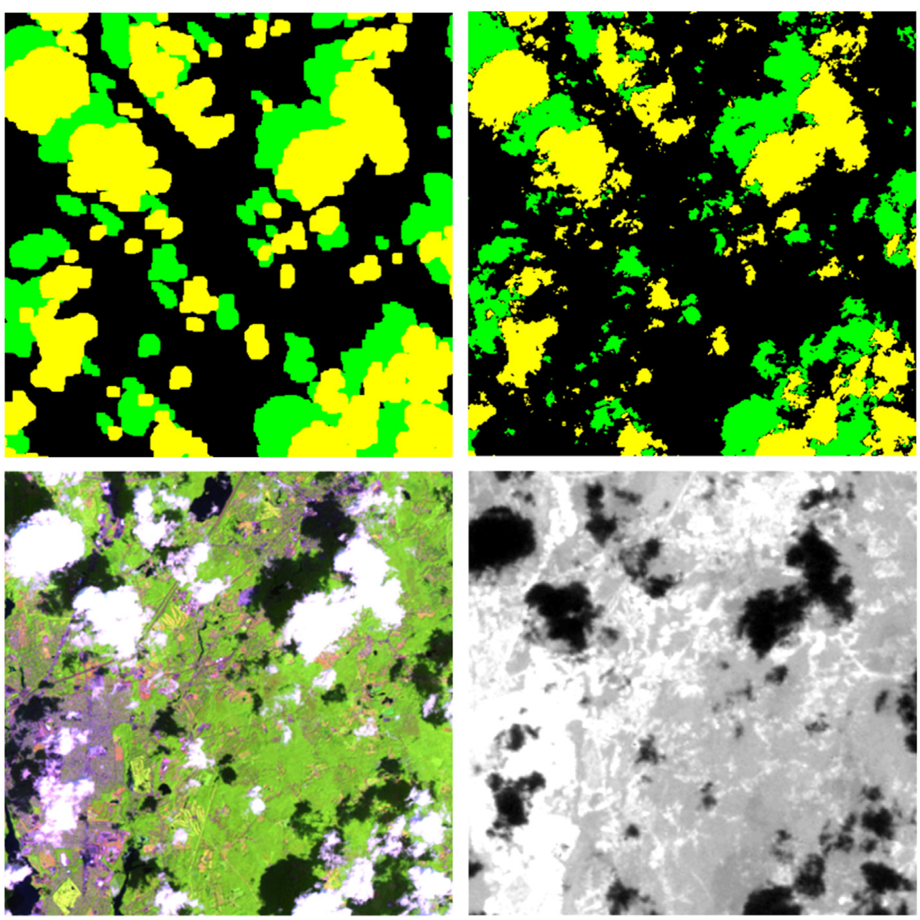

Fig. 6. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired August 11th 2006 — a scenario where two algorithms agree for thick clouds and their shadows. There are four 12×12 km^2 area subset images covering the same area. The upper left image is the Fmask result, and the upper right image is the Tmask result. Clouds are in yellow, and cloud shadows are in green. The lower left image is a composite of Landsat Bands 5, 4, and 2. The lower right image is a Landsat Band 6 (Brightness Temperature) image. The identified clouds, cloud shadows, and snow in Fmask will be a little larger in extent than the Tmask results because of the three pixel dilation of clouds, cloud shadows, and snow. This dilation is necessary for Fmask as most of the time the pixels surrounding clouds, cloud shadows, and snow will still be influenced by the thin edges of clouds and their shadows and pixels that are partially covered with snow. Dilation is unnecessary for Tmask because the time series approach is capable of detecting thin clouds and their shadows and even mixed snow pixels. Therefore, Fmask has a spatial resolution of 3×3 pixels while Tmask is a pixel-level mask. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

separation of clouds, cloud shadows, and snow, and they are all treated as the same kind of outliers.

3.3.1. Tmask model

The "clear" observations of Band 2, 4, and 5 TOA reflectance identified in the initial masking either by Fmask or the backup algorithm are the input for the Tmask algorithm. A simple time series model consisting of sines and cosines is used to estimate Band 2, 4, and 5 TOA reflectance (Eq. 2). The form of this time series model is similar to those we have used in the past for predicting future Landsat observations in the context of monitoring land cover change (Zhu & Woodcock, 2014; Zhu et al., 2012). We made the time series model as simple as possible because it may be influenced by outliers if the model has too many coefficients. Therefore, Tmask uses two coefficients to capture the seasonality and one coefficient to capture the overall reflectance. The last two coefficients ($a_{2,i}$, $b_{2,i}$) are used to allow the time series model to respond to different kinds of land cover change. By including the land cover change in the time series model, Tmask is less likely to exclude land cover change when comparing model estimates and satellite observations.

$$\hat{\rho}(i,x) = a_{0,i} + a_{1,i}\cos\left(\frac{2\pi}{T}x\right) + b_{1,i}\sin\left(\frac{2\pi}{T}x\right) + a_{2,i}\cos\left(\frac{2\pi}{NT}x\right) + b_{2,i}\sin\left(\frac{2\pi}{NT}x\right) \quad (2)$$

where,

x the Julian date
i the Landsat Band i TOA reflectance ($i = 2$, 4, and 5)
T the number of days per year ($T = 365$)
N the number of years (rounds to the nearest integer greater than or equal to N)
$a_{0,i}$ the coefficient for the overall value for Landsat Band i TOA reflectance
$a_{1,i}$, $b_{1,i}$ the coefficients for intra-annual change for Landsat Band i TOA reflectance
$a_{2,i}$, $b_{2,i}$ the coefficients for inter-annual change for Landsat Band i TOA reflectance.

To illustrate the importance of the two coefficients used for modeling land cover change, Fig. 2 shows the model estimation results for Landsat Band 2, 4, and 5 TOA reflectance with and without the last two coefficients for a deforestation pixel. In this Figure, a total of 88 Landsat images between 2005 and 2007 are used. The black dots are clear observations, and the black circles are either clouds, cloud shadows, or snow previously identified by Fmask. The red line is the simple model that only uses the first three coefficients in Eq. (2). The blue line is the Tmask model estimate that includes the last two coefficients in Eq. (2) for modeling land cover change. The simple model (red line) fails to capture the deforestation signal, and the Tmask model (blue line) successfully captures most of the seasonality differences and land cover changes in the time series data. This situation is especially obvious for Band 5 TOA reflectance, as it is more sensitive to forest disturbance.

3.3.2. Tmask model estimation

As the Fmask algorithm is not perfect, it is likely that cloud, cloud shadow, or snow observations exist within the rest of the "clear" pixels.

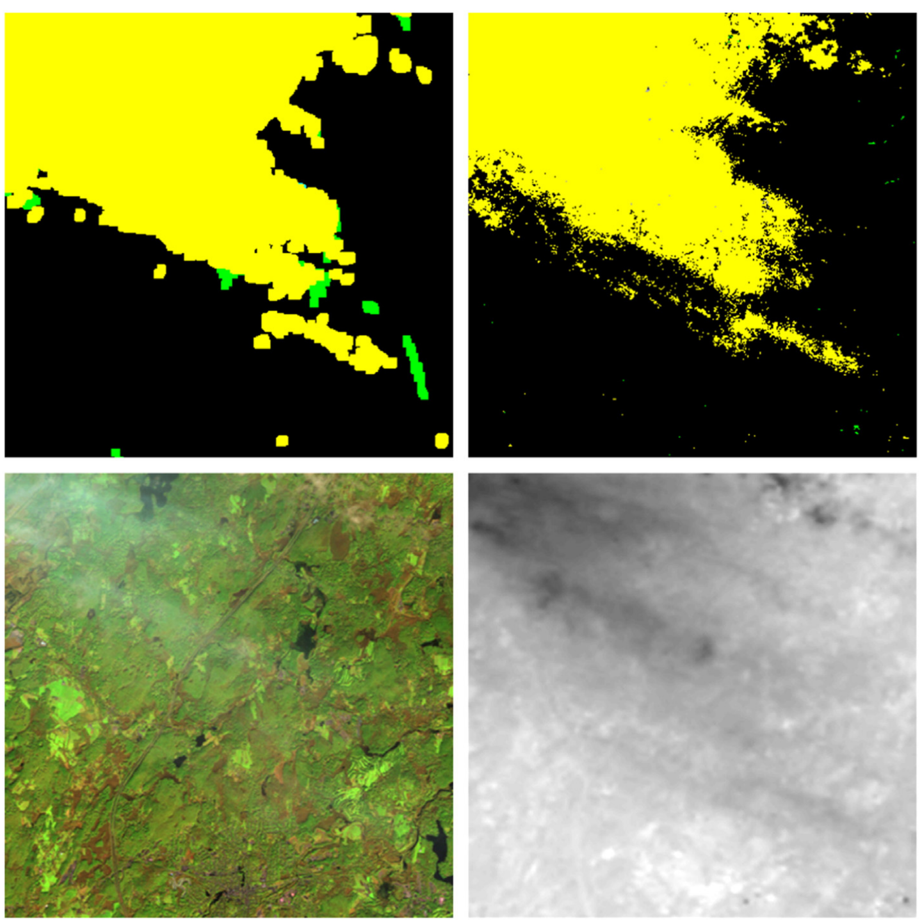

Fig. 7. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired October 22nd 2006 — a scenario where the two algorithms agree for extremely thin clouds (the images follow the description in the caption for Fig. 6).

The regression results will be greatly influenced by these outliers if using the Ordinary Least Square (OLS) method. To overcome this problem, a Robust Iteratively Reweighted Least Squares (RIRLS) method is used for estimating the time series model (DuMouchel & O'Brien, 1989; Holland & Welsch, 1977; O'Leary, 1990; Street, Carroll, & Ruppert, 1988). This algorithm is capable of reducing the weight of the outliers via an iterative process that is capable of providing accurate estimates even when outliers exist.

The weighting function in RIRLS is based on the *Tukey biweight*, also known as the *bisquare weight* (Heiberger & Becker, 1992). The weights are a function of the residuals from the previous iteration (Eq. 3). If the absolute value of the normalized residual ($abs(r)$) is less than 1, the weights (w) are negatively related to the magnitude of normalized residual (r). Otherwise, if the magnitude is larger than 1 (mostly because of clouds, cloud shadows, and snow), the weights are forced to zero. In this way, most of the outliers will have little or no impact on the final results.

$$w = \begin{cases} \left(1-r^2\right)^2 & \text{for } abs(r) \le 1 \\ 0 & \text{for } abs(r) > 1 \end{cases} \tag{3}$$

where:

r — $resid\,/\,(0.4685 \times s \times sqrt(1-h))$
s — $MAD\,/\,0.6745$
resid — is the vector of residuals from the previous iteration
h — is the vector of leverage values from a least-square fit
MAD — Median Absolute Deviation (*MAD*) of the residual from their median.

One of the most important parameters associated with the use of RIRLS is the maximum number of iterations. Usually, the more iteration used, the more robust is the method to outliers. However, this also increases computation time. In our initially screened dataset, it is assumed that Fmask will make mistakes occasionally, but will not make mistakes on consecutive dates. To test how the number of maximum iterations responds to different numbers of consecutive Fmask mistakes, we let the Fmask algorithm make mistakes for one, three, and five consecutive observations. Fig. 3 shows time series of Band 2 TOA reflectance screened by Fmask. Clear pixels are black dots. Clouds, cloud shadows, and snow are black circles. Fmask mistakes are blue circles. The model estimates with a maximum of one, three, and five iterations correspond to the red, green, and blue lines. If Fmask only makes one mistake (Fig. 3A), the model estimates with a maximum of three and five iterations are almost the same (not influenced by the outlier), but the model estimates with a maximum of one iteration are significantly influenced by this single outlier. If Fmask makes three mistakes in a row (Fig. 3B), the model estimates with a maximum of three and five iterations are still very similar, but the model estimates with maximum of one iteration is seriously influenced by the three consecutive outliers. Finally, if Fmask makes five mistakes in a row (Fig. 3C), the model estimates with a maximum of five iterations is still not influenced by the outliers, but the model estimates with a

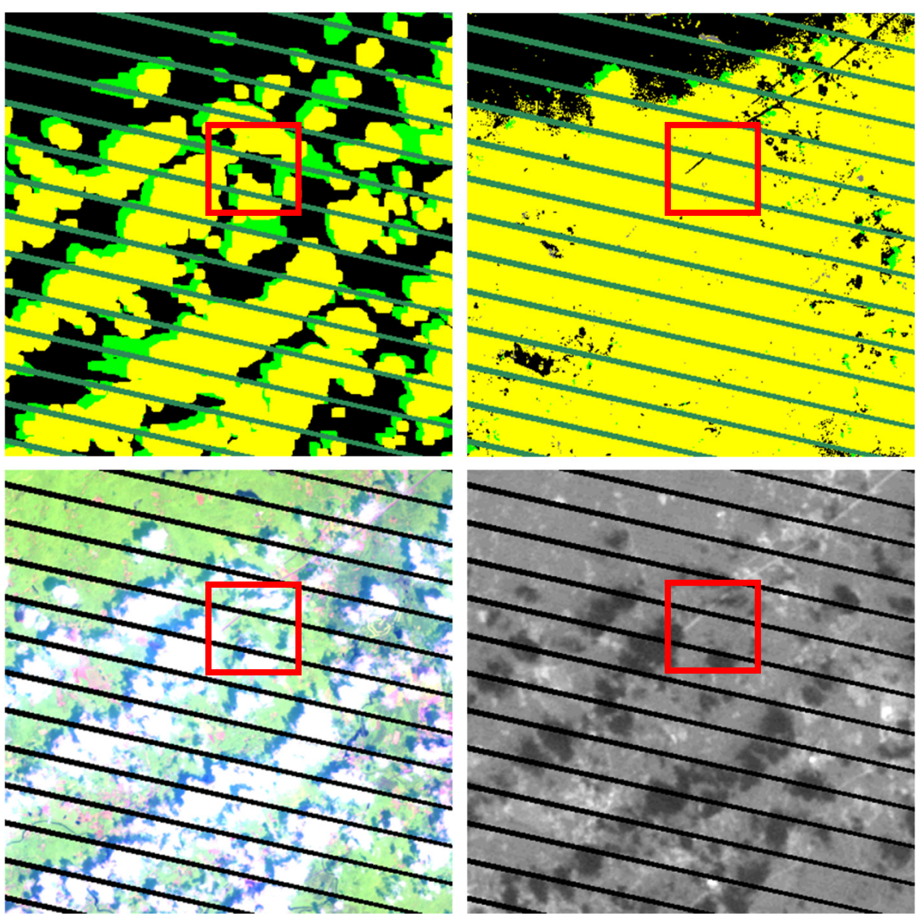

Fig. 8. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired May 7th 2006 — a scenario where two algorithms disagree for extremely thin clouds (Tmask works and Fmask fails). Note that the extremely thin clouds in the center of the red rectangle are missed by Fmask but captured by Tmask. (The images follow the description in the caption for Fig. 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

maximum of one and three iterations are all influenced. Assuming that Fmask will not miss clouds more than five consecutive times for the same pixel, Tmask algorithm uses a maximum of five iterations to find clouds, cloud shadows, and snow that are occasionally missed in the first step.

3.3.3. Tmask spectral differencing

The Tmask algorithm detects clouds, cloud shadows, and snow by differencing the “predicted” images (based on model estimates) with the observed images. A simple decision tree is generated based on the difference between observed and estimated values and the spectral characteristics of clouds, cloud shadows, and snow (Fig. 4). A threshold of 0.04 is chosen for Tmask spectral differencing. This threshold is derived based on a test of all 88 images for the study area by visually examining cloud, cloud shadow, and snow masks when using different thresholds. For most of clear pixels, the difference between “predicted” and the observed Band 2, 4, and 5 TOA reflectance is always less than 0.04, while if there are clouds, cloud shadows, and snow present, the difference is usually larger than 0.04. If the threshold is larger than 0.04, some thin clouds, cloud shadows, and snow pixels may not be able to be detected and if it is smaller than 0.04, noise in the data and other subtle changes such as soil wetness may show up as clouds, cloud shadows, or snow. Therefore, this threshold of 0.04 balances the omission and commission errors for detecting cloud, cloud shadow, and snow.

If the observed Band 2 TOA reflectance minus model estimated Band 2 TOA reflectance is larger than 0.04, this pixel is identified as cloud or snow, as both make visible bands brighter. Cloud and snow separation is based on changes in Band 4 and Band 5 TOA reflectance. The biggest difference is magnitude and direction of change in Band 5 TOA reflectance, as clouds tend to make Band 5 TOA reflectance brighter while snow has the opposite effect. However, there are some ice clouds that can have relatively low Band 5 TOA reflectance, and snow mixed with trees reduces the darkening effect of snow in this band (Xin et al., 2012). Therefore, Tmask uses a pixel-based threshold (*T_snow*) derived from a modified Norwegian Linear Reflectance-to-Snow-Cover (NLR) algorithm (Andersen, 1982). All snow pixels should have changes in Band 5 TOA reflectance less than this threshold and clouds will be larger. TOA reflectance of Band 4 is also used to help separate snow from clouds because snow has a much higher reflectance than clear pixels in the winter, while for clouds the Band 4 TOA reflectance is not always higher than the clear observations, especially during the growing season when vegetation reflects strongly in the Near-Infrared.

On the other hand, for Band 2, if the difference between the observed reflectance and the “predicted” is less than 0.04, this pixel is either clear or cloud shadow, as neither case will make the visible bands brighter. The only difference between clear pixels and cloud shadow is that cloud shadow will make Band 4 and Band 5 TOA reflectance lower, while clear pixels will be similar to the estimated value. Therefore, if the observed Band 4 and Band 5 TOA reflectance are 0.04

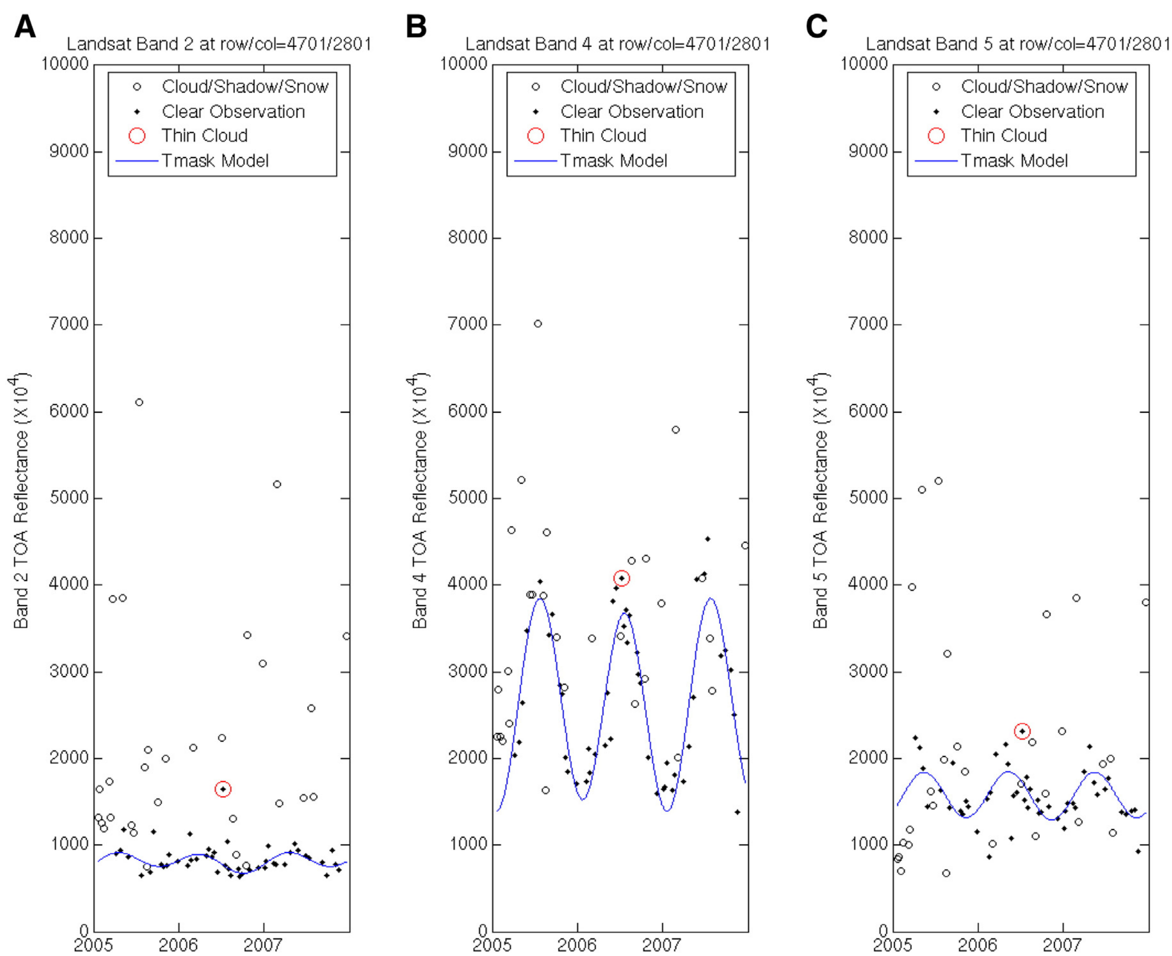

Fig. 9. Times series of Band 2 (A), 4 (B), and 5 (C) TOA reflectance of the pixel located at the center of the red rectangle in Fig. 8. Clear pixels are black dots. Clouds, cloud shadows, and snow are black circles, and the Fmask mistake is shown in the red circle. The blue lines represent the model-estimated values. Note that the cloud observation missed by Fmask but captured by Tmask changes the spectral signal significantly, especially in Band 2 TOA reflectance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lower than the “predicted” values, the pixel is identified as cloud shadow. Fig. 4 is the detailed flow chart of Tmask spectral differencing algorithm.

3.3.4. Tmask snow threshold

As clouds tend to make Band 5 TOA reflectance higher and snow tends to make it lower, a spatially and temporally adjusted Band 5 threshold (*T_snow*) is determined by a modified version of the NLR algorithm. The NLR algorithm is an empirical reflectance-to-snow-cover model originally proposed by Andersen (1982) to estimate snow in Advanced Very High Resolution Radiometer (AVHRR) observations. The NLR algorithm assumes a linear relationship between the observed reflectance (or DN or radiance) and the fractional snow cover for a pixel (Fig. 5). The basic principle of the NLR algorithm is similar to linear spectral unmixing based on a single spectral band. The two spectral components are snow and non-snow background, and the reflectance of the spectral components are calibrated based on 100% fractional snow cover and 0% fractional snow cover pixels. This algorithm works for the visible and near infrared bands. In Tmask, we use green band reflectance for retrieving fractional snow cover for each pixel. The NLR algorithm has good accuracies for homogeneous areas, but it is not accurate for complex landscapes due to its simple assumption of constant reflectance of snow-free pixels (Vikhamar & Solberg, 2003).

To make this algorithm work for heterogeneous environments, we modified the NLR algorithm as follows: the reflectance of snow-free pixels (0% fractional snow coverage) will be no longer constant, but is “predicted” based on the time series model in Eq. (2). Therefore, we can estimate the snow cover percentage for each pixel using the difference between the observed and “predicted” Band 2 reflectance, as shown in Eq. (4). In this equation, $\rho(2, PureSnow)$ is a constant for 100% fractional snow cover pixel in TOA Band 2 reflectance. Due to atmospheric influences and the saturation of Landsat visible bands for bright objects, the Band 2 TOA reflectance of the pure snow pixels (100% fractional snow cover) is usually around 0.4. Therefore, the value of $\rho(2, PureSnow)$ is set to 0.4 in the modified NLR algorithm.

$$Fractional\,snow\,cover = \frac{\rho(2,x)-\hat{\rho}(2,x)}{\rho(2,PureSnow)-\hat{\rho}(2,x)} \tag{4}$$

where:

$\rho(2,x)$ is the observed Landsat Band 2 TOA reflectance at Julian date x
$\hat{\rho}(2,x)$ is the estimated Landsat Band 2 TOA reflectance at Julian date x
$\rho(2, PureSnow)$ is the Landsat Band 2 TOA reflectance of pure snow (a constant value of 0.4).

Similarly, we can apply the modified NLR algorithm to Band 5 TOA reflectance based on the same assumption of a linear relationship between the reflectance and the fraction of snow cover, except that for TOA Band 5 reflectance this relationship is negative. Band 5 TOA reflectance is generally bright for most land covers except for snow, which is usually very low. The reason that this band is not used for the NLR algorithm for estimating fractional snow cover is that it is also

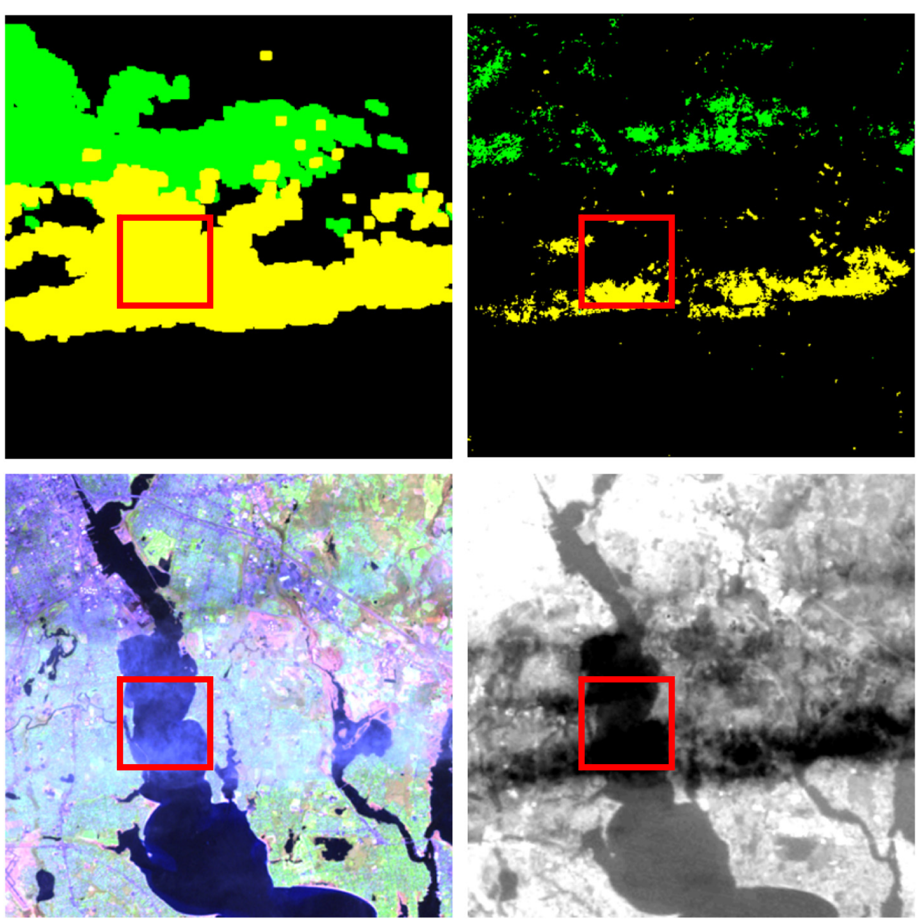

Fig. 10. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired May 7th 2006 — scenario where two algorithms disagree for extremely thin clouds (Fmask works and Tmask fails). Note that the extremely thin clouds in the center of the red rectangle are captured by Fmask but missed by Tmask (the images follow the description in the caption for Fig. 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sensitive to snow grain size. For pure snow pixels, Band 5 reflectance can vary from 0 to 0.12, when the grain radii vary from 1.00 mm to 0.05 mm (Dozier & Painter, 2004). However, in Tmask algorithm, we are not interested in estimating the fractional snow cover. Instead, we want to calculate the optimal threshold for separating snow and cloud. Therefore, we can still derive an equation similar to Eq. (4) for Band 5 TOA reflectance using the modified NLR in Eq. (5).

$$\text{Fractional snow cover} = \frac{\rho(5,x) - \hat{\rho}(5,x)}{\rho(5,\text{PureSnow}) - \hat{\rho}(5,x)} \tag{5}$$

where:

$\rho(5,x)$ is the observed Landsat Band 5 TOA reflectance at Julian date x
$\hat{\rho}(5,x)$ is the estimated Landsat Band 5 TOA reflectance at Julian date x
$\rho(5, PureSnow)$ is the Landsat Band 5 TOA reflectance of pure snow (varies from 0 to 0.12).

By combing Eqs. (4) and (5), we can derive Eq. (6), which provides the estimated change between the observed and estimated Band 5 TOA reflectance for snow pixels.

$$\text{Estimated delta}(B5) = \frac{(\rho(5,\text{PureSnow}) - \hat{\rho}(5,x))x(\rho(2,x) - \hat{\rho}(2,x))}{\rho(2,\text{PureSnow}) - \hat{\rho}(2,x)} \tag{6}$$

As clouds tend to have larger *delta*($B5$) values and snow has smaller *delta*($B5$) values, we need an upper bound of the estimated *delta*($B5$) value to serve as the threshold in Tmask (*T_snow*). Based on Eq. (6), the *estimated delta*($B5$) value is directly related to $\rho(5, PureSnow)$, and the higher the value of $\rho(5, PureSnow)$, the higher the value of *estimated delta*($B5$). Therefore, we use the maximum value of (5, *PureSnow*), which is 0.12 to calculate the snow threshold. In this case, *T_snow* in Fig. 4 is calculated in Eq. (7): if the observed Band 5 TOA reflectance minus the estimated Band 5 TOA reflectance is less than *T_snow*, this pixel is identified as snow, otherwise it is identified as cloud.

$$T\_snow = \frac{(0.12 - \hat{\rho}(5,x)) \times (\rho(2,x) - \hat{\rho}(2,x))}{0.4 - \hat{\rho}(2,x)}. \tag{7}$$

4. Results

We did not compute the accuracy for the Tmask results against a manual mask as we did for the Fmask algorithm (Zhu & Woodcock, 2012) because manually identifying clouds, cloud shadows, and snow in the Landsat images with high accuracy is very difficult. The manual masks that Fmask used to assess its accuracy were derived by the USGS for validating its Automated Cloud-Cover Assessment (ACCA) algorithm (Irish et al., 2006). Even for the carefully derived USGS cloud masks, the average difference of the overall accuracy of the 11 scenes examined by three analysts was around 7% (Oreopoulos et al., 2011). Considering the high overall accuracy already achieved by Fmask algorithm (96.4%), it would be difficult to make a manual mask accurate enough to compare with an algorithm that is probably more accurate than Fmask. Moreover, most of the USGS dataset only has one reference image for each Landsat scene. To validate the Tmask

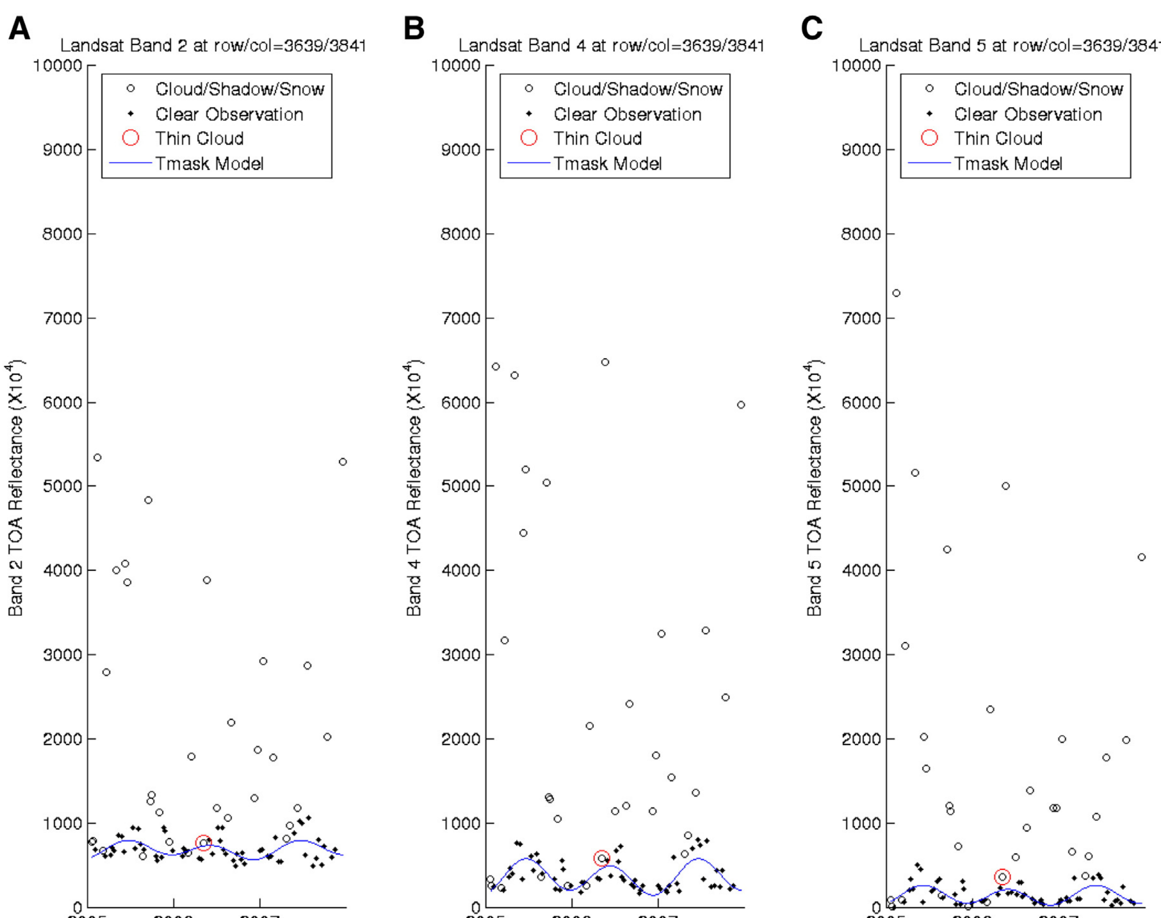

Fig. 11. Times series of Band 2 (A), 4 (B), and 5 (C) TOA reflectance of the pixel located in the center of red rectangle in Fig. 10. Clear pixels are black dots. Clouds, cloud shadows, and snow are black circles, and the Fmask mistake is a red circle. The blue lines represent the model-estimated values. Note that the cloud observation missed by Tmask but captured by Fmask does not change the spectral signal significantly for all three bands. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

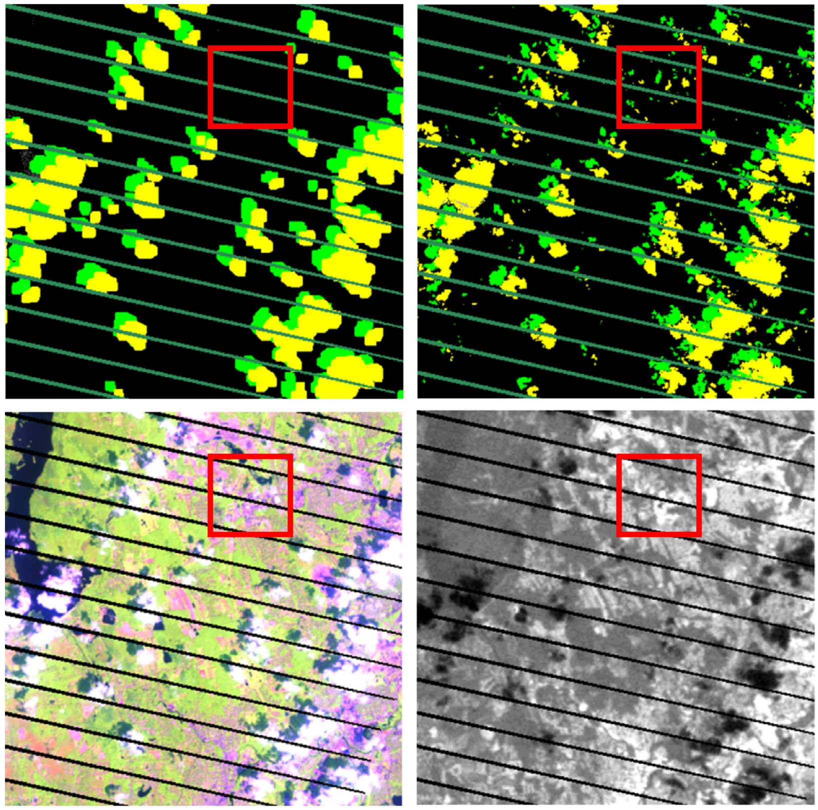

Fig. 12. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired July 26th 2006 — a scenario where the two algorithms disagree for very small clouds and their shadows (Tmask works and Fmask fails). Note that the small clouds and their shadows located within the red rectangle are missed by Fmask but captured by Tmask (the images follow the description in the caption for Fig. 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

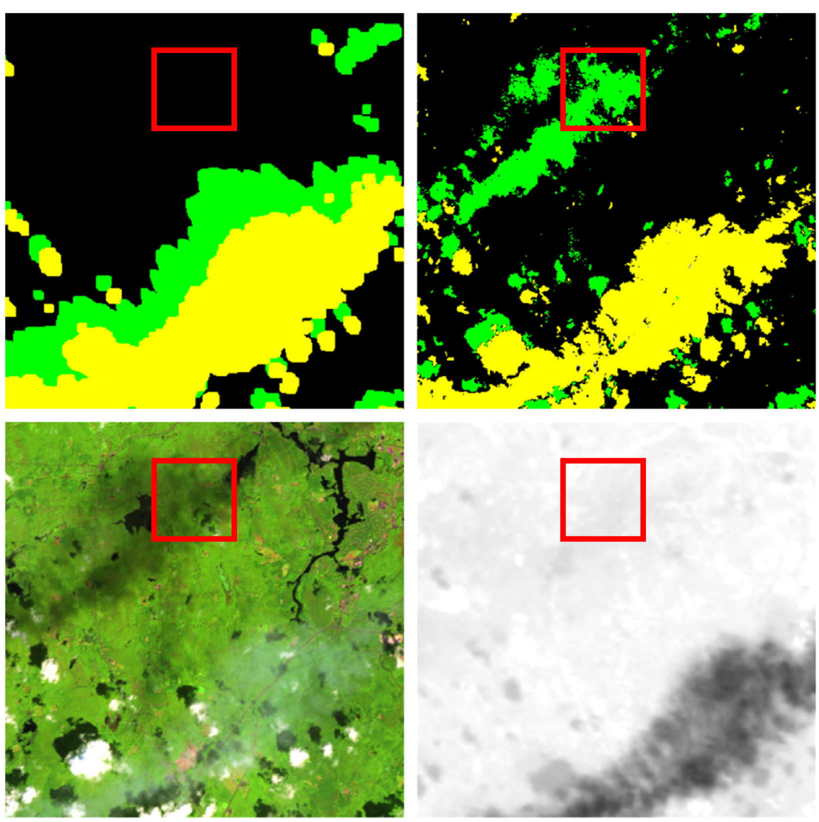

Fig. 13. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired August 19th 2006 — a scenario where two algorithms disagree for shadows from clouds that are both thin and high (Tmask works and Fmask fails). Note that the cloud shadows located within the red rectangle are missed by Fmask but captured by Tmask (the images follow the description in the caption for Fig. 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

algorithm, we need a time series of accurate manual masks, which would be almost impossible to make. Therefore, in this study, the Tmask results are only compared to the Fmask results at different times for the study area. As most of the time Fmask and Tmask results are very similar, in this comparison we focus on results where the two algorithms differ. For each comparison shown here, there are four 12×12 km^2 subsets of Landsat images covering the same area. The upper left image is the Fmask result, and the upper right image is the Tmask result. Clouds are in yellow, and cloud shadows are in green. If there is snow, it is in cyan. The lower left image is a composite of Landsat Bands 5, 4, and 2. The lower right image is Landsat Band 6 (Brightness Temperature). We include the thermal band to help visualize some of the extremely thin clouds that are hard to see in the optical bands but are sometimes apparent in the thermal band. Generally, both the Fmask and Tmask algorithms are accurate in detecting thick clouds and their shadows (Fig. 6). For most of the thin clouds, the two algorithms also show similar results, including the extremely thin clouds that are hard to see in the color composites but obvious in the thermal band (Fig. 7).

For some of the extremely thin clouds, Fmask and Tmask show different results. Sometimes, Tmask is capable of identifying extremely thin clouds that Fmask cannot (Fig. 8) and sometimes the opposite is true (Fig. 10). Though Tmask does not always find all the thin clouds, the thin clouds identified by Tmask have changed the spectral signal significantly (Fig. 9). The clouds missed by Tmask but captured by Fmask are clouds that are generally too thin to make TOA reflectance deviate significantly from the model-estimated values (Fig. 11). Therefore this kind of omission error in Tmask is less likely to cause problems for remote sensing activities like change detection. The pixel located at the center of the red rectangle in Fig. 8 is a cloud pixel that is hardly visible in the Band 5, 4, and 2 composite and the brightness temperature image, and we can only find the pixel located at the center of the red rectangle is slightly brighter compared to the clear pixels located at the upper-left corner of the color composite image. Fig. 9 shows the times series for Band 2, 4, and 5 TOA reflectance for this pixel. In this Figure, clear pixels are black dots, and clouds, cloud shadows, and snow are black circles (Fmask results). The cloud observation missed by Fmask is a red circle. The blue lines represent the model estimated values. Significant differences exist between the model estimates and observations in all three bands, especially in Band 2, for the cloud observation missed by Fmask (red circle). If we miss screening this cloud observation, it will be confused with land cover change based on spectral differencing. On the other hand, for the thin clouds captured by Fmask but missed by Tmask (Fig. 10), there is only a subtle difference between observed and estimated values, and this difference is still within the range of most change detection algorithms (Fig. 11).

As Tmask is a pixel-based algorithm, it is able to identify very small clouds and their shadows, which are omitted by object-based algorithms like Fmask if the size of the cloud object is less than 9 pixels (Zhu & Woodcock, 2012). In Fig. 12, Fmask fails to identify the small clouds and their shadows located within the red rectangle but Tmask successfully identifies them.

The biggest benefit of Tmask compared to Fmask is better identification of cloud shadows. Cloud shadow detection in Fmask is based on a

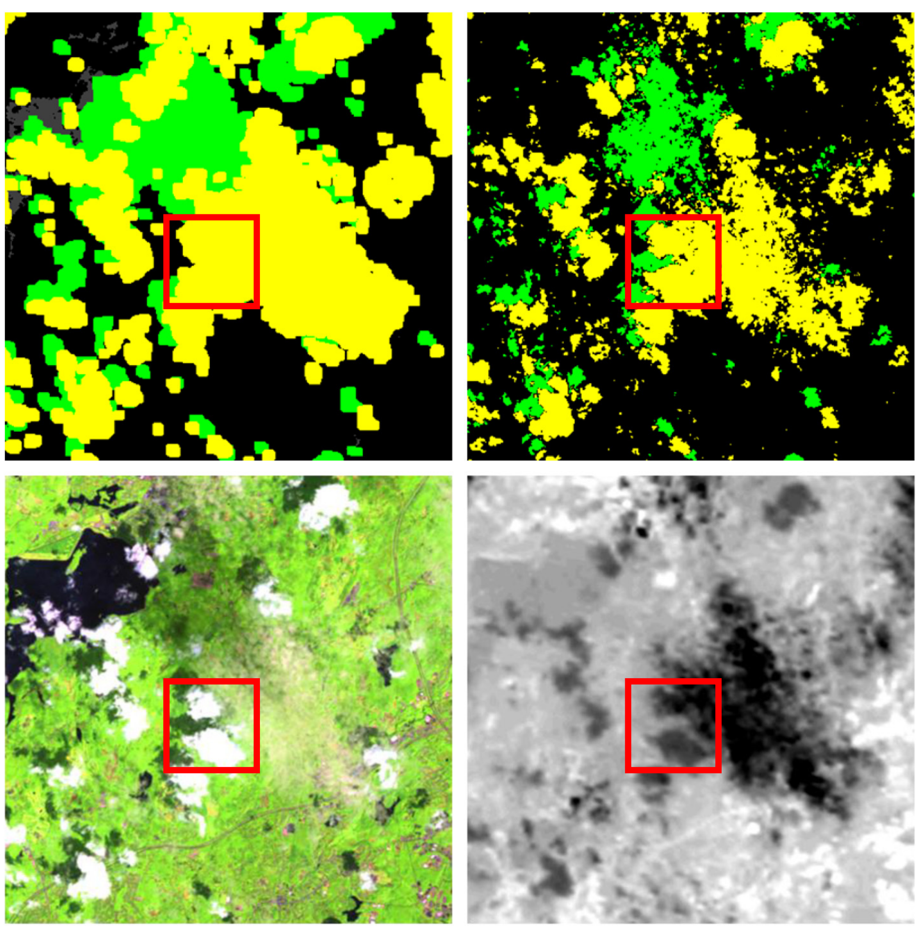

Fig. 14. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired August 3rd 2006 — a scenario where the two algorithms disagree for cloud shadows because of overlapping clouds with different heights (Tmask works and Fmask fails). Note that the cloud shadows located within the red rectangle are missed by Fmask but captured by Tmask (the images follow the description in the caption for Fig. 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

similarity match between clouds and their shadows, which may not be accurate in the following situations: 1) when a cloud is both thin and high; 2) when overlapping cloud objects are at different altitudes; 3) when another dark object is located between the clouds and their shadows; and 4) when clouds are sliced into pieces because of the Scan Line Corrector (SLC-off) problem. In Fig. 13, the cloud is both thin and high; Fmask does not match the shadow for it correctly, but Tmask identifies the cloud shadow successfully (see the red rectangle in Fig. 13).

Problems can also occur when clouds at different heights overlap each other. In Fig. 14, two small thick clouds are overlapped with a large thin cloud (see the red rectangle), and Fmask assumes that they are the same "cloud object" and only matches one cloud shadow for them, but Tmask identifies the cloud shadow successfully.

If there are other dark objects between clouds and their shadows, they can also confuse the cloud and cloud shadow match in Fmask and cause errors in cloud shadow identification, but Tmask identifies the cloud shadow successfully (see the red rectangle in Fig. 15).

Finally, the SLC-off problem can slice one cloud into many pieces and the matched cloud shadows may also show up in small pieces, but Tmask identifies the cloud shadow successfully (see red rectangle in Fig. 16).

For snow and cloud separation, the Tmask algorithm also shows better results than Fmask. For some of the snow covered forest areas, Fmask frequently identifies snow as clouds, but Tmask is capable of separating them (Fig. 17).

Though the Tmask algorithm has significant advantages over the single-date algorithm, it also has disadvantages. For example, the Tmask algorithm assumes all ephemeral changes are caused by clouds, cloud shadows, and snow. However, other ephemeral changes such as flooding or soil wetness may change TOA reflectance for short periods of time. In such cases, the time series model will not be able to respond to this kind of ephemeral change, and comparison of model estimates with Landsat observations may result in the changes being labeled as clouds, cloud shadows, or snow (Fig. 18). For example, the pixel located in the center of the red rectangle in Fig. 18 is a clear pixel that was especially wet on June 16th 2006. It is correctly identified as a clear pixel in Fmask but misidentified as a cloud shadow in Tmask because the observed values in the time series of this pixel (Fig. 19) are much lower than the model estimated value on June 16th 2006 (the red circle). This kind of commission error will remove some good observations; however, for detecting land cover change, this kind of commission error will not cause problems, and instead it will remove ephemeral changes that are easily confused with land cover change.

5. Discussion and conclusions

Though the single-date algorithm (Fmask) achieves high accuracy in detecting clouds (overall accuracy of 96.4% (Zhu & Woodcock, 2012)), for some specific applications such as change detection even a small amount of error can be a problem because the amount of land cover change present at any time may be as small as a fraction of a percent of the image. Thus removal of observations that are actually land cover change when they look like clouds can result in significant underestimation of land cover change.

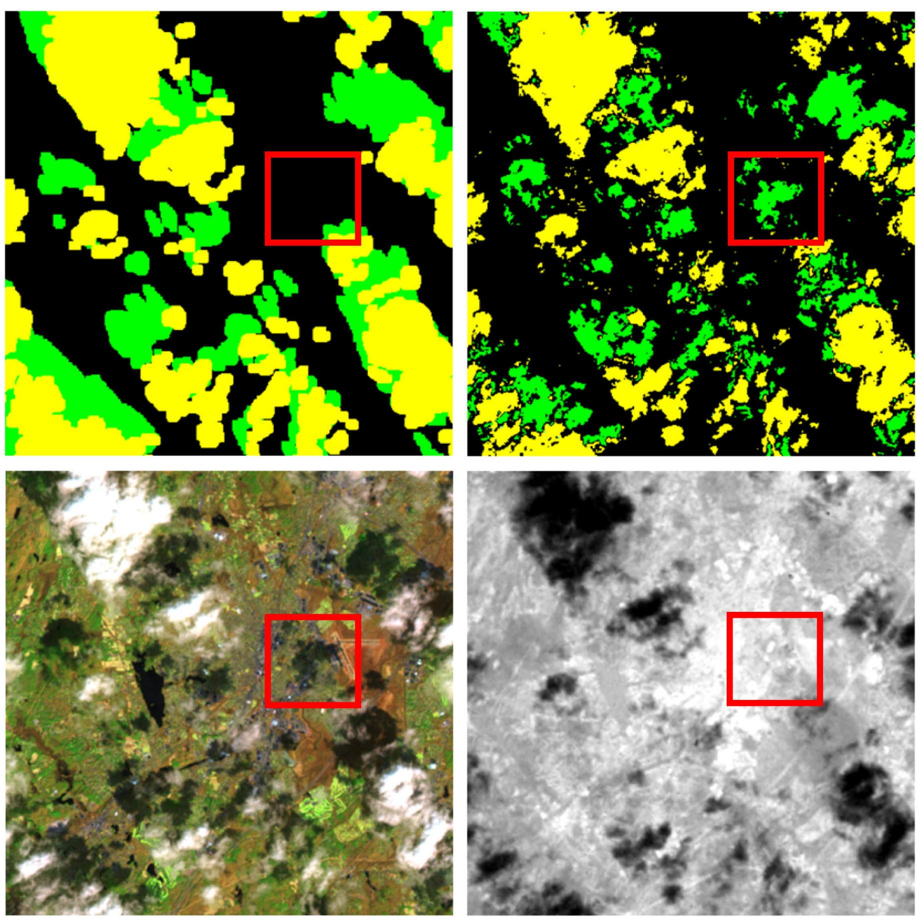

Fig. 15. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired March 20th 2006 — a scenario where two algorithms disagree for cloud shadows because of dark objects located between clouds and their shadows (Tmask works and Fmask fails). Note that the cloud shadows located within the red rectangle are missed by Fmask but captured by Tmask (the images follow the description in the caption for Fig. 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

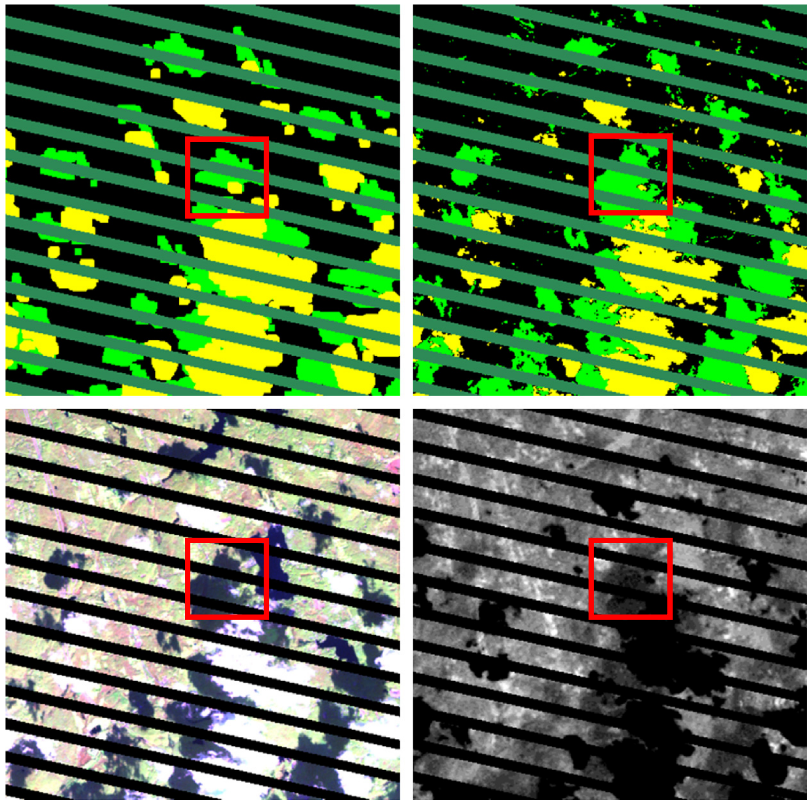

Fig. 16. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired October 14th 2006 — a scenario where two algorithms disagree for cloud shadows because of SLC-off problem (Tmask works and Fmask fails). Note that the cloud shadows located within the red rectangle are sliced into small pieces in Fmask but correctly identified in Tmask (the images follow the description in the caption for Fig. 6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

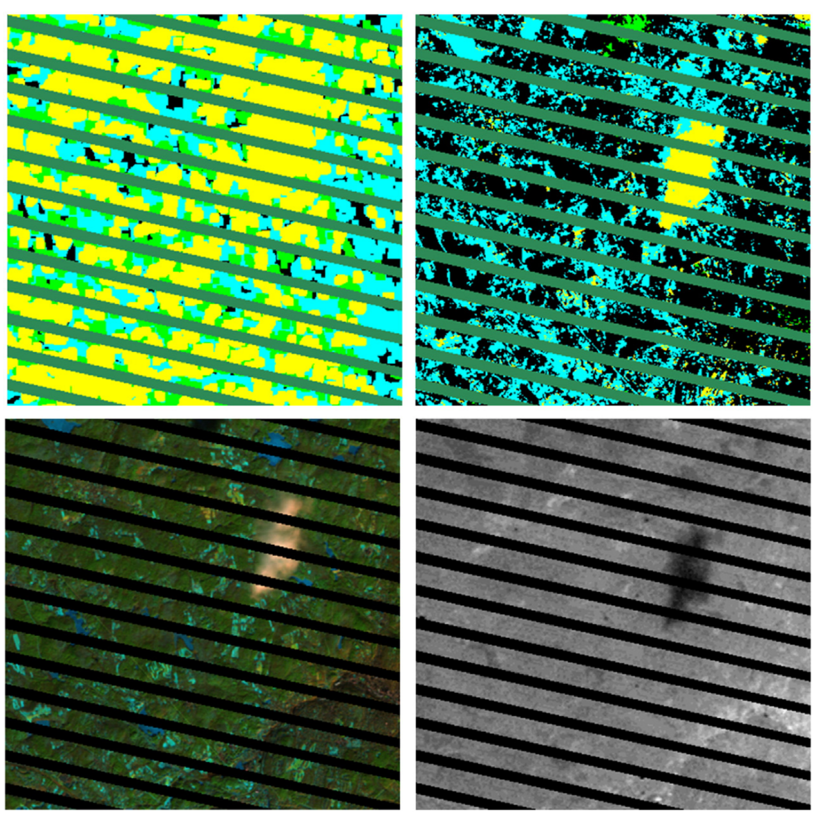

Fig. 17. Comparison of Fmask and Tmask results for a subset of Landsat images at Path 12 Row 31 acquired December 30th 2005 — a scenario where two algorithms disagree for clouds and snow (Tmask works and Fmask fails) (the images follow the description in the caption for Fig. 6).

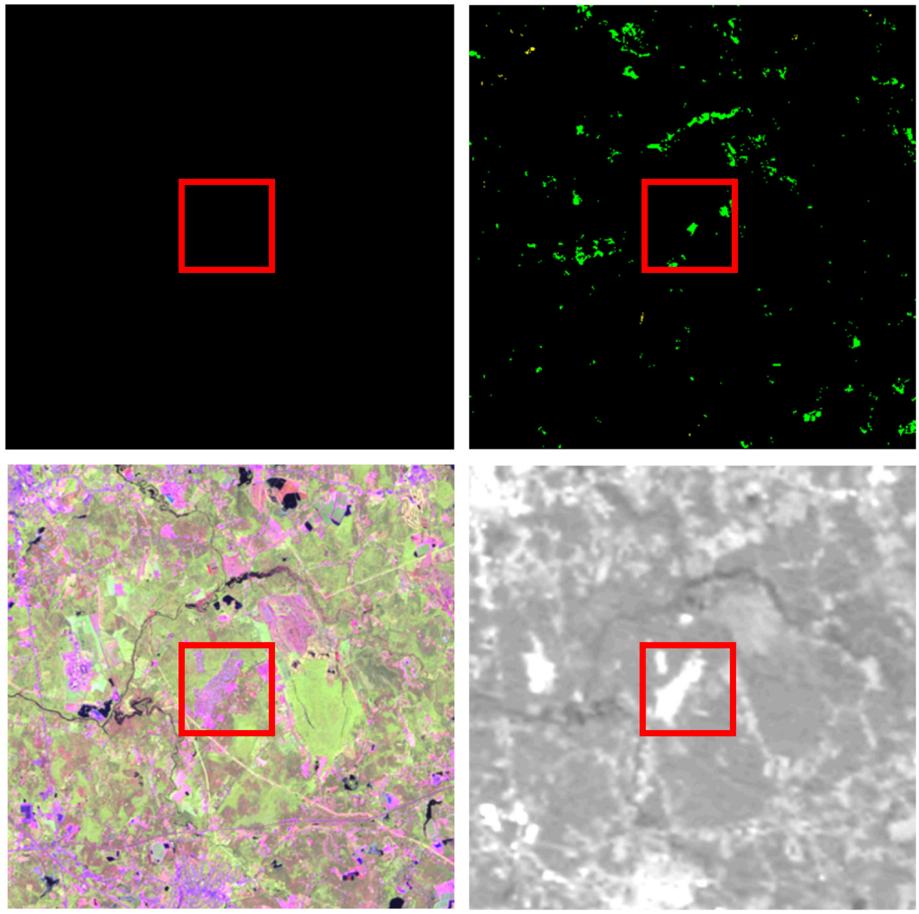

Fig. 18. Comparison of Fmask and Tmask results for a subset of Landsat image at Path 12 Row 31 acquired June 16th 2006 — scenario where two algorithms disagree for cloud shadows (Fmask works and Tmask fails). Note that within the red rectangle Tmask falsely identifies temporary wetness change as cloud shadow and Fmask does not. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Tmask does not use any global spectral thresholds as Fmask does; instead, it is location and time specific. By including spatial and temporal context (and with the help of the single-date algorithm), Tmask is able to identify all kinds of clouds at the pixel level. It can identify clouds that cause changes in Landsat observations large enough to influence time series analysis. Tmask is able to capture some of the extremely thin clouds that Fmask cannot, clouds that significantly change the spectral signal. The pixel-based character of the Tmask algorithm makes it capable of detecting clouds and cloud shadows as small as one Landsat pixel, information excluded by the object-based Fmask algorithm. Furthermore, with the added temporal information, only three optical bands are used, making the thermal band no longer necessary. This means Tmask will also work for Landsat-like sensors (i.e. Sentinel 2A and 2B) that do not have thermal bands.

The most significant improvement in Tmask is cloud shadow detection. Due to the complexity of object-based cloud and cloud shadow matching in Fmask, cloud shadow detection in Fmask is less accurate than cloud detection (Zhu & Woodcock, 2012). Cloud shadow detection in Tmask is not influenced by geometry-based matching of clouds and cloud shadows. The cloud shadows are detected by finding observations that are lower than model predictions in Band 4 and Band 5 TOA reflectance. Though the Tmask algorithm may not be able to identify some cloud shadows that do not make the two spectral bands dark enough to be identified as cloud shadows, this kind of omission error will not cause serious problems for most of remote sensing activities, including change detection, as the missed cloud shadows will not change the reflectance significantly.

Separating snow and clouds is often difficult for optical remote sensing because of their similar spectral signatures. The Tmask algorithm uses the extra temporal information coupled with a modified NLR model to separate snow and clouds. Though both clouds and snow will make Band 2 TOA reflectance higher, the influence of snow and clouds on Band 5 TOA reflectance is quite different. Tmask calculates a snow threshold (*T_snow*) for Band 5 TOA reflectance for each pixel based on a modified NLR model. Tmask showed better results in snow and cloud detection compared to Fmask. This modified NLR model may also be used to estimate fractional snow cover with better results than the traditional NLR algorithm because instead of assuming that there is only one constant value for snow-free pixels, we know what the Earth surface looks like in any particular place and time without snow.

There are also limitations in the Tmask algorithm. First, in order to estimate a time series model, Tmask needs 15 “clear” observations (the observations can be from different years), making it less applicable in places with high cloud cover or persistent snow. Second, Tmask may falsely identify some ephemeral changes as clouds or cloud shadows. While this may benefit detecting more persistent land cover changes, it may not be good for applications that need to monitor more ephemeral changes. Third, a delay will always exist in the Tmask algorithm because it needs more observations to respond to future land cover change. Therefore, it is advised to use Tmask to generate cloud, cloud shadow, and snow masks for the images in the middle of the time series. For example, if we have 30 Landsat images used for this Tmask algorithm, only cloud, cloud shadow, and snow masks for the 10 images

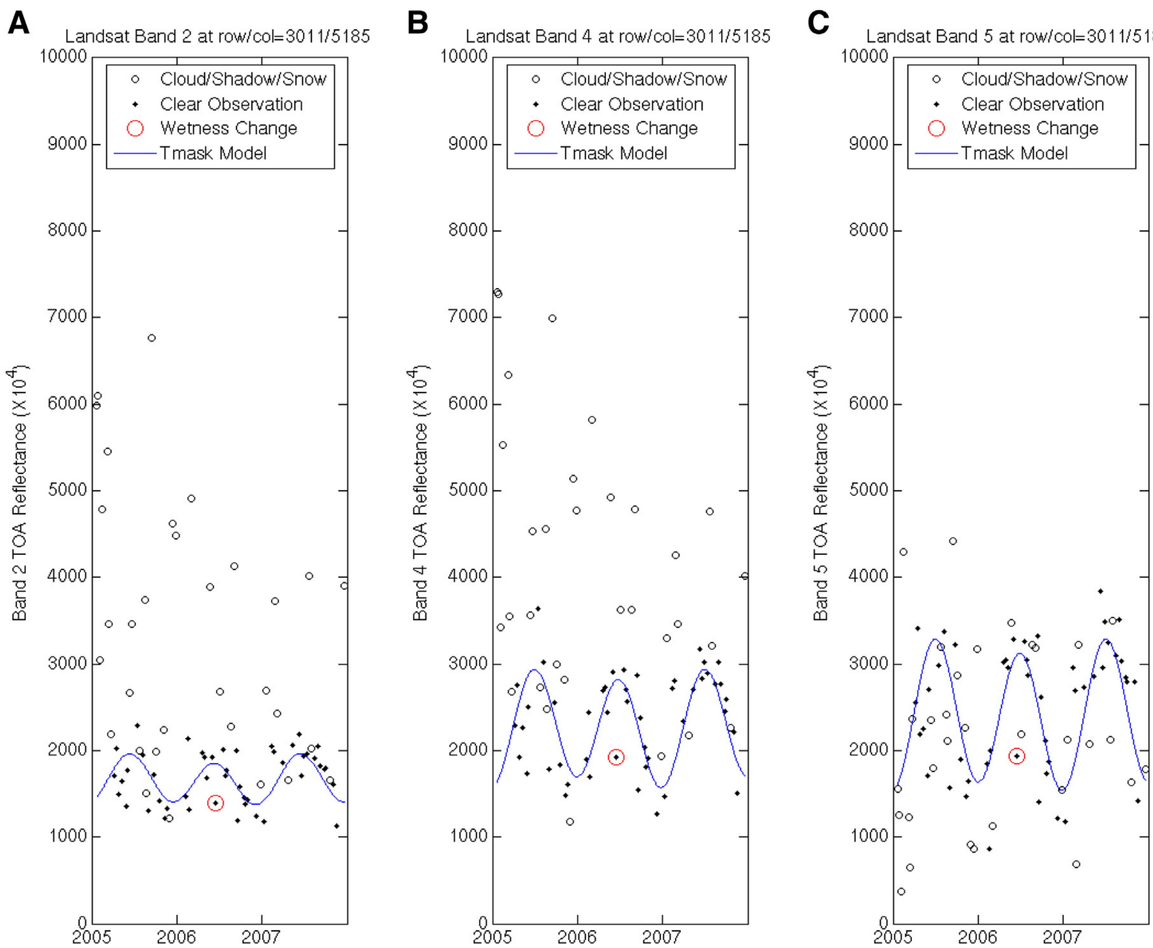

Fig. 19. Times series of Band 2 (A), 4 (B), and 5 (C) TOA reflectance of the pixel located at the center of red rectangle in Fig. 18. Clear pixels are black dots. Clouds, cloud shadows, and snow are black circles, and the Tmask mistake is red circle. The model estimated values are in the blue line. Note that the temporary wetness change (red circle) has made all three bands change significantly, but the next observation is back to normal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

located in the middle of the time period is suggested for use in remote sensing activities like change detection. Finally, Tmask also relies on the assumption that only one land cover change takes place during model estimation because the coefficients responsible for land cover change can only respond to land cover change once. Therefore, Tmask should be applied for masking multitemporal images within a few years (1–5 years) to avoid problems in areas where multiple land cover changes occur. One important future effort will be to test the Tmask algorithm for other regions with different environments. In this paper, the Tmask algorithm works well for one of the New England site, but it hasn't been tested for areas that are more variable, such as semi-arid areas or croplands. For places where the Earth surface is not stable, the time series model estimation may not be accurate and we may need the single-date algorithm results for places where the data variation is large.

In conclusion, adding the use of multitemporal images improves detection of clouds, cloud shadows and snow in Landsat images. Free access to the archive of Landsat images at Earth Resources Observation and Science (EROS) Data Center is revolutionizing many applications using Landsat data, and this is yet another example. In essence, what this algorithm does is use the history of a place to help evaluate whether it is covered by clouds, cloud shadows or snow in an individual image. This sort of contextual information, in this case in the form of a time series model, particularly helps with detection of thin clouds and their shadows, and separation of clouds and snow.

Acknowledgments

We gratefully acknowledge the supports of NASA Earth Science U.S. Participating Investigator Program for Enhancing Compatibility of Sentinel 2 and Landsat Products for Improved Monitoring of the Earth System (grant number NNX11AE18G) and the supports of USGS Landsat Science Team Program for Better Use of the Landsat Temporal Domain: Monitoring Land Cover Type, Condition, and Change (grant number G11PS00422).

References

Andersen, T. (1982). Operational snow mapping by satellites. *Hydrological aspects of alpine and high mountain areas. Proceedings of the Exeter symposium.* (pp. 149–154).
Chander, G., Markham, B.L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, *113*(5), 893–903.
Collins, J. B., & Woodcock, C. E. (1996). An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing of Environment*, *56*, 66–77.
Derrien, M., & Le Gléau, H. (2010). Improvement of cloud detection near sunrise and sunset by temporal-differencing and region-growing techniques with real-time SEVIRI. *International Journal of Remote Sensing*, *31*(7), 1765–1780.
Dozier, J. (1984). Snow reflectance from Landsat-4 thematic mapper. *IEEE Transactions on Geoscience and Remote Sensing*(3), 323–328.
Dozier, J. (1989). Spectral signature of alpine snow cover from the Landsat Thematic Mapper. *Remote Sensing of Environment*, *28*, 9–22.
Dozier, J., & Painter, T. H. (2004). Multispectral and hyperspectral remote sensing of alpine snow properties. *Annual Review of Earth and Planetary Sciences*, *32*, 465–494.

DuMouchel, W. H., & O'Brien, F. L. (1989). Integrating a robust option into a multiple regression computing environment. *Computer science and statistics: Proceedings of the 21st symposium on the interface*. Alexandria, VA: American Statistical Association.

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 Sensing of Environment*, *134*, 50–65.

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 Sensing of Environment*, *114*(8), 1747–1755.

Healey, S. P., Cohen, W. B., Yang, Z., & Krankina, O. N. (2005). Comparison of Tasseled Cap-based Landsat data structure for use in forest disturbance detection. *Remote Sensing of Environment*, *97*, 301–310.

Heiberger, R. M., & Becker, R. A. (1992). Design of an S function for robust regression using iteratively reweighted least squares. *Journal of Computational and Graphical Statistics*, *1*(3), 181–196.

Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Theory and Methods*, *A6*, 813–827.

Huang, C., Goward, S. N., Masek, J. G., Thomas, N., Zhu, Z., & Vogelmann, J. E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, *114*, 183–198.

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

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

Jin, S., Homer, C., Yang, L., Xian, G., Fry, J., Danielson, P., et al. (2013). Automated cloud and shadow detection and filling using two-date Landsat imagery in the USA. *International Journal of Remote Sensing*, *34*(5), 1540–1560.

Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B. C., Li, R. R., & Flynn, L. (1997). The MODIS 2.1-μm channel-correlation with visible reflectance for use in remote sensing of aerosol. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(5), 1286–1298.

Kennedy, R. E., Cohen, W. B., & Schroeder, T. A. (2007). Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sensing of Environment*, *110*(3), 370–386.

Klein, A. G., Hall, D. K., & Riggs, G. A. (1998). Improving snow cover mapping in forests through the use of a canopy reflectance model. *Hydrological Processes*, *12*(10-11), 1723–1744.

Lee, D. S., Storey, J. C., Choate, M. J., & Hayes, R. W. (2004). Four years of Landsat-7 on-orbit geometric calibration and performance. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(12), 2786–2795.

Liu, R., & Liu, Y. (2013). Generation of new cloud masks from MODIS land surface reflectance products. *Remote Sensing of Environment*, *133*, 21–37.

Lyapustin, A., Wang, Y., & Frey, R. (2008). An automatic cloud mask algorithm based on time series of MODIS measurements. *Journal of Geophysical Research [Atmospheres]*, *113*(D16) (1984–2012).

Markham, B.L., Storey, J. C., Williams, D. L., & Irons, J. R. (2004). Landsat sensor performance: History and current status. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(12), 2691–2694.

Masek, J. G., Honzak, M., Goward, S. N., Liu, P., & Pak, E. (2001). Landsat-7 ETM+ as an observatory for land cover: Initial radiometric and geometric comparisons with Landsat-5 Thematic Mapper. *Remote Sensing of Environment*, *78*(1), 118–130.

Masek, J. G., Huang, C., Wolfe, R., Cohen, W., Hall, F., Kutler, J., et al. (2008). North American forest disturbance mapped from decadal Landsat record. *Remote Sensing of Environment*, *112*, 2914–2926.

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

O'Leary, D. P. (1990). Robust regression computation using iteratively reweighted least squares. *Society for Industrial and Applied Mathematics*, *11*(3), 466–480.

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

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

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

Salomonson, V. V., & Appel, I. (2004). Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sensing of Environment*, *89*(3), 351–360.

Scaramuzza, P. L., Bouchard, M.A., & Dwyer, J. L. (2012). Development of the Landsat data continuity mission cloud-cover assessment algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(4), 1140–1154.

Street, J. O., Carroll, R. J., & Ruppert, D. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *American Statistical Association*, *42*(2), 152–154.

Tseng, D. C., Tseng, H. T., & Chien, C. L. (2008). Automatic cloud removal from multi-temporal SPOT images. *Applied Mathematics and Computation*, *205*(2), 584–600.

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

Vikhamar, D., & Solberg, R. (2003). Subpixel mapping of snow cover in forests by optical remote sensing. *Remote Sensing of Environment*, *84*(1), 69–82.

Vogelmann, J. E., Tolk, B., & Zhu, Z. (2009). Monitoring forest changes in the southwestern United States using multitemporal Landsat data. *Remote Sensing of Environment*, *113*, 1739–1748 (110, 370–386).

Wang, B., Ono, A., Muramatsu, K., & Fujiwara, N. (1999). Automated detection and removal of clouds and their shadows from Landsat TM images. *IEICE Transactions on Information and Systems*, *82*(2), 453–460.

Warren, S. G., & Wiscombe, W. J. (1980). A model for the spectral albedo of snow. II: Snow containing atmospheric aerosols. *Journal of the Atmospheric Sciences*, *37*(12), 2734–2745.

Wiscombe, W. J., & Warren, S. G. (1980). A model for the spectral albedo of snow. I: Pure snow. *Journal of the Atmospheric Sciences*, *37*(12), 2712–2733.

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

Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, *21*(3), 311–332.

Wulder, M.A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, *122*, 2–10.

Xin, Q., Woodcock, C. E., Liu, J., Tan, B., Melloh, R. A., & Davis, R. E. (2012). View angle effects on MODIS snow mapping in forests. *Remote Sensing of Environment*, *118*, 50–59.

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

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

Zhu, Z., Woodcock, C. E., & Olofsson, P. (2012). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, *122*, 75–91.