

CLOUD DETECTION BY INTER-BAND PARALLAX AND A-CONTRARIO VALIDATION

R. Grompone von Gioi^{1,*}, C. Hessel^{1,2}, T. Dagobert¹, J. M. Morel¹, C. de Franchis^{1,2}

¹ Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, Gif-sur-Yvette, France

² Kayrros SAS

KEY WORDS: cloud detection, inter-band parallax, satellite optical images, a-contrario theory

ABSTRACT:

The presence of clouds often causes detection errors in automatic optical satellite images analysis. Cloud detection is thus an important pre-processing step. This work exploits the apparent movement of clouds relative to the ground that is observed between bands due to the parallax effect in push-broom satellites. A simple optical flow is computed by correlation of the gradient direction on a local window. Then, a statistical validation test based on the a-contrario theory is used to reject spurious detections. The proposed method achieves performances in the same order of magnitude as the state of the art, while relying on less input information.

1. INTRODUCTION

Assessing ground visibility is an important step in optical satellite images analysis. Indeed, the presence of clouds and haze concealing the surface of the Earth often causes detection errors in automatic image analysis. This task is usually addressed as a cloud detection problem, where the image pixels are classified into classes such as ground, clouds, cirrus, snow, haze, cloud shadows, etc. (Chandran and Jojo, 2015, Mahajan and Fataniya, 2019). Satellite cloud detection often exploits spectral bands specifically designed for cloud detection (Irish, 2000, Zhang et al., 2001, Irish et al., 2006, Scaramuzza et al., 2012, Chandran and Jojo, 2015, Taravat et al., 2015, Hollstein et al., 2016).

Another approach for cloud detection is based on temporal information (Hagolle et al., 2010, Goodwin et al., 2013, Zhu and Woodcock, 2014, Frantz et al., 2015, Dagobert et al., 2019b, Dagobert et al., 2020c, Grompone von Gioi et al., 2020). However, time-series-based method are not adapted, for example, to change detection, as changes can be mis-classified as clouds. Relying on yet another technique, the cloud detectors proposed in (Rodriguez et al., 2021) uses a convolutional neural network to detect clouds in satellite images with a single band.

In push-broom satellites, the inter-band delay allows cloud detection by parallax analysis of the spectral bands (Shin and Pollard, 1996, Panem et al., 2005, Manizade et al., 2006, Wu et al., 2016, Sinclair et al., 2017, Frantz et al., 2018, Dagobert et al., 2020a). In this kind of satellite, the different bands are acquired by linear captors placed at slightly different physical positions in the focal plane; thus, the same point on Earth is acquired with a short time delay between bands, corresponding to different viewpoints, see Figure 1. The bands are then registered to produce a coherent multi-band image. Clouds are at different height than the ground; by the parallax effect, they will appear as displaced relative to the ground, and the more the higher the altitude. As a consequence, clouds appear as objects with an apparent displacement between bands. The general idea to detect clouds by inter-band parallax is thus to estimate an optical flow between spectral bands and then to identify regions with a coherent apparent movement, see Figure 2.

The United States Geological Survey (USGS) classification maps distributed for Landsat products are currently generated by the

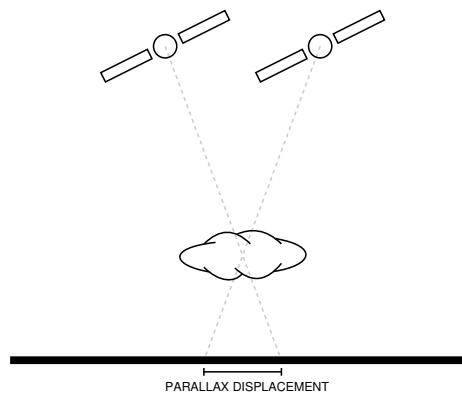


Figure 1. The clouds, at higher altitude than the ground, result in an apparent displacement relative to the ground due to parallax.

Fmask algorithm. It was introduced in 2012 (Zhu and Woodcock, 2012) for the Landsat satellites and improved on the ACCA algorithm (Irish, 2000, Irish et al., 2006). It consists in a series of tests on the bands values, on band ratios, or on normalized difference indices. It was then revised and improved in a series of papers. In (Zhu et al., 2015), the threshold were refined and support for Landsat-8 and Sentinel-2 was added, notably with the inclusion of the cirrus band. The Cloud Displacement Index (CDI) was later introduced for Sentinel-2 (Frantz et al., 2018) using bands B07, B08 and B8A. Bands B08 and B8A are closer spectrally than bands B07 and B8A, but have a stronger parallax effect. The ratio B08/B8A is thus expected to be uniform and close to one on the ground but not on the clouds; on the contrary, the ratio B07/B8A is expected to be uniform and close to one on the clouds but not on the ground. The authors therefore proposed a normalized difference index based on the local variations of these ratios. The latest updates (Qiu et al., 2017, Qiu et al., 2019) revised the thresholds and included auxiliary data: a water occurrence map, and a digital elevation model used to normalize the cirrus values and improve detection in mountainous areas.

The Sentinel-2 products at level 2A distributed by the European Space Agency (ESA) contain cloud masks generated by the Sen2cor processor (Müller-Wilm, 2012, Main-Knorn et al., 2017). Similarly to Fmask, the method classifies the pixels in several classes including three for clouds (medium probability, high probability, and thin cirrus), cloud shadow, vegetation, wa-

* Corresponding author (grompone@ens-paris-saclay.fr)

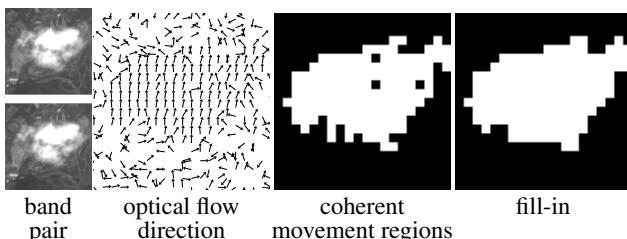


Figure 2. Steps of the proposed method.

ter, snow, and other classes for unclassified or unusable pixels. Using top-of-atmosphere reflectance values, a series of thresholds are applied on the bands values and bands ratios. The parallax effect is not exploited for the cloud detection.

The Sentinel-2 cloud detector proposed in (Dagobert et al., 2020b) uses five criteria, including three based on the parallax effect. Optical flow is used to estimate the inter-band displacement. Empirical distributions are obtained on the Hollstein dataset (Hollstein et al., 2016), features are merged and the a-contrario framework is used to set a single detection threshold.

Here we propose an algorithm for cloud detection relying entirely on the parallax effect between bands, see Figure 2. The method is contrast-invariant, so it can be used on any set of bands, regardless of whether they share the same spectral profile or not, provided that the parallax effect is indeed present.

Cloud parallax is not the only possible source of apparent displacement between bands. Any high altitude object, such as mountains, will also produce apparent motion by parallax. In addition, really moving objects such as airplanes or cars result in displacement between bands. (Clouds could also be moving due to wind and in very rare cases these two effects cancel one another out.) Finally, the band registration is not perfect; the residual registration error is a constant spatial shift of bands, and thus equivalent to an inter-band displacement. The fundamental hypothesis is thus that the bands are well registered up to a certain precision; observed inter-band displacements larger than the registration precision are considered parallax effects.

There are two key aspects in cloud detection by parallax: the computation of the optical flow and the criterion for deciding whether the displacement of a region is coherent enough. There is a large literature on optical flow; this application requires an efficient (given the large size of satellite images) and accurate method. An additional requirement in our case is to obtain spatially decorrelated values (as needed for the proposed statistical validation). We propose to use a simple optical flow computed by gradient direction correlation on a local window.

Concerning the control of false detections, the a-contrario statistical approach (Desolneux et al., 2000, Desolneux et al., 2008) was introduced to provide a well founded mechanism for setting detection thresholds while producing few false detections. The method described here uses a new a-contrario statistical test to decide whether a group of pixels shows a coherent movement from one band to another.

This paper is organized as follows. Section 2 describes the simple optical flow method used to assess the apparent inter-band movement. Then, Section 3 introduces briefly the a-contrario approach which is then applied in Section 4 for deciding whether a group of pixels shows a coherent apparent movement between bands. The proposed method is illustrated with some experiments in Section 5. Finally, Section 6 concludes the paper.

2. OPTICAL FLOW

The optical flow is the distribution of apparent motion between two images (Horn and Schunck, 1981). There is an extensive literature and variety of methods to compute the optical flow (Garamendi et al., 2019, Dagobert et al., 2019a, Gamonal et al., 2019, Monzón et al., 2016, Garrido and Kalmoun, 2015, Jara-Wilde et al., 2015, Meinhardt-Llopis et al., 2013, Sánchez Pérez et al., 2013). In our case, there are two characteristics of the present problem which determine our choice.

Most optical flow methods use some kind of regularization, which helps to compute an accurate result. However, the validation criterion described below to decide whether a cloud is present or not is based on the coherence of the apparent movement directions. Using regularization in the computation of the optical flow makes this criterion harder, as it requires deciding if an observed coherence is due to image contents, or to the imposed regularization. To simplify our setting, the proposed approach uses a simple optical flow computed independently at each point, without any regularization.

The second characteristic of our setting is that we need to compute optical flow between different spectral bands, while most optical flow methods are designed to work on successive frames of a video. The usual setting allows to assume that the images to be compared have a similar dynamic. In our case, however, bands are specially designed to capture different information, which implies that the dynamics are usually different. One possible solution is to normalize bands before applying the optical flow. Instead, we decided to use a simple approach by computing the optical flow with a contrast-invariant similarity measure.

This similarity measure is the correlation of the gradient direction on a window. Given a single band image I , the normalized gradient at pixel (x, y) is the unit vector $\vec{g}(x, y)$ given by

$$\vec{g}(x, y) = \frac{\nabla I(x, y)}{|\nabla I(x, y)|}. \quad (1)$$

The gradient is computed by centered differences. Given two bands of an image I_a and I_b , the corresponding normalized gradients are noted \vec{g}_a and \vec{g}_b , respectively. Given a window size $(2W + 1) \times (2W + 1)$, the correlation between position (x, y) of band a and position (x', y') of band b is defined by

$$c_{ab}(x, y, x', y') = \sum_{i=-W}^W \sum_{j=-W}^W \vec{g}_a(x+i, y+j) \cdot \vec{g}_b(x'+i, y'+j),$$

where \cdot denotes the dot product. As the dot product of two unit vectors takes values in $[-1, 1]$, the correlation at each point takes values between $-(2W + 1)^2$ and $(2W + 1)^2$. A value $(2W + 1)^2$ corresponds to a perfect match, when the gradients \vec{g}_a and \vec{g}_b have exactly the same directions in the whole window. A value $-(2W + 1)^2$ would be observed in case of a perfect contrast inversion. A correlation close to zero indicates a bad match. On digital images this correlation is only defined for coordinates (x, y) and (x', y') in the discrete domain of the images; here we will assume that these coordinates are integer values.

The optical flow, or apparent movement, at (x, y) is now obtained as the displacement $(\Delta x, \Delta y)$ with the best correlation:

$$\vec{m}_{ab}(x, y) = \arg \max_{(\Delta x, \Delta y) \in \mathbb{N}^2} c_{ab}(x, y, x + \Delta x, y + \Delta y). \quad (2)$$

In practice, we are only interested in the displacement produced by parallax, which is limited by the speed and height of the satellite, and by the maximal possible height of clouds. This allows to restrict the search to $(\Delta x, \Delta y) \in [-D, D]^2$, for an appropriate D , greatly reducing the computational cost.

Notice that the computed correlation coefficient, and thus the apparent movement, only depends on the gradient direction and not on its magnitude. Thus, it is contrast-invariant provided that there is no inversion of contrast. This is particularly important here to handle the different spectral response of different bands.

A final step is to compute a simple sub-pixel refinement on the apparent movement $\vec{m}_{ab}(x, y)$. This interpolation is performed independently for the horizontal and vertical components. Assuming that (x, y) are the integer coordinates of the maximal correlation, the sub-pixel interpolation is given by:

$$M_{ab}^x = m_{ab}^x + \frac{m_{ab}^x(x-1, y) - m_{ab}^x(x+1, y)}{2m_{ab}^x(x-1, y) - 4m_{ab}^x(x, y) + 2m_{ab}^x(x+1, y)}$$

$$M_{ab}^y = m_{ab}^y + \frac{m_{ab}^y(x, y-1) - m_{ab}^y(x, y+1)}{2m_{ab}^y(x, y-1) - 4m_{ab}^y(x, y) + 2m_{ab}^y(x, y+1)}$$

where (m_{ab}^x, m_{ab}^y) are the components of \vec{m}_{ab} and the estimated sub-pixel apparent movement is $\vec{M}_{ab} = (M_{ab}^x, M_{ab}^y)$. This corresponds to the addition, to the initial estimation, of a second term computed as a quadratic interpolation; this sub-pixel interpolation was used for example by Devernay in the context of gradient magnitude interpolation (Devernay, 1995).

The sub-pixel interpolation step is mainly used to reject displacements with magnitude smaller than the accuracy of the registration between bands, which we note T . Indeed, those displacements could be just the result of an imprecise registration of bands, and in such a case, the apparent movement vectors $\vec{m}_{ab}(x, y)$ will be of course coherent. When $|\vec{M}_{ab}(x, y)| < T$ the apparent movement is declared undefined at position (x, y) .

On multiple-band images, the apparent movement can be computed between any pair of bands sharing the same resolution. A pair of bands of different resolution can also be used by down-sampling the one with higher resolution or by up-sampling the one with lower resolution. More exploration is needed to determine the relative performances of these options.

3. THE A-CONTRARIO APPROACH

The a-contrario theory (Desolneux et al., 2000, Desolneux et al., 2008) is a statistical framework used to set detection thresholds automatically in order to control the number of false detections. It is based on the non-accidentalness principle (Witkin and Tenenbaum, 1983, Lowe, 1985) which informally states that an observed structure is meaningful only when the relation between its parts is too regular to be the result of an accidental arrangements of independent parts. In the words of D. Lowe, “we need to determine the probability that each relation in the image could have arisen by accident, $P(a)$. Naturally, the smaller that this value is, the more likely the relation is to have a causal interpretation” (Lowe, 1985, p. 39).

A stochastic background model \mathcal{H}_0 needs to be defined, where the structure of interest is not present and can only arise as an accidental arrangement. For example, many image matching methods are based on the orientation of the image gradient (Lowe, 2004, Cao et al., 2008, Rabin et al., 2009, Grompone von Gioi

and Pătrăucean, 2015). Some geometrical feature detection methods such as line segments (Grompone von Gioi, 2014) are also based on the image gradient. In our present case we will evaluate the orientation of the apparent movements. In all these cases, a good background model \mathcal{H}_0 assumes that the orientations at each pixel are independent random variables, uniformly distributed in $[-\pi, \pi]$. Under this background model, a region of the image where the orientation follows a regular structure would be a rare accident and is detected as such.

We also need to define a family of events of interest \mathcal{T} . For feature detection the family of events is the set of all the geometrical events considered, i.e., all the line segments considered in the image domain. Then, we need to assess the accidentalness of a candidate feature. For example, if a line segment is present in an image, the gradient orientation at the corresponding position would be orthogonal to the line segment (Grompone von Gioi, 2014). Given a candidate line segment, one measures how well the image gradient corresponds to the candidate event, and evaluates the probability of observing such a good agreement by chance in the background model \mathcal{H}_0 . A rough agreement could arise just by chance and thus does not correspond to an interesting event; conversely, a very good agreement would be rare and suggests the presence of a structure instead of just a lucky accident. In other words, when this probability is small enough, there exists evidence to reject the null hypothesis and declare the event meaningful. However, one needs to consider that multiple candidates are tested. If 1000 tests were performed, for example, it would not be surprising to observe among them one event that appears with probability 0.001 under random conditions. The number of tests N_T (i.e., the size of the family \mathcal{T}) needs to be included as a correction term, as it is done in the statistical multiple hypothesis testing framework (Gordon et al., 2007).

Following the a-contrario methodology (Desolneux et al., 2000, Desolneux et al., 2008), we define the *Number of False Alarms* (NFA) of an event e observed up to a given precision ρ as

$$\text{NFA}(e, \rho) = N_T \cdot \mathbb{P}_{\mathcal{H}_0}(e_\rho), \quad (3)$$

where the right hand term is the probability of obtaining the event e up to precision ρ in the background model \mathcal{H}_0 . The smaller the NFA, the more unlikely the event e is to be observed by chance in the background model \mathcal{H}_0 ; thus, the more meaningful. The a-contrario approach prescribes accepting as valid detections the candidates with $\text{NFA} < \varepsilon$ for a predefined value ε . It can be shown (Desolneux et al., 2000, Desolneux et al., 2008) that under \mathcal{H}_0 , the expected number of tests with $\text{NFA} < \varepsilon$ is bounded by ε . As a result, ε gives an a priori estimate of the mean number of false detections under \mathcal{H}_0 . In many practical applications, including the present one, the value $\varepsilon = 1$ is adopted. Indeed, it allows for less than one false detection on an image or on a set of images, which is usually quite tolerable.

4. CLOUD DETECTION

Given an apparent movement \vec{M}_{ab} between bands a and b , we would like to know whether the direction of movement is coherent in the connected region of pixels R . As we will see, regions with coherent apparent movement according to the proposed approach are large enough and usually correspond to clouds.

For this, we will consider regions R in which all the pixels share a given apparent movement orientation θ , up to a given angular tolerance ρ . This is the event e_ρ to be considered in our

a-contrario formulation using Eq. 3. As explained before, a natural background model \mathcal{H}_0 for this problem is that the apparent movement orientations at each pixel are independent random variables, uniformly distributed in $[-\pi, \pi]$. Nevertheless, given that the apparent movement \tilde{M}_{ab} was computed by the procedure described in Section 2, there is some correlation between neighboring pixels. To ensure that the independence hypothesis is reasonable, the apparent movement \tilde{M}_{ab} field will be sub-sampled at distances W (half of the correlation window size). If the original images were of size $X \times Y$, then the sub-sampled ones are $U \times V$, with $U = X/W$ and $V = Y/W$.

In this setting, the probability under \mathcal{H}_0 that all the pixels of a region R have an orientation coherent to θ up to precision ρ is

$$\mathbb{P}_{\mathcal{H}_0}(e_\rho) = \left(\frac{\rho}{\pi}\right)^{|R|}, \quad (4)$$

where $|R|$ is the number of pixels in R counted considering the sub-sampling at distance W . Following the a-contrario framework, and according to Eq. 3, we will define the NFA associated to a candidate region R as

$$\text{NFA}(R, \theta, \rho) = N_T \cdot \mathbb{P}_{\mathcal{H}_0}(e_\rho) = N_T \cdot \left(\frac{\rho}{\pi}\right)^{|R|}. \quad (5)$$

To complete the formulation we still need to specify the family of tests \mathcal{T} and the corresponding number of tests N_T . Our aim is detecting clouds, which have connected shapes. We will use the notion of *4-connectivity*, in which a pixel (x, y) is connected to the pixels at coordinates $(x \pm 1, y)$ and $(x, y \pm 1)$. Groups of pixels connected under 4-connectivity correspond to the figures called *polyominoes* (Golomb, 1994, Gardner, 1960), see Figure 3. The exact number b_n of different polyomino configurations of given size n is not known in general, but there are good approximations of this number (Jensen and Guttmann, 2000). In our case, it is enough to use an estimate of the order of magnitude, so the approximate formula given in (Jensen and Guttmann, 2000) is sufficient for our needs. It reads

$$b_n \approx \alpha \frac{\beta^n}{n}, \quad (6)$$

where $\alpha \approx 0.316915$ and $\beta \approx 4.062570$. We need to consider that each particular polyomino may be placed at any position in the image; thus we need to multiply b_n by UV to consider all possible placements (this estimation is not exact as it counts some polyominoes extending outside the image domain; nevertheless, it gives the order of magnitude). Also, we consider connected regions of different sizes, from 1 to UV , the latter being the case where all the pixels are part of one polyomino. This calculation is not exact due to the restrictions imposed by the total size of the image; e.g., among the b_{UV} polyominoes of size UV , only one is actually possible, the one using all the pixels in the images; again, this rough estimation is useful for our case. The number of tolerated false detections ε is divided into UV categories, and to each one we allow to produce $\frac{\varepsilon}{UV}$ false detections. In this way, the total number of false detections, including the ones obtained for polyomino regions of size from 1 to UV , will add up to $\sum_{n=1}^{UV} \frac{\varepsilon}{UV} = \varepsilon$, as desired. Thus, for a given region size, the test becomes

$$UV b_{|R|} \cdot \mathbb{P}_{\mathcal{H}_0}(e_\rho) < \frac{\varepsilon}{UV}. \quad (7)$$

This is equivalent to setting the number of tests to $U^2 V^2 b_{|R|}$ and

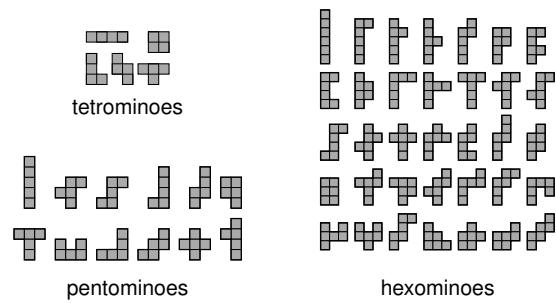


Figure 3. Polyominoes of four (tetrominoes), five (pentominoes) and six (hexominoes) elements. When rotations and reflections are not considered, there are 5 tetrominoes, 12 pentominoes and 35 hexominoes, as shown here. When rotations and reflections are also considered distinct, there are 19 tetrominoes, 63 pentominoes and 216 different hexominoes. Eq. (6) gives the right order of magnitude of these numbers: 21.6 for tetrominoes, 70.1 for pentominoes, and 237.4 for hexominoes.

comparing directly to ε . The final number of tests is then

$$N_T = U^2 V^2 b_{|R|}. \quad (8)$$

The NFA of a candidate region R is thus computed by

$$\text{NFA}(R, \theta, \rho) \approx U^2 V^2 \cdot \alpha \frac{\beta^{|R|}}{|R|} \cdot \left(\frac{\rho}{\pi}\right)^{|R|}. \quad (9)$$

As mentioned before, we set $\varepsilon = 1$. When a regions has $\text{NFA}(R, \theta, \rho) < 1$, we say that its orientations are significantly coherent. Each such regions is declared as a cloud detection.

As described, all possible connected regions R in the image should be evaluated, which would take an impossible amount of time. Instead, a greedy algorithm is used. Starting from a given pixel, the 4-connected neighbors are added when they agree with the orientation θ up to precision ρ . This process is iterated until no further pixel is added. Then, the NFA is computed and the region is kept as a detection or not depending on the NFA test described above. To accelerate the algorithm, after a region is evaluated, regardless of whether it is validated or not, its pixels are marked and they cannot be used again in further regions.

There are several options to specify the reference angle θ . When the direction of the satellite trajectory is known, then this should determine the reference angle θ . But this angle is not always easy to obtain. In that case, several angles could be tried, which would correspond to further testing in the a-contrario framework. A simple solution is to use the angle of the first pixel as the reference angle. In such a case, the first pixel cannot be counted in Eq. 4, where $|R|$ must be replaced by $|R| - 1$, counting all pixels in the region but excluding the first one.

Concerning the angular tolerance ρ , instead of selecting a particular value, we propose to use several values. After the optical flow is computed, which is the most expensive part of the method, the greedy search for regions is repeated for several values of ρ . In our implementation and experiments we use $\frac{\rho}{\pi} \in \{\frac{1}{40}, \frac{1}{20}, \frac{1}{10}, \frac{1}{5}, 0.3, 0.4\}$. The regions found on all those iterations are then merged to produce the final detection. In the a-contrario formulation, regions evaluated with different angular tolerance ρ must be counted as different tests. Then, the number of angular tolerance P (in our implementation $P = 6$) needs to be added as additional factor to the number of tests N_T .

The algorithm can also use more than two bands. If bands a , b and c are available, we can compute, for instance, the apparent movement fields \vec{M}_{ab} and \vec{M}_{bc} . To comply with the independence assumption on the computed apparent movement vectors, each apparent movement field should use at least one band not used on the others. In the previous example, the apparent movement field \vec{M}_{ac} would be correlated with other two and should be excluded. The decision of which band pairs to use depends on characteristics of each satellite. First, the resolution of the bands should be the same (or be properly re-sampled). Second, the inter-band delay of different pairs are to be evaluated. If bands a , b , c , d , e and f are available and are captured in that order, we could compute the apparent movement fields \vec{M}_{ab} , \vec{M}_{cd} and \vec{M}_{ef} ; this results in three apparent movement fields with similar inter-band delay. Another option is computing the apparent movement fields \vec{M}_{ab} , \vec{M}_{ac} , \vec{M}_{ad} , \vec{M}_{ae} and \vec{M}_{af} , in which the inter-band delay is larger except in the first one. Nevertheless, it is important to compute the apparent movement in all pairs in a coherent order, as \vec{M}_{ab} and \vec{M}_{ba} will have opposite directions.

Given a set of N band pairs and the corresponding apparent movement fields \vec{M}_1 to \vec{M}_N , the same greedy algorithm is used but now verifying that all the N apparent movements are coherent with θ before adding a pixel to the region R . This implies two changes to the NFA formulation. First, as N apparent movement vectors need to agree with θ up to precision ρ for every pixel of $|R|$, the probability of the event in \mathcal{H}_0 is now

$$\mathbb{P}_{\mathcal{H}_0}(e_\rho) = \left(\frac{\rho}{\pi}\right)^{N|R|}. \quad (10)$$

(Notice that there is no need to multiply $|R|$ by N in the count of polyominoes; the region size is the same.) Second, each of the N apparent movements of the initial pixel could be used as the reference angle θ , so the number of band pairs N needs to be multiplied to the number of tests N_T .

All in all, the general formulation of the NFA is

$$\text{NFA}(R, \theta, \rho) \approx U^2 V^2 \cdot N \cdot P \cdot \alpha \frac{\beta^{|R|}}{|R|} \cdot \left(\frac{\rho}{\pi}\right)^{N|R|-1}, \quad (11)$$

when one pixel of one pair determines the reference θ , or

$$\text{NFA}(R, \theta, \rho) \approx U^2 V^2 \cdot P \cdot \alpha \frac{\beta^{|R|}}{|R|} \cdot \left(\frac{\rho}{\pi}\right)^{N|R|}, \quad (12)$$

when the reference angle θ is provided by the known direction of satellite's movement.

There is a minimal region size $|R|$ which can lead the detection condition $\text{NFA} < 1$ which depends on U , V , P , N and ρ . Let consider an image of size 10000×10000 and $W = 10$; then $U = 1000$ and $V = 1000$. Using $P = 6$, $N = 1$ and the very fine precision $\frac{\rho}{\pi} = \frac{1}{40}$, the minimal region size $|R|$ leading to a detection is 12 pixels. This corresponds to a region of about 3×4 pixels on the sub-sampled image, or a region of size 30×40 pixels on the actual image. At Sentinel-2 resolution of 10m this corresponds, in turn, to a region of about 300×400 meters, which is a large object to have a coherent movement. This is an extremely precise case; in more typical cases, regions leading to detections are much larger. Such large objects with coherent apparent movement are almost always clouds.

As a final step, holes in the detection map are filled-in using a morphological closing (a dilation followed by an erosion) (Serra,

	Recall	Precision	B. Accuracy	Accuracy
Fmask 4.4	91.6	88.2	94.4	96.1
CDI	61.7	84.5	79.6	90.7
Dagobert	70.3	97.6	84.9	94.1
Ours	88.0	81.4	91.7	94.0

Table 1. Recall, Precision, Balanced Accuracy and Accuracy (all in %) of cloud detectors evaluated on 20 Sentinel-2 tiles with mainly opaque clouds from the dataset of (Baetens et al., 2019).

	Recall	Precision	B. Accuracy	Accuracy
Fmask 4.4	80.9	90.0	89.0	93.3
CDI	40.5	85.1	69.1	84.4
Dagobert	50.4	97.9	75.0	87.9
Ours	61.2	83.0	78.6	87.7

Table 2. Recall, Precision, Balanced accuracy and Accuracy (all in %) of cloud detectors evaluated on the 31 Sentinel-2 tiles from the dataset of (Baetens et al., 2019).

1982), see Figure 2. The structuring element is a square of width $2(W + 1) + 1$ (two pixels larger than the correlation window).

5. EXPERIMENTS

We evaluated our method on the Sentinel-2 cloud dataset from (Baetens et al., 2019). This dataset has seven classes: high clouds, low clouds, cloud shadows, land, water, snow, and no data. For the evaluation we merged the high and low clouds into one single cloud class, and all other categories into a second non-cloud class. We evaluated the Fmask 4.4 (Qiu et al., 2019) method and also measured the performance of the CDI criterion alone, which is used in Fmask and exploits the parallax effect in Sentinel-2 images. This index takes values in $[-1, 1]$; so as to get a binary map, all pixels with $\text{CDI} < -0.8$ were assigned to the cloud category. The same threshold is used in Fmask. In addition, we evaluated the parallax-based cloud detector from (Dagobert et al., 2020a). We evaluated the proposed algorithm with band pairs (B08, B07) and (B08, B8A). The parameters used were $W = 10$, $D = 20$ and $T = 0.2$. The code of the proposed method is publicly available¹.

We evaluated the different methods in terms of precision and recall. These measures are also called, respectively, user's accuracy and producer's accuracy. They are defined by

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (13)$$

where TP, FP are the true and false positives, and TN, FN the true and false negatives. As a global measure of performance we use the balanced accuracy, as is appropriate when the population are unbalanced (Sokolova et al., 2006). We also provide the Accuracy, since it is often used in the literature.

$$\text{B. Accuracy} = \frac{\text{TPR} + \text{TNR}}{2}, \quad \text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{P} + \text{N}}, \quad (14)$$

where TPR is the true positive rate (the recall) and TNR the true negative rate TN/N , with $\text{N} = \text{TN} + \text{FP}$ and $\text{P} = \text{TP} + \text{FN}$.

Quantitative results are reported in Table 1. Since our detector cannot detect transparent clouds, we evaluated all methods only on tiles containing a majority of opaque clouds. We manually selected tiles that met this criterion. We nevertheless provide

¹ <https://github.com/rafael-grompone-von-gioi/pcd>

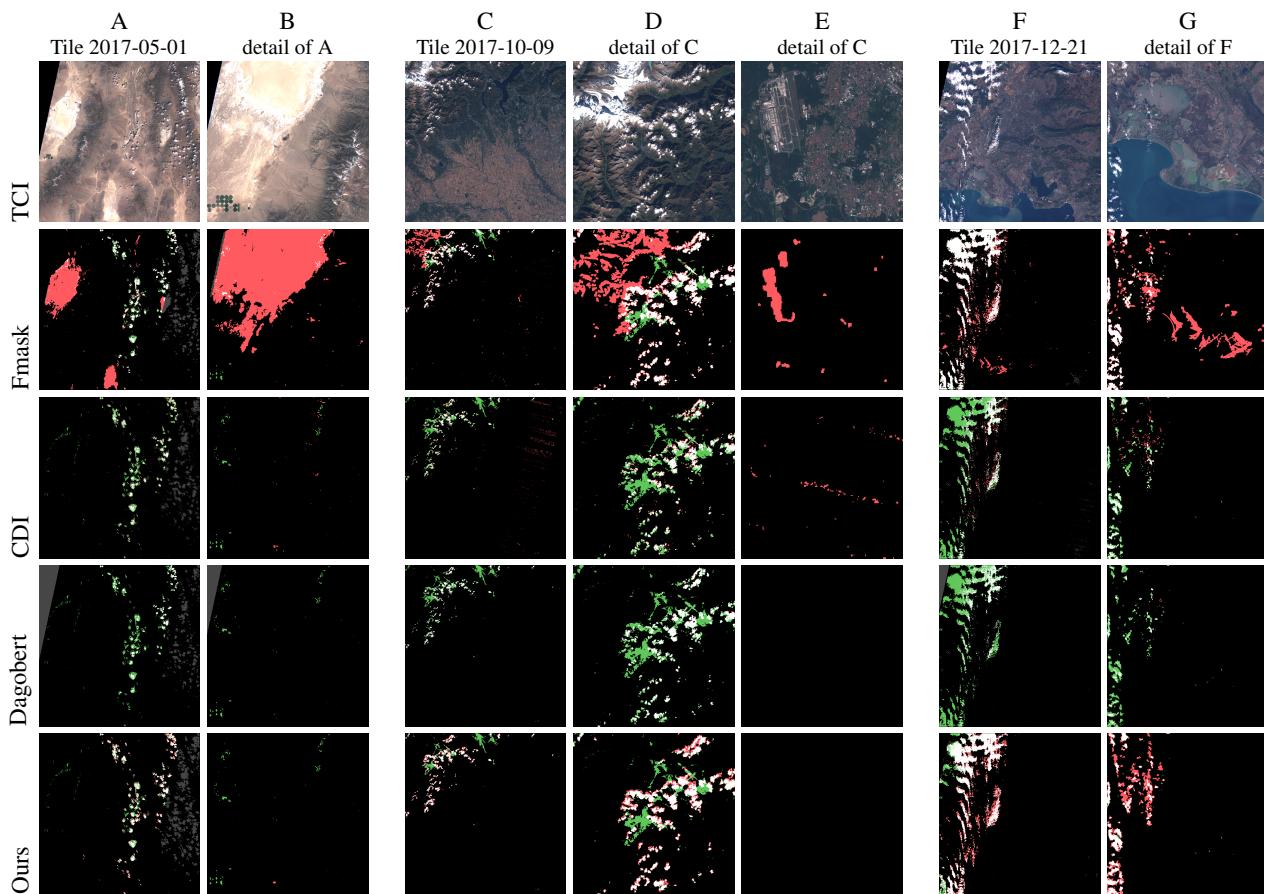


Figure 4. From top to bottom: True Color preview of the Sentinel-2 tile (TCI), Fmask, CDI, Dagobert, and our result. The true positives are displayed in white, and the true negatives in black. The false positives are in red and the false negatives in green. Gray indicates detections not accounted for in the evaluation (nodata class in the ground truth). Details are displayed next to each of the three tiles.

the metrics obtained with the entire dataset in Table 2. The definition Baetens *et al.* use to differentiate high from low clouds is not equivalent to opaque and transparent clouds. Indeed, clouds are classified high if they can be seen in band 10. This is true for transparent clouds but also for a large part of opaque clouds. Since the ground truth classification is given at 60 m/px, we downsampled all detection masks at this resolution. We used average downsampling and a binarization threshold at 0.5. Furthermore, a margin of a few pixels from the image borders was masked in the evaluation. This is because the proposed method cannot make detections on the borders as the optical flow needs at least the size of half window (W) and of the maximal displacement (D) to operate ($W + D$ pixels on each side, which become 5 pixels in the 60 m/px masks with the parameters used).

According to Table 1, our method does not perform as well as Fmask. This is expected, as Fmask relies on many criteria (including the cirrus band specifically designed for cloud detection). Nonetheless, the algorithm described in this paper achieved a reasonably high balanced accuracy. It is notable that the recall of our method is significantly higher than the recall of the CDI. This suggests that the index proposed in (Frantz *et al.*, 2018) does not fully exploits the parallax effect.

We present three tiles (and some detailed views) and the associated cloud masks in Figure 4. Each tile is 10980 × 10980 pixels at a resolution of 10 m/px. Columns A, C and F show full tiles from dates 2017-05-01, 2017-08-15 and 2017-12-21, respectively. In the first tile (column A), Fmask detected bright sand in

many places (detail on column B). Neither CDI nor the proposed parallax cloud detector produced these false positives. In the second tile (column C), Fmask detected snow on mountains (see detail on column D), as well as many small and bright regions, including an airport (see detail on column E). The CDI does not have false positives on the snow, but has false positives in surprising stripe patterns on the right of the image. Our method does not have false positives on the snow nor on bright areas. However, we miss the transparent clouds (see detail on column D). In the third tile (column F), Fmask classified some coastal regions as clouds (see detail on column G). The other parallax-based methods did not produce these false positives. Compared to Dagobert detector, our method has fewer false negatives.

Although our algorithm does not struggle with bright regions on the ground (*e.g.* snow, sand), we noticed some mis-classification of opaque clouds at very low altitude; in particular, clouds in valleys of mountainous areas. These clouds had almost no parallax. Also, Sentinel-2 images are composed from several detectors, whose orientation alternate; this yields parallax with opposite directions, creating some errors near the frontiers. Another limitation is that transparent clouds are not detected because the optical flow step correlates the ground and not the clouds.

The proposed method can also be applied to Sentinel-2 bands B02, B8A and B12, which are the ones with the largest acquisition delay (excluding bands at 60 m/px). The results are comparable to those obtained with bands B07, B08 and B8A. Cloud detection on Landsat-8 and 9 is also possible, since they use push-broom sensors. Preliminary experiments showed good

detection results. Likewise, good results were obtained with images from the PlanetScope constellation. Proper exploration and evaluation of our method on these other data sources is left for future work.

6. CONCLUSION

An algorithm for cloud detection by inter-band parallax was described. The method computes inter-band optical flow by correlation of the gradient direction on a window. This contrast-invariant procedure does not require any normalization for different band dynamics provided that there is no contrast inversion. Then, regions with coherent movement, as validated by a statistical test according to the a-contrario framework, are detected as clouds. The algorithm presented can be applied to images of any push-broom sensor, with any number of band pairs. The results are better than other parallax based methods and only slightly inferior to other state of the art methods. This suggests that the proposed approach could improve on the state of the art when using complementary information.

ACKNOWLEDGEMENTS

Work partly financed by Office of Naval research grant N00014-20-S-B001, DGA Astrid project “Filmer la Terre” no. ANR-17-ASTR-0013-01 and Kayrros.

REFERENCES

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. *Remote Sensing*, 11(4).

Cao, F., Lisani, J. L., Morel, J. M., Musé, P., Sur, F., 2008. *A Theory of Shape Identification*. Springer.

Chandran, G., Joji, C., 2015. A Survey of Cloud Detection Techniques For Satellite Images. *International Research Journal of Engineering and Technology (IRJET)*, 2(9), 2485–2490.

Dagobert, T., Grompone von Gioi, R., de Franchis, C., Morel, J.-M., Hessel, C., 2020a. Cloud Detection by Luminance and Inter-band Parallax Analysis for Pushbroom Satellite Imagers. *Image Processing On Line*, 10, 167–190.

Dagobert, T., Grompone von Gioi, R., de Franchis, C., Morel, J.-M., Hessel, C., 2020b. Cloud Detection by Luminance and Inter-band Parallax Analysis for Pushbroom Satellite Imagers. *Image Processing On Line*, 10, 167–190.

Dagobert, T., Grompone von Gioi, R., Morel, J.-M., de Franchis, C., 2020c. Temporal Repetition Detector for Time Series of Spectrally Limited Satellite Imagers. *Image Processing On Line*, 10, 62–77.

Dagobert, T., Monzón, N., Sánchez, J., 2019a. Comparison of Optical Flow Methods under Stereomatching with Short Baselines. *Image Processing On Line*, 9, 329–359.

Dagobert, T., Morel, J. M., de Franchis, C., Grompone von Gioi, R., 2019b. Visibility Detection in Time Series of PlanetScope Images. *IEEE International Geoscience and Remote Sensing Symposium*, 1673–1676.

Desolneux, A., Moisan, L., Morel, J.-M., 2000. Meaningful Alignments. *International Journal of Computer Vision*, 40(1), 7–23.

Desolneux, A., Moisan, L., Morel, J.-M., 2008. *From Gestalt theory to image analysis: a probabilistic approach*. 34, Springer.

Devernay, F., 1995. A Non-Maxima Suppression Method for Edge Detection with Sub-Pixel Accuracy. *INRIA Research Report*, 2724.

Frantz, D., Haß, E., Uhl, A., Stoffels, J., Hill, J., 2018. Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sensing of Environment*, 215, 471 - 481.

Frantz, D., Roder, A., Udelhoven, T., Schmidt, M., 2015. Enhancing the Detectability of Clouds and Their Shadows in Multitemporal Dryland Landsat Imagery: Extending Fmask. *IEEE Geoscience and Remote Sensing Letters*, 12(6), 1242–1246.

Gamonal, F. P., Ballester, C., Haro, G., Meinhardt-Holzapfel, E., Palomares, R. P., 2019. An Analysis and Speedup of the FALDOI Method for Optical Flow Estimation. *Image Processing On Line*, 9, 94–123.

Garamendi, J. F., Lazcano, V., Ballester, C., 2019. Joint TV-L1 Optical Flow and Occlusion Estimation. *Image Processing On Line*, 9, 432–452.

Gardner, M., 1960. More about the shapes that can be made with complex dominoes (Mathematical Games). *Scientific American*, 203(5), 186–201.

Garrido, L., Kalmoun, E. M., 2015. A Line Search Multilevel Truncated Newton Algorithm for Computing the Optical Flow. *Image Processing On Line*, 5, 124–138.

Golomb, S. W., 1994. *Polyominoes*. 2. ed. edn, Princeton University Press.

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.

Gordon, A., Glazko, G., Qiu, X., Yakovlev, A., 2007. Control of the mean number of false discoveries, Bonferroni and stability of multiple testing. *The Annals of Applied Statistics*, 1(1), 179–190.

Grompone von Gioi, R., 2014. *A contrario line segment detection*. Springer.

Grompone von Gioi, R., Hessel, C., Dagobert, T., Morel, J.-M., de Franchis, C., 2020. Temporal repetition detection for ground visibility assessment. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020, 829–835.

Grompone von Gioi, R., Pătrăucean, V., 2015. A contrario patch matching, with an application to keypoint matches validation. *2015 IEEE International Conference on Image Processing (ICIP)*, 946–950.

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.

Hollstein, A., Segl, K., Guanter, L., Brell, M., Enesco, M., 2016. Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. *Remote Sensing*, 8(8).

Horn, B. K., Schunck, B. G., 1981. Determining optical flow. *Artificial Intelligence*, 17(1), 185–203.

Irish, R. R., 2000. Landsat 7 automatic cloud cover assessment. *Proceedings of SPIE*, 4049, 348–355.

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 & Remote Sensing*, 72(10), 1179–1188.

Jara-Wilde, J., Cerdá, M., Delpiano, J., Härtel, S., 2015. An Implementation of Combined Local-Global Optical Flow. *Image Processing On Line*, 5, 139–158.

Jensen, I., Guttman, A. J., 2000. Statistics of lattice animals (polyominoes) and polygons. *Journal of Physics A: Mathematical and General*, 33(29), L257.

Lowe, D., 1985. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

Mahajan, S., Fataniya, B., 2019. Cloud detection methodologies: variants and development—a review. *Complex & Intelligent Systems*.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2. L. Bruzzone, F. Bovolo, J. A. Benediktsson (eds), *Image and Signal Processing for Remote Sensing XXIII*, SPIE, 3.

Manizade, K. F., Spinhirne, J. D., Lancaster, R. S., 2006. Stereo Cloud Heights From Multispectral IR Imagery via Region-of-Interest Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9), 2481–2491.

Meinhardt-Llopis, E., Sánchez Pérez, J., Kondermann, D., 2013. Horn-Schunck Optical Flow with a Multi-Scale Strategy. *Image Processing On Line*, 3, 151–172.

Monzón, N., Salgado, A., Sánchez, J., 2016. Robust Discontinuity Preserving Optical Flow Methods. *Image Processing On Line*, 6, 165–182.

Müller-Wilm, U., 2012. Sentinel-2 msi-level 2a products algorithm theoretical basis document. Technical Report ESA Report 2012, ref S2PAD-ATBD-0001 Issue 2.0, European Space Agency, Paris, France.

Panem, C., Baillarin, S., Latry, C., Vadon, H., Dejean, P., 2005. Automatic cloud detection on high resolution images. *IEEE International Geoscience and Remote Sensing Symposium*, 1.

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 Sensing of Environment*, 199, 107–119.

Qiu, S., Zhu, Z., He, B., 2019. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment*, 231(May), 111205.

Rabin, J., Delon, J., Gousseau, Y., 2009. A Statistical Approach to the Matching of Local Features. *SIAM Journal on Imaging Sciences*, 2(3), 931–958.

Rodriguez, M., Anger, J., De Franchis, C., Hessel, C., Facciolo, G., Grompone, R., Morel, J.-M., 2021. A CNN Cloud Detector for Panchromatic Satellite Images. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 4964–4967.

Sánchez Pérez, J., Meinhardt-Llopis, E., Facciolo, G., 2013. TV-L1 Optical Flow Estimation. *Image Processing On Line*, 3, 137–150.

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.

Serra, J., 1982. *Image Analysis and Mathematical Morphology*. Academic Press.

Shin, D., Pollard, J. K., 1996. Cloud height determination from satellite stereo images. *IEE Colloquium on Image Processing for Remote Sensing*, 4/1–4/7.

Sinclair, K., van Diedenhoven, B., Cairns, B., Yorks, J., Wasilewski, A., McGill, M., 2017. Remote Sensing of Multiple Cloud Layer Heights using Multi-Angular Measurements. *Atmospheric Measurement Techniques Discussions*, 1–23.

Sokolova, M., Japkowicz, N., Szpakowicz, S., 2006. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. A. Sattar, B.-h. Kang (eds), *AI 2006: Advances in Artificial Intelligence*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1015–1021.

Taravat, A., Proud, S., Peronaci, S., Del Frate, F., Oppelt, N., 2015. Multilayer Perceptron Neural Networks Model for Meteosat Second Generation SEVIRI Daytime Cloud Masking. *Remote Sensing*, 7(2), 1529–1539.

Witkin, A. P., Tenenbaum, J. M., 1983. On the role of structure in vision. J. Beck, B. Hope, A. Rosenfeld (eds), *Human and Machine Vision*, Academic Press, 481–543.

Wu, T., Hu, X., Zhang, Y., Zhang, L., Tao, P., Lu, L., 2016. Automatic cloud detection for high resolution satellite stereo images and its application in terrain extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 121, 143 – 156.

Zhang, W. D., He, M. X., Mak, M. W., 2001. Cloud detection using probabilistic neural networks. *IEEE 2001 International Geoscience and Remote Sensing Symposium*, 5, 2373–2375.

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 Sensing of Environment*, 159, 269–277.

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

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 Sensing of Environment*, 152, 217–234.