Cloud Detection over Snow and Ice with Oxygen A- and B-band Observations from the Earth Polychromatic Imaging Camera (EPIC)

Yaping Zhou[1,2], Yuekui Yang[1], Meng Gao[1,3], Peng-Wang Zhai[4]

[1]NASA Goddard Space Flight Center, Greenbelt, MD.

[2]JCET/University of Maryland Baltimore County, Baltimore, MD

[3]SSAI/NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Greenbelt, MD, USA

[4]JCET/Department of Physics, University of Maryland Baltimore County, Baltimore, MD, 21250, USA

Correspondence to: Yaping Zhou (yaping.zhou-1@nasa.gov)

Abstract

Satellite cloud detection over snow and ice has been difficult for passive remote sensing instruments due to the lack of contrast between clouds and the bright and cold surfaces; cloud mask algorithms often heavily rely on shortwave IR channels over such surfaces. The Earth Polychromatic Imaging Camera (EPIC) onboard the Deep Space Climate Observatory (DSCOVR) does not have infrared channels, which makes cloud detection over snow/ice even more challenging. This study investigates the methodology of applying EPIC's two oxygen absorption band pair ratios in A-band (764 nm, 780 nm) and B-band (688 nm, 680 nm) for cloud detection over the snow and ice surfaces. An elevation and zenith angle-dependent threshold scheme has been developed based on radiative transfer model simulations. The new scheme achieves significant improvement over the existing algorithm that imposes fixed thresholds for the A-band and B-band ratios. The positive detection rate nearly doubled from around 36% to 70% while the false detection rate dropped from 50% to 15% in January 2016 and 2017. The improvement during the summer months is less significant due to relatively better performance in the current algorithm. The new algorithm is applicable for all snow and ice surfaces including Antarctic, sea ice, high-latitude snow, and high-altitude glacier regions. This method is less

reliable when clouds are optically thin or below 2.5 km because the sensitivity is low in oxygen band ratios for these cases.

1. Introduction

The Earth Polychromatic Imaging Camera (EPIC) onboard the Deep Space Climate Observatory (DSCOVR) was launched in 2015 after a long haul in the late 1990s as it was initially designed for the Triana mission. The unique orbit of Triana, which was renamed DSCOVR, allows the instrument to take continuous measurements of the entire sunlit face of the

10 Earth from backscattering direction (scattering angles between 168.5° and 175.5°) from the first Lagrangian (L1) point of the Earth-Sun orbit, approximately 1.5 million km away. The EPIC instrument has 10 narrow spectral channels in the UV and Vis/NIR (317-780 nm) spectral range that enable retrieval of atmospheric ozone, cloud, and surface vegetation information. The focal plane of the EPIC system is a 2048×2048 pixel charge-coupled device

15 (CCD) array that covers the entire disk with a nadir resolution of 8 km. However, due to limited transmission capacity, all channels except the 443 nm channel are reduced to 1024 x 1024 arrays through onboard processing and interpolated back to full resolution after being downlinked. The operation of instrument and the downlink speed limit the temporal frequency of measurements to be approximately once every 1.5 hour in winter and 2.5 hour in summer. Detailed descriptions of

20 the EPIC instrument can be found in Herman et al. (2018), Marshak et al. (2018), and Yang et al. (2019).

The EPIC cloud product, including cloud mask (CM), cloud effective pressure (CEP), cloud effective height (CEH), and cloud optical thickness (COT), are developed with fewer spectral

25 channels available compared with many spectroradiometers currently onboard the polar and geostationary satellites (Yang et al., 2019). For example, the Moderate-resolution Imaging Spectroradiometers (MODIS) cloud algorithm uses simultaneous two-channel retrievals of COT and cloud effective radius (CER) with cloud phase determined by a series of spectral tests. Since EPIC does not have a particle size-sensitive channel, and has limited capability to

30 determine the cloud phase, the EPIC COT retrieval uses a single channel and derives two sets of COT, one for assumed ice phase and one for assumed liquid phase, each with fixed CER (Yang

et al., 2019; Meyer et al., 2016). CEP is derived based on two oxygen (O_2) band pairs, each consisting of an absorption and a reference channel (A-band: 764 nm and 780 nm; and B-band: 688 nm and 680 nm), respectively. The O_2 absorption bands are sensitive to cloud height because the presence of clouds, especially thick clouds, reduces the absorbing air mass that light

5   travels through, and the ratio of the absorbing and reference bidirectional reflectance functions (BRF) becomes larger. Since O_2 absorption at 764 nm is stronger than 688 nm, the A-band ratio has higher sensitivity than the B-band ratio (Yang et al., 2013).

Cloud detection algorithms usually use the contrast between clouds and the underlying earth

10   surface. Clouds are generally higher in reflectance and lower in temperature than the surface, which makes simple threshold approaches in the visible and infrared window channels effective in cloud detection (e.g., Saunders and Kriebel, 1988; Rossow and Garder, 1993; Yang et al., 2007; Ackerman et al., 2010). However, there are many situations when simple visible and infrared threshold tests are not able to separate cloud from surface or from atmospheric aerosols

15   such as heavy dust or smoke. The contrast between cloud and surface is usually weak in visible channels when the surface is bright, and in IR channels when surface temperature is very cold or cloud is very low. Additionally, partially cloudy pixels due to small-scale cumulus or cloud edge also increase difficulty of detection. The official MODIS CM algorithm uses more than 20 spectral channels to detect clouds in various situations. In particular, it heavily relies on

20   shortwave infrared channels in 1.38, 1.6, 2.1μm and thermal channels in 11 and 13.6 μm for cloud detection over snow and ice (Frey et al., 2008; Ackerman et al., 2010)

The lack of infrared and near-infrared channels in EPIC makes cloud detection very challenging, especially over snow and ice surfaces. The current EPIC CM algorithm adopts a

25   general threshold method, which uses two sets of spectral tests for each of the three scene types: ocean, land, and ice/snow (Yang et al., 2019). Over ocean, the 680 nm and 780 nm channels are used for cloud detection, because clouds and sea surface contrast well in both channels. Over land, because of large variations in surface reflectivity in the 680 nm and 780 nm, these two channels can no longer be used alone for cloud detection. Instead, the algorithm

30   uses the 388 nm channel and the A-band reflectivity ratio, i.e., R_{764}/R_{780} for cloud detection. The 388 nm channel is used because of its low reflectivity over land surface. The A-band ratio is used

based on the same mechanism as the cloud height retrieval because clouds reduce O_2 band absorption by increasing the height of effective reflective layer. The A-band ratio of a cloudy pixel is thus expected to be higher than that of a clear pixel in the otherwise same situation. The A-band ratio is selected for use over land surface because it has higher sensitivity than the B-band ratio. Over snow- and ice-covered regions, the O_2 A- and B-band ratios are used for cloud detection since the contrast between surface and clouds is small in the visible and UV channels.

Preliminary evaluation of the EPIC cloud products with collocated cloud retrievals from a composite cloud product (Khlopenkov et al., 2017) from geosynchronous earth orbit (GEO) and low earth orbit (LEO) satellites shows that the EPIC cloud algorithms are performing reasonably well and are consistent with theoretical expectations. The EPIC CM has an overall 80.2% accuracy rate and 85.7% correct cloud detection rate, but large discrepancy is found over the snow/ice regions, where the EPIC algorithm significantly underestimates cloud fraction, especially over ice and snow-covered Antarctic (Yang et al., 2019).

The current work aims to provide a better understanding of the variability of the O_2 band ratios under various clear and cloudy conditions over snow/ice surface to improve the CM performance. Radiative transfer model simulations and observed reflectance will be examined to derive dynamic thresholds for the O_2 band ratios so that the new algorithm is applicable to all snow/ice surface, i.e., Antarctic, Greenland, snow in high latitude and glaciers over high mountains. Section 2 provides an analytical discussion on the relationship between the O_2 band ratios with the relative airmass and surface elevation. Section 3 conducts sensitivity studies through radiative transfer modeling, and derives an algorithm from the model simulations. Section 4 describes the new cloud mask algorithm for the EPIC instrument over snow and ice. Section 5 reports on the new algorithm validation. Finally, Section 6 provides a brief summary and discussion.

2. An analytical guide with monochromatic radiative transfer

Oxygen absorption has been applied to remote sensing of cloud and aerosol extensively (e.g., Grechko, et al. 1973; Fischer, J. and Grassl, 1991; Min et al. 2004; Stammes et al., 2008; Wang

et al., 2008; Vasilkov et al. 2008; Ferlay et al., 2010; Yang et al. 2013; Ding et al. 2016; Richardson et al, 2019). The underlining physics is based on well-known and well-mixed atmospheric O_2 gaseous absorption, therefore, changes in observed radiance in the expected O_2 band contains information on how clouds or atmospheric aerosols interrupt the normal

5 absorption photon path and/or provide additional scattering at different vertical levels. The cloud detection using the O_2 absorption band ratios is based on the fact that clouds decrease the photon path within the atmosphere. Hence, clouds reduce the oxygen absorption optical thickness while their impact on the nearby reference channels is negligible. As a result, everything being equal, the BRF ratios between the absorption and the reference channels are expected to be larger for

10 cloudy sky than clear sky. In reality, photon paths can be very complicated: Yang et al. (2013) listed six pathways for a photon to reach the sensor. To simplify the discussion, we only focus on completely clear or cloudy cases. To determine a threshold for separating clear sky and cloudy sky, the first step is to understand factors that affect the clear O_2 band ratios so that clear sky O_2 band ratios can be well predicted. The second step is to understand how O_2 band ratios change

15 with the presence of different kinds of clouds. This step helps determine where thresholds can be drawn between clear sky and cloudy sky and what kind of sensitivity or uncertainty can be expected with this method.

The radiances entering the sensor consist of many components including the directly
20 reflected sun light by clouds, aerosol, surface and Rayleigh scattering through single- and multiple-scattering processes. Rayleigh optical thickness at the Oxygen A- and B-band regions are about 0.02 and 0.04, respectively. Hence, for clear sky over a bright surface, we can neglect the contribution of single and multiple scattering. Thus, the monochromatic BRF at the top of atmosphere can be related to the column optical depth via Beer's Law as:

$$R_{abs} = T_{abs}^{dn} * \alpha_{abs} * T_{abs}^{up} = \alpha_{abs} e^{-\left(\tau(z)+\tau_{ray}(z)\right)\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)} \qquad (1)$$

$$R_{ref} = T_{ref}^{dn} * \alpha_{ref} * T_{ref}^{up} = \alpha_{ref} e^{-\tau_{ray}(z)\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)} \qquad (2)$$

30 $$m = \frac{1}{\mu} + \frac{1}{\mu_0} = \frac{1}{\cos\theta} + \frac{1}{\cos\theta_0} \qquad (3)$$

where R_{abs} and R_{ref} are the BRF for the oxygen band and its reference band, respectively. BRF at the top of atmosphere is a product of downward transmittance (T^{dn}), spectral surface reflection albedo α, and upward transmittance (T^{up}). τ and τ_{ray} are optical thickness values due

5 to O₂ absorption and Rayleigh scattering at nadir, respectively, and are functions of surface elevation Z. m is the total airmass accounting for the slant path for both incoming (T^{dn}) and reflected light (T^{up}). The absorption channels are subject to both absorption and Rayleigh scattering while the reference channels only incur Rayleigh scattering. The ratio of R_{abs} and R_{ref} led to cancellation of Rayleigh scattering and surface albedo since the two channels are very

10 close, such that

$$\frac{R_{abs}}{R_{ref}} = e^{-\tau(z)\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)} = e^{-\tau(z)*m} \qquad (4)$$

The absorption optical thickness at a given location decreases exponentially with surface elevation following the approximate relationship in Eq. (5) (Petty, 2006):

$$\tau(z) = K_a w_1 \rho_0 H \exp\left(-\frac{Z}{H}\right) = c * \exp\left(-\frac{Z}{H}\right) \qquad (5)$$

Here H is the scale height, and K_a, w_1, ρ_0 are the mass absorption coefficient, mixing ratio of oxygen, and density of air at sea level, respectively. To relate the O₂ band ratios directly to

20 surface elevation and zenith angles in two separate terms, we take a double logarithm on both sides of Eq. (4), and substitute τ with Eq. (5), which leads to

$$\ln\left(\frac{R_{abs}}{R_{ref}}\right) = c * \exp\left(-\frac{Z}{H}\right) * m \qquad (6)$$

$$Define \; dln\left(\frac{R_{abs}}{R_{ref}}\right) = \ln\left\{-\ln\left(\frac{R_{abs}}{R_{ref}}\right)\right\} \qquad (7)$$

$$dln\left(\frac{R_{abs}}{R_{ref}}\right) = c_0 - \frac{Z}{H} + \ln m \qquad (8)$$

Here *dln* refers to double logarithm, and the minus sign before the second logarithm function is added to avoid negative value. Eq. (8) decouples the effect of elevation and zenith angles in $dln(\frac{R_{abs}}{R_{ref}})$, which allows estimation of coefficients in Eq. (8) with simple linear regression using two independent terms Z and $\ln m$:

$$dln \approx c_0 + c_1 \ln m + c_2 Z \qquad (9)$$

Once $dln(\frac{R_{abs}}{R_{ref}})$ is solved, the O_2 band ratios can be derived with Eq. (10):

$$\frac{R_{abs}}{R_{ref}} = e^{-e^{dln}} \qquad (10)$$

The above derivation shows that the clear sky O_2 band ratios can be analytically predicted using surface elevation and zenith angles. Of course, many approximations have been used such as cancellation of Rayleigh extinction and surface BRDF for the pair channels and constant absorption scale height. Due to large surface albedo, contributions of Rayleigh scattering are also
15 neglected. The contribution of Rayleigh scattering in the reflectance is about 0.01-0.02, and this may cause an uncertainty of 1% to 2% in the band ratio for bright surface. In cases of dark surfaces such as over ocean, the surface albedo is so small (~0.05) that the Rayleigh scattering starts to dominate the observed reflectance, and the simple equations derived here will result in large bias. However, with albedos relatively large (around 0.8), our sensitivity studies find the
20 ratios relatively stable, even though the single channel reflectances change in proportion to the surface albedo. The coefficients in Eq. (9) can be derived from either radiative transfer model simulations or real observational data from EPIC. The advantage of the former is the exact knowledge of model's atmosphere and clear or cloudy conditions. Conversely, its disadvantage is a limited number of atmospheric profiles and sometimes simplistic or even unrealistic cloud
25 input to the model. The advantage of using observational data is the abundant radiance measurements that could be used as training dataset while the disadvantage is the limited knowledge on atmospheric profiles and uncertainties in clear pixel identification. A common practice for developing a cloud mask algorithm is to use retrievals of simultaneous measurements from other better-equipped instruments or ground observations as the truth. Exact
30 same-time overpass is quite rare even with the vast data volume from the polar orbiting satellites

such as Terra and Aqua, and cloud detection over snow and ice from instruments such as MODIS is itself subject to large uncertainty. This could lead to some false cloud/clear identification in the training dataset and bias the results. Based on the above reasoning, we first derive the O_2 band ratio thresholds with both model simulations and observations, and then

5   determine which set of coefficients to use for the EPIC cloud mask algorithm.

3. Radiative transfer simulations

We used a radiative transfer simulator for EPIC (Gao et al., 2019) to generate the A-band and

10   B-band reflectances over snow and ice surface. The EPIC simulator is built upon a radiative transfer model (Zhai et al., 2009, 2010) that solves multiple scattering of monochromatic light in the atmosphere and surface system. Gas absorptions due to ozone, oxygen, water vapor, nitrogen dioxide, methane, and carbon dioxide are incorporated in all EPIC bands. In the O_2 A- and B-bands, radiances from line-by-line simulations are convoluted with the instrument

15   response function at the top of the atmosphere. The model atmosphere assumes a one-layer cloud with a molecular layer both above and beneath.

For this particular study, the cloud layer has varied optical thickness ranging from 0.2 to 30 and cloud top height from 2.5 km to 15 km. The cloud geometric height is 1 km. The cloud

20   droplets assume a gamma size distribution with an effective radius of 10 µm and an effective variance of 0.1. Four atmospheric vertical profiles from 1976 US standard atmosphere, mid-latitude winter, subarctic summer and subarctic winter atmospheres are used. Surface albedo is set at 0.8 to represent over snow or ice surface. The model simulates a variety of cases with 17 solar zenith angles ranging from 0° to 80°, 18 view zenith angles from 0° to 85°, and 37 azimuth

25   angles from 0° to 180°, all with an increment of 5°. In addition to the varying sun-sensor geometry, the reflecting surface elevation is set from 0 to 15 km with a 2.5 km increment for the clear sky sensitivity tests while the cloudy sky simulations are performed for the whole atmospheric column.

We first examine whether the radiative transfer model simulations duplicate the quantitative relationship between the O_2 band ratios and surface elevation and total airmass as discussed above (Eq. 9).

A direct inspection of O_2 band ratios at fixed view zenith angle and relative azimuth angles

5   with surface elevation indicates a nearly linear relationship between the two (Fig. 1a, 1b). Higher elevation means shorter photon path length and larger O_2 band ratios. However, the relationship is not strictly linear; the largest rate of change appears at 12.5 km for the A-band and at 7.5 km for the B-band. The relationship also depends on the solar zenith angle. At higher solar zenith angle, not only the ratios are lower at all surface elevations but also the change with height ($\frac{\partial r}{\partial Z}$)

10   is larger. However, the same relationship can be expressed as a quasi-linear relationship between Z and double logarithm of O_2 band ratios at fixed zenith angles as indicated by Eq. (9) (Fig. 1c, 1d).

The variation of O_2 band ratios with sun and viewing geometry has been discussed in Yang et

15   al. (2013) and Gao et al. (2019). Here we show a more quantitative dependence of O_2 band ratios as a function of the total relative airmass (m) defined in Eq. (3) at fixed surface elevation (sea level in this case, Fig. 1e, 1f). The inverse relationship of O_2 band ratios with m is evident. Although EPIC is positioned close to the backscattering direction, there is a small difference in θ_s and θ_v, generally smaller than 6°. The red dots show the simulations when the difference

20   between θ_s and θ_v is smaller than 6° to mimic the EPIC sun-view geometry. The relationship derived from samples with restricted view zenith angles is not much different from that of all samples. Figures 1g-h further project this relationship as logarithm of m versus double logarithm of O_2 band ratios as shown in Eq. (9). We notice that the linear relationship holds very well except for very large relative airmass (ln (m) > 2.5, which corresponds to zenith angles > 80°).

To account for both elevation and zenith angle effect, we use Z and ln (m) as two independent terms to fit the $dln(\frac{R_{abs}}{R_{ref}})$ for the simulations with view zenith angle difference less than 6° as suggested in Eq. (9). The results indicate high confidence of the fitting with multi-correlation coefficients reaching 0.998 for both A-band and B-band (Fig. 1i, 1j). The coefficients

30   c_0, c_1, and c_2 are listed in Table 1. Table 1 also lists the set of coefficients derived from

observations utilizing information from collocated GEO/LEO pixels. Details will be discussed in the next section.

Table 1. Regression coefficients to equation (9) using model simulation data and observations.

	A-band			B-band		
	c0	c1	c2	c0	c1	c2
Simulations	-0.2706	-0.1471	0.5180	-0.9589	-0.1373	0.4328
Observations	-0.1764	-0.1152	0.4542	-0.8672	-0.1185	0.3995

The coefficients in Table 1 can be applied to Eq. (9) to compute an expected clear sky band ratio. In order to test the feasibility of using the derived clear sky band ratios as the thresholds for clear and cloudy pixel separation, we first evaluate the sensitivity of O_2 band ratios to cloud properties. This is done by adding clouds of different optical thickness and cloud top height in

10   the radiative transfer simulations, and then comparing the O_2 band ratios of cloudy sky with those of clear sky under the same sun-view geometry. The results for solar and view zenith angles of 30° and 60° and relative azimuth angle of 160° are shown in Figure 2, with the corresponding clear sky values shown as the filled and open triangles, respectively. We notice that the O_2 band ratios generally increase with the optical thickness and are higher for cloudy sky

15   than the clear sky but with certain exceptions. At low zenith angles (< 30°), Figure 2 shows very low sensitivity of O_2 band ratios on cloud optical depth when cloud top height is 2.5 km. Even at 5 km, the cloudy sky ratios are observably higher than clear sky value only when cloud optical thickness (COT) is greater than 3. Note that the figure shows that adding a layer of optically thin cloud (COT < 3) actually decreases the ratio for the 30° zenith angle case. The reason is that

20   under this circumstance the reflectance of reference channel increases more than the absorption channel, which indicates an increase in the photon path. The causes of photon path increase include multiple scattering inside the cloud and surface-cloud interaction. The strong surface-cloud interaction over the bright surface of snow/ice partly contributes to the low sensitivity of O_2 band ratios for the low and thin clouds compared with relatively darker surfaces. The

25   sensitivity of O_2 band ratios to cloud optical thickness and height increases with solar and view

zenith angles, as can be seen from the SZA = VZA = 60° curves. These results show that O_2 band ratios can be used to detect clouds that are thick and/or high with much confidence over snow/ice surfaces. Difficulties still exist in detecting thin clouds or low clouds at low zenith angles (<30°). Note that A-band has better sensitivity than B-band as expected.

4. EPIC cloud mask over snow/ice

As discussed in Section 2, we can derive the thresholds using either radiative transfer simulations or satellite observations. The previous section discussed the path of using modeling

10 results, here we attempt to derive the thresholds based on the real EPIC data using the same theoretical guide as provided in Section 2.

For this purpose, a collocated dataset from GEO/LEO composite cloud product and EPIC L1B data for January and July of 2017 are used. The GEO/LEO composite dataset was

15 generated by the Clouds and the Earth's Radiant Energy System (CERES) team at the NASA Langley Research Center by projecting the GEO/LEO retrievals to the EPIC grid at each EPIC observing time (Khlopenkov et al., 2017). This ensures that every EPIC image/pixel has a corresponding GEO/LEO composite image/pixel with approximately same size and observation time. The LEO satellites include NASA Terra and Aqua MODIS and NOAA AVHRR while

20 geosynchronous satellite imagers include Geostationary Operational Environmental Satellites (GOES) operated by NOAA, Meteosat satellites by EUMETSAT, and the Multifunctional Transport Satellites (MTSAT) and Himawari-8 satellites operated by Japan Meteorological Agency (JMA). The cloud retrievals in the composite data follows Minnis et al. (2011). The cloud fraction and surface scene types from the composite dataset are used to select the clear

25 pixels (100% clear) over snow/ice surface (when 90% of the major scene type is permanent snow or ice). To reduce the uncertainties, we further restrict the observations in the composite files to be within 5 minutes of the EPIC image. We also restrict the analysis on pixels with view zenith angle less than 80°. The surface elevation data is from the National Geophysical Data Center (NGDC) TerrainBase Global Digital Terrain Model (DTM), version 1.0 (Row and Hastings,

30 1994).

The same type of regression is performed for the clear sky pixels using the elevation and logarithm of total relative airmass as independent variables, and the double logarithm of the O_2 band ratios as the dependent variables as suggested by Eq. (9). The derived regression coefficients (Table 1) are quite close to those derived from the model simulations with slightly larger scatter (Fig. 3a, 3b). In addition, clear sky thresholds predicted from observational data have to be adjusted to provide a better overall performance since the regression model is designed to predict the median rather than the upbound of clear sky band ratios. The same regression coefficients applied to cloudy sky samples indicate many overlapping of O_2 band ratios from clear sky and cloudy sky pixels (Fig. 3c, 3d). A threshold value too high will guarantee the clear sky identification but underestimate cloudy pixels, and too low will lead to overestimation of cloudy pixels. To achieve the best overall clear sky and cloudy sky performance, we set the threshold value by increasing the ratios derived from Eq. (10) by 0.025 so that the cloud mask threshold is close to the upper quantile of the clear sky values (red dashed line in Fig. 3c and 3d).

Results show that using the set of coefficients derived from the model simulations captures most of the clear sky samples without being adjusted (Figures not shown). We found that even though the thresholds derived from the observational data performs slightly better when applied back to the same dataset, they underperform the model derived algorithm when applied to a different dataset. One likely reason is that the cloud identification in the observational training dataset has its own non-negligible uncertainties. These uncertainties will not affect the performance in the training dataset but affect the algorithm performance in an independent one. For this purpose, we adopt the algorithm derived from the model simulations for the rest of this paper.

Following the current EPIC cloud mask algorithm, we also set an upper and a lower threshold that is 0.02 above or below the model predicted threshold (RT_0). A cloud mask (CM) confidence level is determined for each pair of the O_2 band ratios based on whether the ratios fall between these intervals/thresholds:

$$CM = \begin{cases} 4 & \text{Ratio} > RT_0 + 0.02; & \text{CldHC} \\ 3 & RT_0 < \text{Ratio} < RT_0 + 0.02; & \text{CldLC} \\ 2 & RT_0 - 0.02 < \text{Ratio} < RT_0; & \text{ClrLC} \\ 1 & \text{Ratio} < RT_0 - 0.02; & \text{ClrHC} \end{cases}$$

Here, CldHC, CldLC, ClrHC, and ClrLC refer to Cloud with High Confidence, Cloud with low Confidence, Clear with High Confidence; Clear with low Confidence, respectively. The final

5   confidence level is determined by combing the two results from the A- and B-band tests according to Table 2. Note that we only define high confident cloud and clear when both tests show cloud and clear with high confidence.

Table 2. The logic table for combining the cloud mask results from the A- and B-band tests.

10   Acronyms CldHC: Cloud with High Confidence; CldLC: Cloud with low Confidence; ClrHC: Clear with High Confidence; ClrLC: Clear with low Confidence.

		A-band Test			
		CldHC	CldLC	ClrLC	ClrHC
B-band test	CldHC	**CldHC**	CldLC	CldLC	CldLC
	CldLC	CldLC	CldLC	CldLC	ClrLC
	ClrLC	CldLC	CldLC	ClrLC	ClrLC
	ClrHC	CldLC	ClrLC	ClrLC	**ClrHC**

An illustration of EPIC O_2 band ratios and the derived cloud mask over Antarctic on Dec.23, 2017 is shown in Figure 4, along with cloud fraction derived from GEO/LEO composite. In this

15   figure, the A-band and B-band ratios show not only the presence of clouds but also the effect of elevation, as the low values over Ross Ice Shelf are clearly influenced by the low elevation in that area. The new cloud mask detects the majority of the cloud area, but some portion of clouds over this region is missing. This indicates the derived thresholds have uncertainties for near sea level situations.

5.   Algorithm validation

Using the thresholds from radiative transfer simulations, we reprocessed the EPIC cloud
mask over snow/ice surface for all the collocated pixels in three months, January 2016, 2017 and
July 2017.

Because of EPIC's large pixel size, one EPIC pixel corresponds to many GEO/LEO pixels,
hence a subpixel cloud fraction is reported in the composite dataset for each of the EPIC pixels.
We divide the GEO/LEO cloud fraction into 4 categories to match with the CM in EPIC:

10 $\qquad$ GEO/LEO CM $= \begin{cases} 4: & \text{cloud fraction} \geq 95\% \\ 3: & 50\% \leq \text{cloud fraction} < 95\% \\ 2: & 5\% \leq \text{cloud fraction} < 50\% \\ 1: & \text{cloud fraction} < 5\% \end{cases}$

Figure 5 shows the distributions of the EPIC cloud mask values corresponding to each of
the GEO/LEO cloud fraction range for the three months. The figure also compares the
performance of the original EPIC algorithm (V01) and the newly developed algorithm. As can be
15 seen from the figure, the original algorithm overestimates the clear sky fraction (blue bars) for
observations in January 2016 and 2017, evidenced by the high blue bar values in both the
GEO/LEO low cloud fraction (<5%) and high cloud fractions (>95%) categories. There are also
considerable amount of pixels in the GEO/LEO low cloud fraction category being classified as
cloudy by the original algorithm (yellow and red bars). Improvement is evident for the new
20 algorithm, where most of the pixels with < 5% cloud fraction have CM = 1 or 2 (high and low
confidence clear, respectively), while pixels with >95% cloud fraction more likely have CM
values of 4 and 3 (high and low confidence cloudy, respectively).

To quantitatively measure the performance of the cloud masking algorithms, we further
define successful retrievals as: a) both algorithms identify the pixel as cloud (including high and
25 low confidence), b) both identify as clear (including high and low confidence), and unsuccessful
retrievals as: c) EPIC identified as clear while GEO/LEO identified as cloud, d) EPIC identified
as cloud and GEO/LEO identifies as clear. Assuming GEO/LEO is the "truth," a number of
parameters as a measure of EPIC's CM accuracy are computed:

$$Accuracy = \frac{a+b}{a+b+c+d} \qquad (11)$$

$$POCD = \frac{a}{a + c} \qquad (12)$$

$$POFD = \frac{d}{b + d} \qquad (13)$$

Here POCD and POFD are the probability of correct detection and probability of false
detection, respectively. For January 2016 and 2017, compared to the current product, the
accuracies have been improved considerably from a low 57-60% to around 83%. The POCD is
nearly doubled (from 36% to 70%) and a significant reduction of POFD (a drop from around
50% to 15%). The original algorithm performs relatively well in July 2017 with a probability of
correct detection (POCD) at 77.5% and a low probability of false detection (POFD) of 16.5%;
hence the improvement for this month is relatively small.

Figure 6 shows the cloud fraction on a 1° x 1° grid for January 2017 over snow/ice covered
Antarctica. Note that here we lift the 5 min time difference limitation and use all available pixels
with view zenith angles less than 75° from the GEO/LEO composites (Khlopenkov et al., 2017)
in order to have a full coverage of the region. The cloud fraction map from GEO/LEO shows a
belt of high cloud fraction originated from mid-latitude storm track reaching the edge of the
continent. Onto the icy plateau of East Antarctica, cloud fraction quickly decreases. High cloud
fraction is found over West Antarctica. The cloud fraction from the original algorithm shows
quite an opposite cloud distribution pattern between the West and East Antarctica. This is likely
due to fixed threshold that is too low for the high elevation in the East Antarctica and too high
for the low elevation in the West Antarctica. By taking the elevation into account, the new
algorithm identifies the regional cloud distribution much better. In addition, the new algorithm
also has a better cloud fraction match around the edge of the Antarctic continent.

To examine the performance of the new algorithm on the global scale, we plotted gridded
cloud fraction over snow/ice surfaces for the entire globe in January 2016 (Fig. 7). The number
of snow/ice pixels used for the map are also shown, because sample numbers affect the quality of
monthly mean. We notice that the number of snow/ice pixels per grid is much higher in January
over Antarctica. There are also considerable amounts of snow/ice pixels in northern hemisphere
high-latitude regions and the southern tip of Andes. There is no retrieval north of 50° N due to no

day light or view zenith angle too large in January (DSCOVR only has observations for the daytime Earth). Comparison show that the new algorithm improves cloud distributions noticeably.

5   Figure 8 shows a similar map but for July 2017. During the boreal summer, the cloud mask algorithm has retrievals over the entire northern hemisphere but not for part of Antarctica south of 65 °S due to the polar night. The GEO/LEO cloud fraction map indicates high cloud fraction (> 80%) over snow/ice surface over most of the regions in July except over Greenland. The original algorithm has similar cloud fraction in most areas over snow and ice surfaces, except

10   over southeast Greenland where it has significantly more cloud than the other part of the Greenland. This is likely due to the original algorithm's failure to take into consideration the high elevation there. On the other hand, the underestimation of cloud fraction at the southern tip of Andes could be due to its failure to take into account the large solar and view zenith angles in summer. The new algorithm detects significantly lower amount of cloud fraction in Greenland

15   and improves the cloud detection in the aforementioned high mountain areas.

         Even though the new cloud mask has achieved high accuracy and general pattern match with the GEO/LEO retrievals, regional differences between the two can still be quite large. This is partly due to the large uncertainty of cloud detection from GEO/LEO over snow/ice itself, and

20   partly due to the intrinsic difficulty of using O_2 band ratios in detecting the low cloud and thin cloud as discussed before. In addition, the time difference between EPIC and GEO/LEO observations can also impact the comparison between the two. Stratifying the performance based on difference in the observation time, we find a larger difference in the observing time leads to slightly lower POPD, higher POFD and an overall decreasing accuracy (Fig. 9).

6. Summary and discussion

         Due to limited spectral channels, especially the lack of infrared and near infrared channels in the DSCOVR EPIC instrument, cloud detection for EPIC over snow and ice poses a great

30   challenge. The existing EPIC cloud mask algorithm employs two oxygen pair ratios in A-band (764 nm, 780 nm) and B-band (688 nm, 680 nm) for cloud detection over the snow and ice

surfaces. The method is based on the fact that cloud blocks some or all of the photon path beneath it and reduces the optical thickness of the oxygen absorption so that the ratios are expected to be larger for cloudy sky than the clear sky for the same sun-view geometry; hence a threshold can be set to separate cloudy pixels from clear pixels. However, clear sky O_2 band

5   ratios depend on a number of factors such as surface elevation and sun/viewing geometry that impact the total absorption airmass; these factors need to be accounted for.

In this study, we use both the radiative transfer theory and model simulations to quantify the relationship between the O_2 band ratios with surface elevation and zenith angles. Thresholds are

10   derived as a function of surface elevation and sun-view geometry based on both model simulation results and observations. Model derived algorithm is chosen because of its stable performance. The new algorithm increases the accuracy of EPIC cloud mask over snow and ice surfaces in winter by more than 20%. This is achieved through a significant reduction of false detection rate from 50% to 15% and nearly doubling of the correct detection rate. The

15   improvement in the summer month July is mild, with the main improvement observed over Greenland. Of course, these performance matrices are based on comparison with GEO/LEO cloud mask which has quite large uncertainty over snow and ice surfaces itself. In addition to significant improvement in cloud detection over Antarctic, the new algorithm also improves cloud detection over Greenland and some mid-latitude high mountain areas.

Limitations of this method include difficulties in identifying thin cloud with optical thickness less than 3 or low cloud below 2.5 km due to the lack of sensitivity in O_2 band ratios under these circumstances. Compared with the infrared-based techniques, one advantage of this oxygen band technique is that it is relatively insensitive to the surface and atmosphere

25   temperature. Therefore, the method presented in this work provides a solution to polar cloud detection when infrared channels are not available. We anticipate that cloud detection using oxygen band technique to be of great value in the future missions.

Acknowledgments

This research was supported by the NASA DSCOVR Earth Science Algorithms program managed by Richard Eckman. The DSCOVR level-1 and level-2 data used in this paper are

5   publicly available from NASA Langley Atmospheric Sciences Data Center (ASDC).

References

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

Ackerman, S. and Frey, R.: MODIS/Aqua Cloud Mask and Spec- tral Test Results 5-Min L2 Swath 250m and 1km, Dataset, doi:10.5067/MODIS/MYD35_L2.006, 2016.

Buriez, J.-C., Vanbauce, C., Parol, F., Goloub, P., Herman, M., Bonnel, B., Fouquart, Y., Couvert, P., and Sèze, G.: Cloud detection and derivation of cloud properties from POLDER, Int. J. Remote Sens., 18, 2785–2813, 1997.

Davis,A.B., N. Ferlay, Q. Libois, A. Marshak, Y. Yang, Q. Min. Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: a physics-based approach. J Quant Spectrosc Radiat Transf, 220 (2018), pp. 84-96, 10.1016/j.jqsrt.2018.09.006

Davis, A.B., G. Merlin, C. Cornet, L.C. Labonnote, J. Rédi, N. Ferlay, *et al.*Cloud information content in EPIC/DSCOVR's oxygen A- and B-band channels: an optimal estimation approach. J Quant Spectrosc Radiat Transfer, 216 (2018), pp. 6-16, 10.1016/j.jqsrt.2018.05.007.

Ding, S., Wang, J., and Xu, X.: Polarimetric remote sensing in oxygen A and B bands: sensitivity study and information content analysis for vertical profile of aerosols, Atmos. Meas. Tech., 9, 2077–2092, https://doi.org/10.5194/amt-9-2077-2016, 2016.

Ferlay, N., Thieuleux, F., Cornet, C., and Davis, A. B.: Toward New Inferences about Cloud Structures from Multidirectional Measurements in the Oxygen A Band: Middle-of-Cloud Pressure and Cloud Geometrical Thickness from POLDER-3/PARASOL. *J. Appl. Meteor. Climatol.*, **49**, 2492–2507. doi: http://dx.doi.org/10.1175/2010JAMC2550.1, 2010.

Fischer, J. and Grassl, H.: Detection of cloud-top height from backscattered radiances within the oxygen A band. Part 1: Theo- retical study, J. Appl. Meteorol., 30, 1245–1259, 1991.

Frey, R.A., S.A. Ackerman, Y. Liu, K.I. Strabala, H. Zhang, J.R. Key, and X. Wang, 2008: Cloud Detection with MODIS. Part I: Improvements in the MODIS Cloud Mask for Collection 5. *J. Atmos. Oceanic Technol.,* **25**, 1057–1072, https://doi.org/10.1175/2008JTECHA1052.1

Gao, M., P.-W. Zhai, Y. Yang, and Y. Hu. 2019. "Cloud remote sensing with EPIC/DSCOVR observations: A sensitivity study with radiative transfer simulations." Journal of Quantitative Spectroscopy and Radiative Transfer, **230:** 56-60 [10.1016/j.jqsrt.2019.03.022]

5   Grechko, Y. I., Dianov-Klokov, V. I., and Malkov, I. P.: Aircraft measurements of photon paths in reflection and transmission of light by clouds in the 0.76 mm oxygen band, Atmos. Ocean Phys., 9, 262–265, 1973.

Herman, J.; Huang, L.; McPeters, R.; Ziemke, J.; Cede, A.; Blank, K. Synoptic ozone, cloud reflectivity, and erythemal irradiance from sunrise to sunset for the whole earth as viewed

10   by the DSCOVR spacecraft from the earth–sun Lagrange 1 orbit. Atmos. Meas. Technol. **2018**, 11, 177–194.

Koelemeijer, R.B. A, P. Stammes, J.W. Hovenier, J.F. de Haan. A fast method for retrieval of cloud parameters using oxygen A band measurements from the Global Ozone Monitoring Experiment. J Geophys Res, 106 (2001), pp. 3475-3490

15   Khlopenkov, K., Duda, D., Thieman, M., Minnis, P., Su, W., and Bedka, K.: Development of multi-sensor global cloud and ra- diance composites for earth radiation budget monitoring from DSCOVR, Proc. SPIE 10424, Remote Sens., 104240K, Warsaw, 2 October 2017, https://doi.org/10.1117/12.2278645, 2017.

Marshak, A., J. Herman, A. Szabo, et al. 2018. "Earth Observations from DSCOVR/EPIC

20   Instrument." Bulletin of the American Meteorological Society, BAMS-D-17-0223.1 [10.1175/bams-d-17-0223.1]

Min, Q. L., Harrison, L. C., Kierdron, P., Berndt, J., and Joseph, E.: A high-resolution oxygen A-band and water vapor band spectrometer, J. Geophys. Res., 109, D02202, doi:10.1029/2003JD003540, 2004.

25   Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., Spangenberg, D. A., Arduini, R. F., Trepte, Q. Z., Smith, W. L. Jr., Ayers, J. K., Gibson, S. C., Miller, W. F., Chakrapani, V., Takano, Y., Liou, K.-N.,Xie, Y. and Yang, P., "CERES Edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data, Part I: Algorithms," IEEE Trans. Geosci. Remote Sens., 49, 4374–4400, doi:

30   10.1109/TGRS.2011.2144601 (2011).

Petty, W. G. (2006). A first course in atmospheric radiation. Sundog
    Pub. ISBN 9780972903318. OCLC 932561283.

Richardson, M., Leinonen, J., Cronk, H. Q., McDuffie, J., Lebsock, M. D., and Stephens, G. L.:
    Marine liquid cloud geometric thickness retrieved from OCO-2's oxygen A-band
5   spectrometer, Atmos. Meas. Tech., 12, 1717–1737, https://doi.org/10.5194/amt-12-1717-
    2019, 2019.

Rossow, W. B., and L. C. Garder, 1993: Cloud detection using satellite measurements of infra-
    red and visible radiances for ISCCP. J. Climate, **6**, 2341-2369.

Row, L. W., III, and D. A. Hastings, 1994: TerrainBase worldwide digital terrain data, release
10   1.0 NOAA/National Geophysical Data Center, accessed 22 June
    2016, ftp://ftp.ngdc.noaa.gov/Solid_Earth/cdroms/TerrainBase_94/data/global/tbase/. Googl
    e Scholar

Saunders, R. W. and K. T. Kriebel, 1988: An improved method for detecting clear sky and
    cloudy radiances from AVHRR data. Int. J. Remote Sens., **9**, 123-150.

15   Stammes, P., Sneep, M., de Haan, J. F., Veefkind, J. P., Wang, P., and Levelt, P. F.: Effective
    cloud fractions from the Ozone Monitoring Instrument: Theoretical framework and
    validation, J. Geophys. Res., 113, D16S38, doi:10.1029/2007JD008820, 2008.

Vasilkov, A., Joiner, J., Spurr, R., Bhartia, P. K., Levelt, P., and Stephens, G.: Evaluation of the
    OMI cloud pressures derived from rotational Raman scattering by comparisons with other
20   satellite data and radiative transfer simulations, *J. Geophys. Res.*, 113, D15S19,
    doi:10.1029/2007JD008689, 2008.

Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendael, M.: FRESCO+: An
    improved O_2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals. *Atmos.*
    Chem. Phys., 8**,** 6565–6576, 2008.

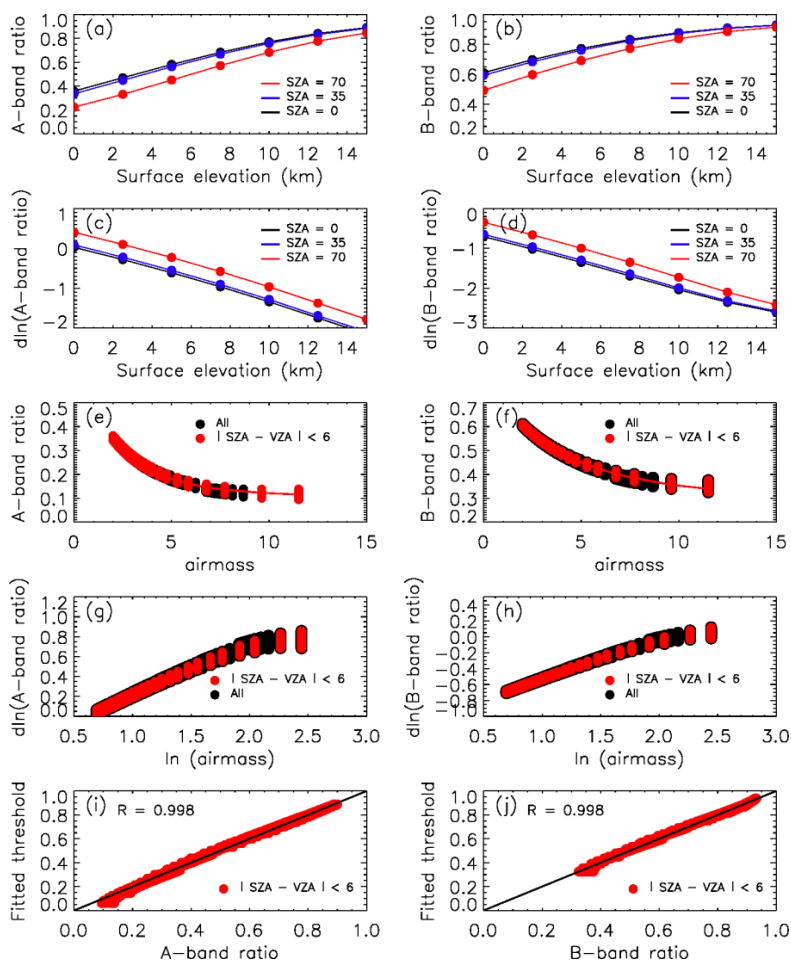
25   Yang, Y., Di Girolamo, L. and Mazzoni, D.: Selection of the automated thresholding algorithm
    for Multi-angle Imaging SpectroRadiometer Radiometric Camera-by-Camera Cloud Mask
    over land. *Remote Sens. Environ.* 107, 159-171, doi:10.1016/j.rse.2006.05.020, 2007.

Yang, Y., Marshak, A., Mao, J., Lyapustin, A., and Herman, J.: A Method of Retrieving Cloud
    Top Height and Cloud Geometrical Thickness with Oxygen A and B bands for the Deep

Space Climate Observatory (DSCOVR) Mission: Radiative Transfer Simulations, J. Quant. Spectrosc. Ra., 122, 141–149, https://doi.org/10.1016/j.jqsrt.2012.09.017, 2013.

Yang, Y., K. Meyer, G. Wind, et al. 2019. "Cloud products from the Earth Polychromatic Imaging Camera (EPIC): algorithms and initial evaluation." Atmospheric Measurement Techniques, **12 (3):** 2019-2031 [10.5194/amt-12-2019-2019]

Zhai, P., Y. Hu, C. R. Trepte, and P. L. Lucker, (2009). A vector radiative transfer model for coupled atmosphere and ocean systems based on successive order of scattering method, Optics Express **17**, 2057-2079.

Zhai, P., Y. Hu, J. Chowdhary, C. R. Trepte, P. L. Lucker, D. B. Josset, (2010) A vector radiative transfer model for coupled atmosphere and ocean systems with a rough interface, J Quant Spectrosc Radiat Transf, **111**, 1025-1040 (2010).

5   Figure 1. Relationships between model simulations of clear sky A-band (left column) and B-band (right column) ratios with surface elevation and relative airmass. a, b) O_2 band ratios as a function of surface elevation; c, d) double logarithm of O_2 band ratios versus surface elevation; e, f) O_2 band ratios as a function of total relative airmass; g, h) double logarithm of O_2 band ratios versus logarithm of total relative airmass; i, j) scatter plot of fitted thresholds and O_2 band

10   ratios. The red points in Panels e-j show the simulations when the difference between θ_s and θ_v is smaller than 6° to mimic the EPIC sun-view geometry. The fitted thresholds are computed with a multivariable linear regression in which double logarithm of O_2 band ratios are expressed as a

function of surface elevation and logarithmic of total relative airmass. The simulations use 4

atmospheric profiles: mid-latitude winter, subarctic summer, subarctic winter, standard US

atmosphere. Surface albedo is set at 0.8 to represent snow and ice surface.

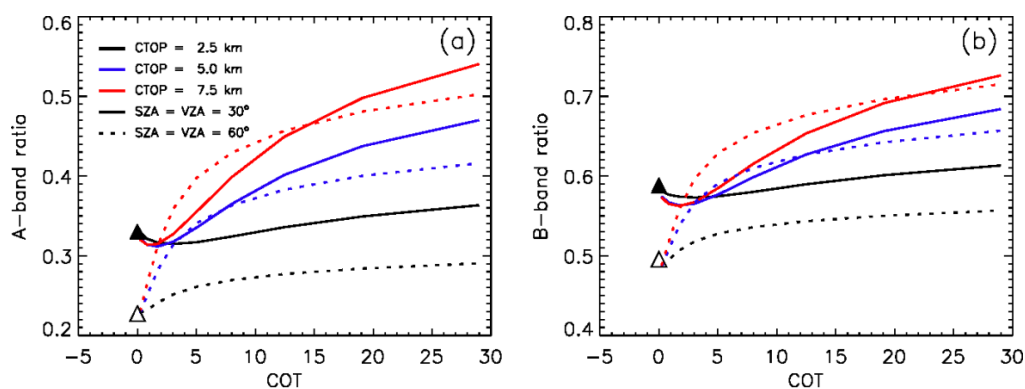
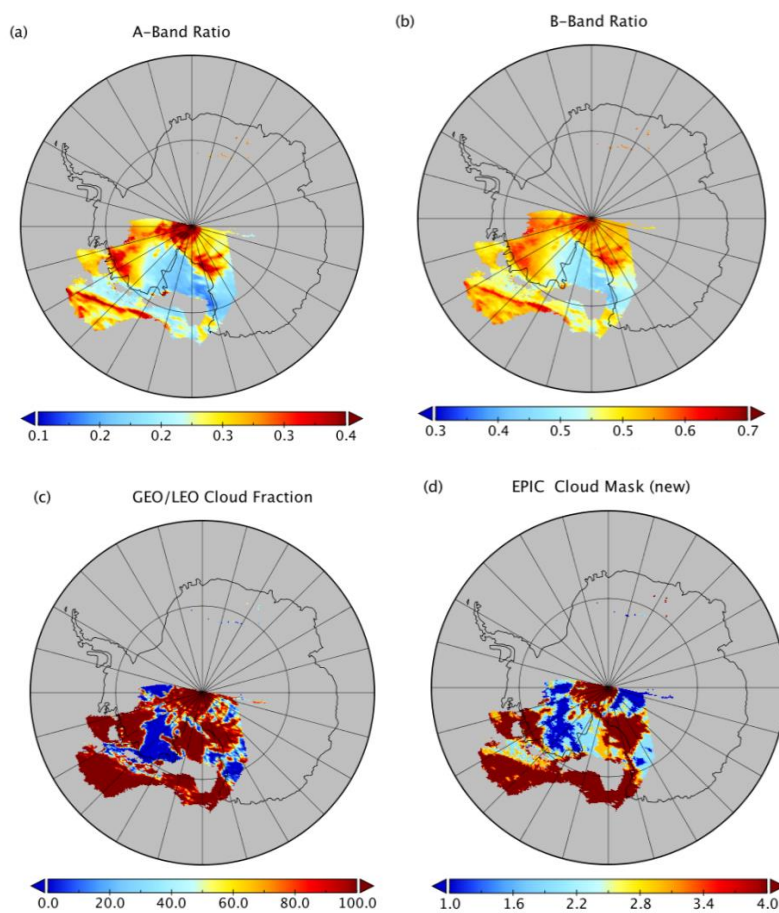



Figure 2. Model simulated Oxygen band ratios as a function of cloud optical thickness (COT) with cloud top height at 2.5 km (black), 5.0 km (blue) and 7.5 km (red) and solar zenith angles at

5   30° (solid line) and 60° (dotted lines), respectively for (a) A-band and (b) B-band. View zenith angle is the same as the solar zenith angle and relative azimuth angle is 160° for all the simulations. The clear sky simulations are marked with filled and unfilled triangles for solar and view zenith angles at 30° and 60°, respectively. Both clear sky and cloudy sky simulations use standard US atmosphere and zero ground elevation. Surface albedo is set at 0.8 to represent snow

10   and ice surface.

Figure 3. Scatter plot of regression fit versus A-band (left) and B-band (right) ratios for clear sky (a, b) and cloudy sky (c, d) pixels from EPIC measurements over global snow and ice surfaces in January and July 2017. The regression is derived with clear sky oxygen band ratio as a function of surface elevation and airmass. The pixels on the right side of black lines will be identified as clear sky and on the left side of the lines cloudy pixels. The dashed lines indicate division of clear and cloud pixels if predicted ratios are increased by 0.025.

5   Figure 4. Section of an EPIC granule on Dec 23, 2017, 1707 UTC time with matching GEO/LEO
overpass within 5 minutes of the EPIC scan over western Antarctic. (a) A-band ratios, (b) B-
band ratio, (c) cloud fraction from GEO/LEO composite, (d) Cloud mask from the new
algorithm.

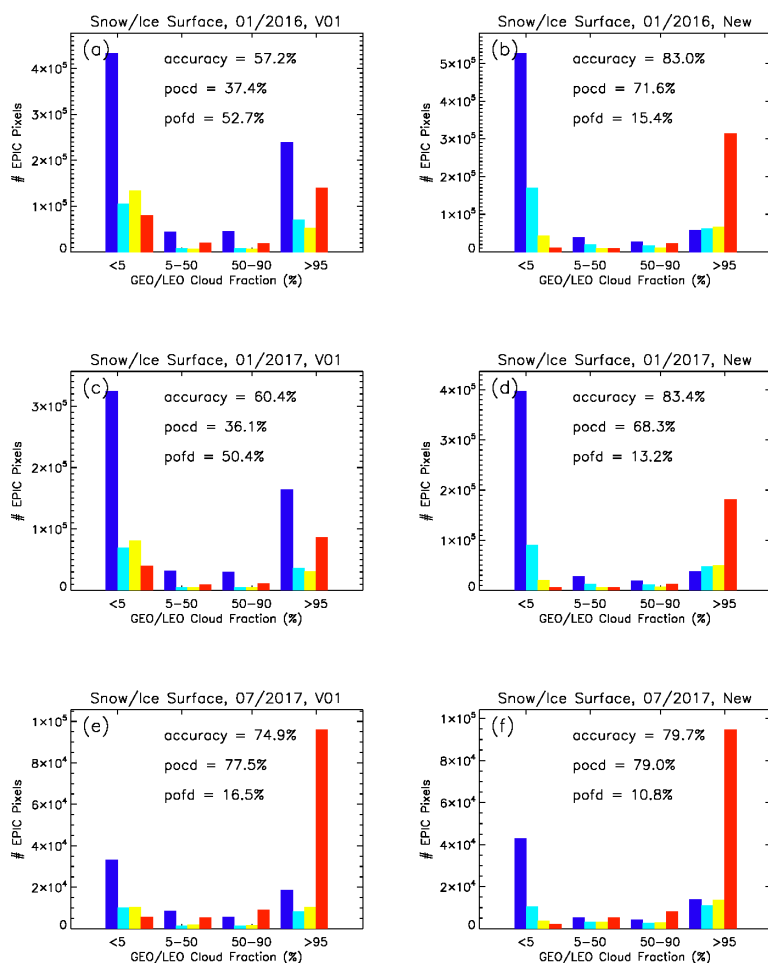



Figure 5. Number of pixels in each pixel-by-pixel matchup category between cloud mask from EPIC and GEO/LEO composite cloud fraction over snow/ice surfaces for January 2016 (a, b), January 2017 (c, d), and July 2017 (e, f). Left is from the current EPIC cloud mask algorithm and the right is from the new algorithm. Blue, cyan, yellow and red bars are for EPIC cloud mask equals to 1, 2, 3, 4, respectively. POCD: probability of correct detection; POFD: probability of false detection.

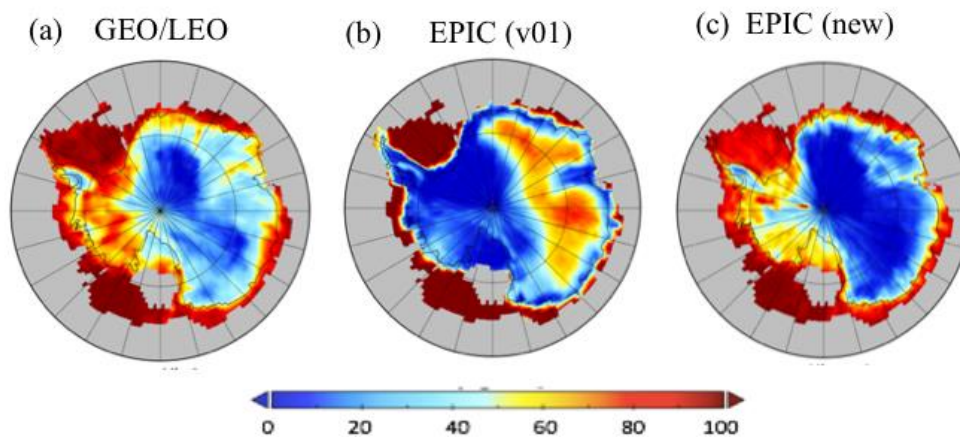



Figure 6. Cloud Fractions derived from (a) composite GEO/LEO retrievals, (b) original EPIC
cloud mask, (c) new EPIC cloud mask over Antarctic in January 2017.

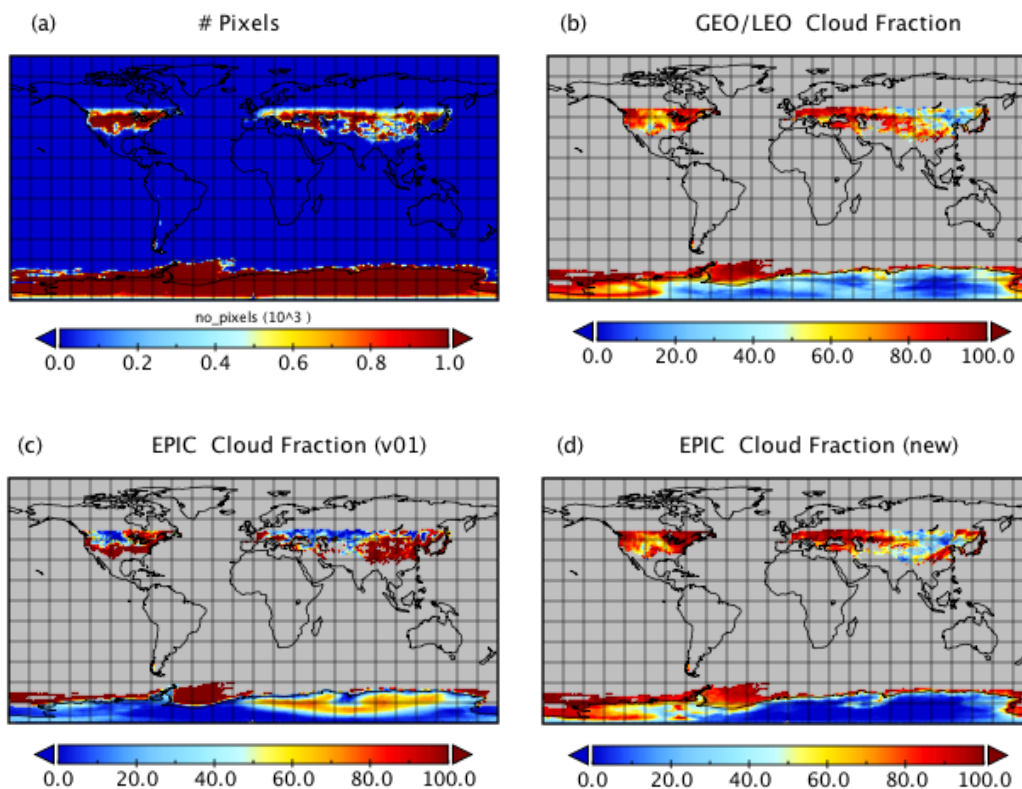
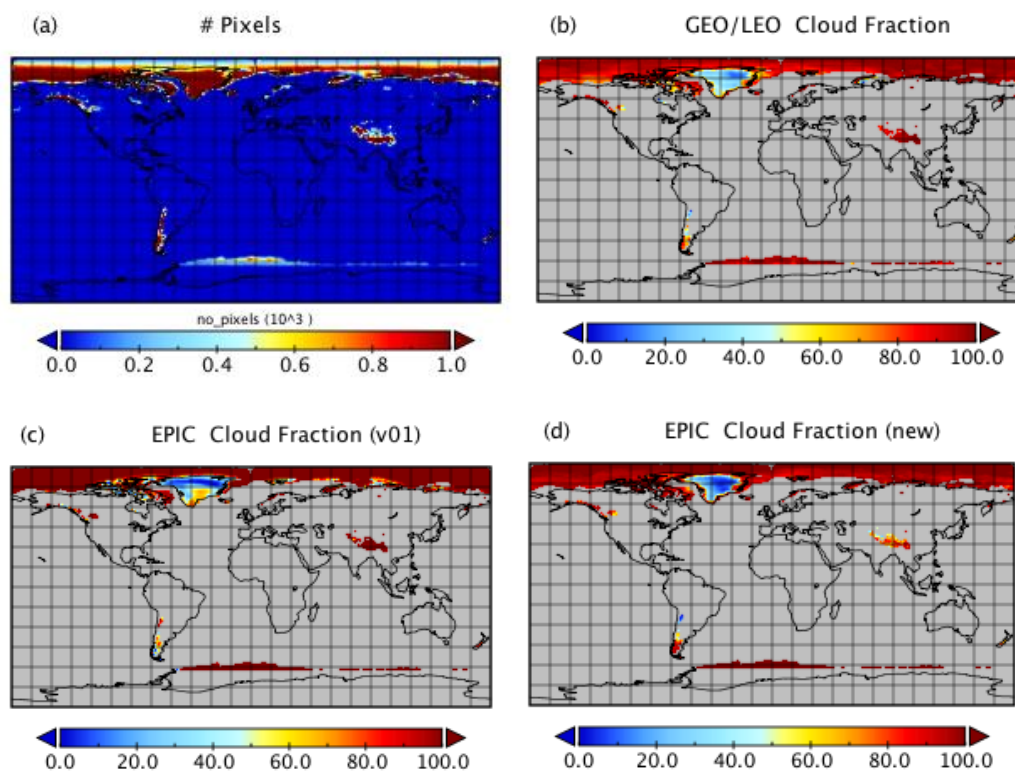
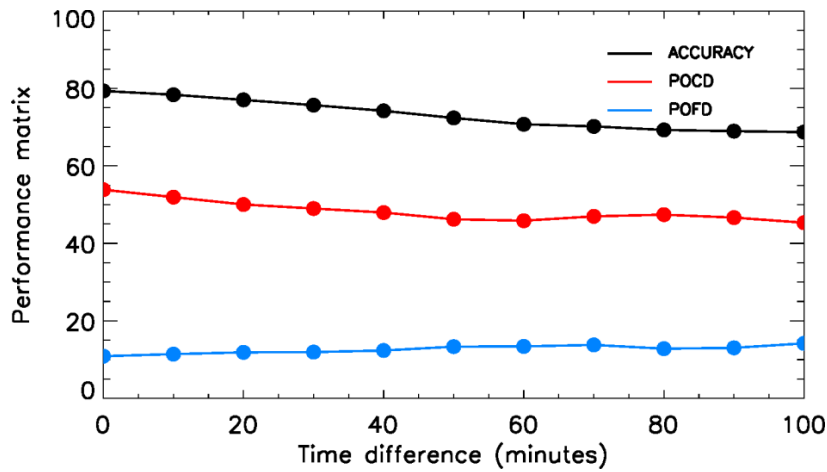



Figure 7. (a) Number of ice/snow pixels and monthly mean cloud fractions derived from (b) GEO/LEO composites, (c) original EPIC cloud mask algorithm, and d) new algorithm in 1° x 1° grids for January 2016.

5   Figure 8. (a) Number of ice/snow pixels and monthly mean cloud fractions derived from (b)
       GEO/LEO composites, (c) original EPIC cloud mask algorithm, and d) new algorithm in 1° x 1°
       grids for July 2017.

5   Figure 9. Performance matrix for January 2017 as a function of time difference between EPIC
and GEO/LEO instrument measurements. POCD: probability of correct detection; POFD:
probability of false detection.